Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 22 (2020), Issue 8 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Original Papers

Proposal

Viewpoints

Short Paper

Letter to the Editor

Corrigenda and Addendas

Review

Conversational Agents in Health Care: Scoping Review and Conceptual Analysis

Lorainne Tudor Car[1,2], MD, MSc, PhD; Dhakshenya Ardhithy Dhinagaran[1], BSc (hons); Bhone Myint Kyaw[1], MBBS, MSc, PhD; Tobias Kowatsch[3,4,5], MSc, PhD; Shafiq Joty[6], MSc, PhD; Yin-Leng Theng[7], PhD; Rifat Atun[8], MBBS, MBA, FRCGP, FFPH, FRCP

[1]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore

[2]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

[3]Future Health Technologies programme, Campus for Research Excellence and Technological Enterprise (CREATE), Singapore-ETH Centre, Singapore

[4]Center for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[5]Center for Digital Health Interventions, Institute of Technology Management, University of St Gallen, St Gallen, Switzerland

[6]School of Computer Sciences and Engineering, Nanyang Technological University Singapore, Singapore

[7]Centre for Healthy and Sustainable Cities, Nanyang Technological University, Singapore

[8]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, United States

Corresponding Author:
Lorainne Tudor Car, MD, MSc, PhD
Family Medicine and Primary Care
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
11 Mandalay Road
Singapore
Phone: 65 69041258
Fax: 65 69041258
Email: lorainne.tudor.car@ntu.edu.sg

Abstract

Background: Conversational agents, also known as chatbots, are computer programs designed to simulate human text or verbal conversations. They are increasingly used in a range of fields, including health care. By enabling better accessibility, personalization, and efficiency, conversational agents have the potential to improve patient care.

Objective: This study aimed to review the current applications, gaps, and challenges in the literature on conversational agents in health care and provide recommendations for their future research, design, and application.

Methods: We performed a scoping review. A broad literature search was performed in MEDLINE (Medical Literature Analysis and Retrieval System Online; Ovid), EMBASE (Excerpta Medica database; Ovid), PubMed, Scopus, and Cochrane Central with the search terms “conversational agents,” “conversational AI,” “chatbots,” and associated synonyms. We also searched the gray literature using sources such as the OCLC (Online Computer Library Center) WorldCat database and ResearchGate in April 2019. Reference lists of relevant articles were checked for further articles. Screening and data extraction were performed in parallel by 2 reviewers. The included evidence was analyzed narratively by employing the principles of thematic analysis.

Results: The literature search yielded 47 study reports (45 articles and 2 ongoing clinical trials) that matched the inclusion criteria. The identified conversational agents were largely delivered via smartphone apps (n=23) and used free text only as the main input (n=19) and output (n=30) modality. Case studies describing chatbot development (n=18) were the most prevalent, and only 11 randomized controlled trials were identified. The 3 most commonly reported conversational agent applications in the literature were treatment and monitoring, health care service support, and patient education.

Conclusions: The literature on conversational agents in health care is largely descriptive and aimed at treatment and monitoring and health service support. It mostly reports on text-based, artificial intelligence–driven, and smartphone app–delivered conversational agents. There is an urgent need for a robust evaluation of diverse health care conversational agents’ formats, focusing on their acceptability, safety, and effectiveness.

(J Med Internet Res 2020;22(8):e17158) doi:10.2196/17158

KEYWORDS

conversational agents; chatbots; artificial intelligence; machine learning; mobile phone; health care; scoping review

Introduction

Background

Conversational agents or chatbots are computer programs that simulate conversations with users [1]. They are increasingly adopted in many different fields, including finance, commerce, marketing, retail, and fitness, with favorable reception from customers [2]. Conversational agents are often deployed via messaging apps, a website, or a mobile phone app. They can also be integrated into cars and television sets or in the form of a stand-alone device such as speakers. They can converse through a range of methods such as text, image, and voice. Conversational agents that can interpret human speech and respond via synthesized voices as well as manage tasks requested by the user are also known as voice assistants. Some of the most popular voice assistants include Apple's Siri, Amazon's Alexa, Google Assistant, and Microsoft's Cortana, mostly delivered using voice-activated or smart speakers such as Amazon's Echo and Google Home. They are utilized for aiding or executing tasks such as web-based shopping, control of smart home devices, and disseminating news or for entertainment [3-5].

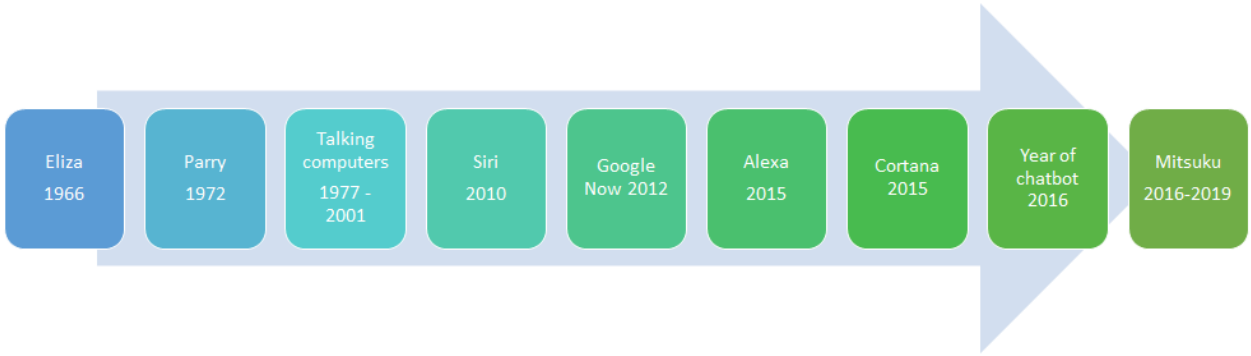
Conversational agents cover a broad spectrum of aptitudes ranging from *simple* to *smart* [2]. Simple conversational agents are *rule based*, meaning that they depend on prewritten keywords and commands programmed by the developer. The user is therefore restricted to predetermined options when answering questions posed by the conversational agents, and there is little or no opportunity for free responses. If a user enters a question or sentence without a single keyword, the conversational agents will be unable to understand the input and will respond with a default message such as "Sorry, I did not understand" [2]. Despite these restrictions, simple conversational agents are increasingly used in executing tasks such as booking appointments, purchasing merchandise, ordering food, and sharing information without the need for human involvement [2].

In contrast, smart conversational agents do not respond with preprepared answers but with adequate suggestions instead. This is enabled by machine learning, a type of artificial intelligence (AI), which allows for broadening of the computer system's capacity through its learning from data (in this case conversations) without being explicitly programmed [2,6]. The process whereby the machine translates human commands into a form in which the computer can understand, process, and revert to the user is called natural language processing (NLP) [6] and natural language understanding or interpretation [6,7]. This degree of programming allows for personalized conversational agents to be generated. Smart conversational agents have the potential to undertake more complex tasks that involve greater interaction, reasoning, prediction, and accuracy. Although the technology behind smart conversational agents is continuously developed, they currently do not have full human-level language abilities, resulting in misunderstanding and users' dissatisfaction [8]. Furthermore, as machine learning algorithms develop, it is becoming increasingly challenging to keep track of their development, evolution, and the reasoning behind their responses. This is known as the *black box effect* [9,10]. Although the black box effect appears to be an unavoidable consequence of the use of AI, there is some emerging research on making AI transparent and explainable [11]. However, at the moment, its use may affect the safety and accuracy of treatment and should be carefully monitored and evaluated when used in health care [9].

The first conversational agent *ELIZA* was developed by Weizenbaum [12] in 1966, with ELIZA taking on the role of a person-centered Rogerian psychotherapist (Figure 1). This was a groundbreaking contribution to the field of AI and was reported to have a positive impact on patients who communicated with the conversational agent [13]. A step up from ELIZA was achieved when *PARRY*, a conversational agent representing a simulated paranoid patient with schizophrenia, was developed [14,15]. These first examples of conversational agents, *chatterbots* (as they were referred to then), in health care were valuable in demonstrating that virtual agents have the potential to mimic human-human conversation and successfully pass the Turing Test, a test of a machine's ability to replicate human intelligence, and the machine passes the test when the tester cannot distinguish it from the human [16].

The literature over the next few decades does not explicitly mention *chatbots* or *conversational agents* in health care, but it does refer to *talking computers* [17-21], a less sophisticated version of today's conversational agents previously used for conducting patient satisfaction surveys [17], altering adult eating habits [18], aiding health care service delivery through diagnosis aid [19], and promoting patient-physician communication [20]. Although not presented in the literature, chatbot Jabberwacky was released in 1988. It was one of the first few AI agents developed for human interaction and entertainment and introduced the shift from text- to voice-operated conversational agents. Soon after, ALICE gained plenty of attention in 1995, after which it went on to win the Loebner Prize 3 times in 2000, 2001, and 2004.

Figure 1. Evolution of conversational agents from 1966 to 2019.



The next big milestone for conversational agents was in 2010 when Apple released *Siri*. The interest in conversational agents increased exponentially at this point as evidenced by Google, Amazon, and Microsoft all developing their own versions over the coming years: Google now, Alexa, and Cortana, respectively [14]. Year 2016 was named the *Year of the Chatbot* as a number of major information technology companies started to use conversational agents: Facebook launched its messenger platform for conversational agents, Google announced its procurement of the conversational agent development tool API.ai, LinkedIn revealed its first messaging bot, and Viber released Public Accounts for chatting with businesses [22-25]. Currently, the title of the world's best conversational agent is held by Mitsuku, a 4-time winner of the Loebner Prize, an annual competition in AI [26].

Health care, which has seen a decade of text messaging on smartphones, is an ideal candidate for conversational agent–delivered interventions. Conversational agents enable interactive, 2-way communication, and their text- or speech-based method of communication makes it suitable for a variety of target populations, ranging from young children to older people. The concept of using mobile phone messaging as a health care intervention has been present and increasingly explored in health care research since 2002 [27]. A series of systematic reviews on the use of text messaging for different health disorders have shown that text messaging is an effective and acceptable health care intervention [28,29]. With a global penetration rate of 96% [28], mobile phones are ubiquitous and avidly used, and can be efficiently harnessed in health care [30]. Conversational agents are increasingly used in diverse fields, including health care, and there is a need to identify different ways and outcomes of the use of conversational agents in health care. Existing reviews on conversational agents focus on a certain subtype of agents such as virtual coaches [31-33] or embodied conversational agents (ECAs) [34] or on specific functionalities of these agents such as behavior change [35] or mental health applications [36,37]. Other reviews report solely on the technical aspects of conversational agents such as system architecture and dialogues [38] or on the funding component of health care conversational interfaces [39].

Objectives

Our objective was to provide a comprehensive overview of the existing research literature on the use of health care–focused conversational agents. We aimed to examine how conversational agents have been employed and evaluated in the literature to date and map out their characteristics. Finally, in line with the observed gaps in the literature, we sought to provide recommendations for future conversational agent research, design, and applications.

Methods

Search Strategy

We adopted methodological guidance from an updated version of the Arksey and O'Malley framework with suggestions proposed by Peters et al [40] in 2015 to conduct our scoping review. To identify literature pertaining to the application of conversational agents in health care, a broad literature search was conducted in April 2019 in MEDLINE (Medical Literature Analysis and Retrieval System Online; Ovid), EMBASE (Excerpta Medica database; Ovid), PubMed, Scopus, and Cochrane Central. Given the novelty of the field, the amount of ongoing research happening in the area, and to increase comprehensiveness, we also searched for the gray literature in the OCLC WorldCat database, ResearchGate, Google Scholar, OpenGrey, and the first 10 pages of Google.

We used an extensive list of 63 search terms, including various synonyms for conversational agents (Multimedia Appendix 1). These synonyms were generated using a web-based search and by identifying specific terms or phrases used in the titles of articles discussing health care conversational agents. The reference list of relevant articles and systematic reviews were also searched for further articles related to the review.

Inclusion and Exclusion Criteria

To map out the current conversational agent applications in health care, we included primary research studies that had conducted an evaluation and reported findings on a conversational agent implemented for a health care–specific purpose. We excluded articles that just presented a proposal for conversational agent development, articles that mentioned conversational agents briefly or as an insignificant part of a review, as well as opinion pieces and articles where primary research was not conducted or discussed. A further point of exclusion was articles with poorly reported data on chatbot assessments where there was minimal or no evaluation data. In addition, we excluded articles concerning ECAs, relational agents, animated conversational agents, or other conversational agents with a visual or animated component.

ECAs are computer-generated virtual individuals with an animated appearance to enable face-to-face interaction between the user and the system [41]. Relational agents are a type of ECA designed to create long-term deep and meaningful relationships with individuals [42]. ECAs are similar to conversational agents in that conversation is central to their function; however, ECAs are more complex as hand movements and facial expressions can be conveyed to the user as well [41]. The user's interaction may be affected by nonverbal behaviors, graphics, and layout of the program, and it was decided that the complexities associated with ECAs are beyond the scope of this review and were therefore excluded.

Screening, Data Extraction, and Analysis

Screening of articles for inclusion was performed in 2 stages: title and abstract review and full article review, undertaken independently by 2 reviewers. Following an initial screening of titles and abstracts, full texts were obtained and screened by 2 reviewers. From the included studies, 2 reviewers independently extracted relevant information in an Excel (Microsoft) spreadsheet. We extracted data on the first author, year of publication, source of literature, title of article, type of literature, study design and methods, geographic focus, health care sector, conversational agent name, accessibility of conversational agent, dialogue technique, input and output modalities, and nature of conversational agent's end goal. We piloted the data extraction sheet on at least five articles. Potential discrepancies in the extracted data were discussed between the authors and resolved through discussion and consensus.

We performed a narrative synthesis of the included literature and presented findings on (1) study specifics, such as study design, geographic focus, and type of literature; (2) conversational agent specifics (ie, conversational agent delivery channel, dialogue technique, personality, etc); (3) conversational agent content analysis; and (4) study evaluation findings.

We used the principles of thematic analysis to analyze the content, scope, and personality traits of the conversational agents. Two researchers familiarized themselves with the literature identified, generated the initial codes in relation to personality and content analysis, applied the codes to the included studies, compared their findings, and resolved any discrepancies via discussion.

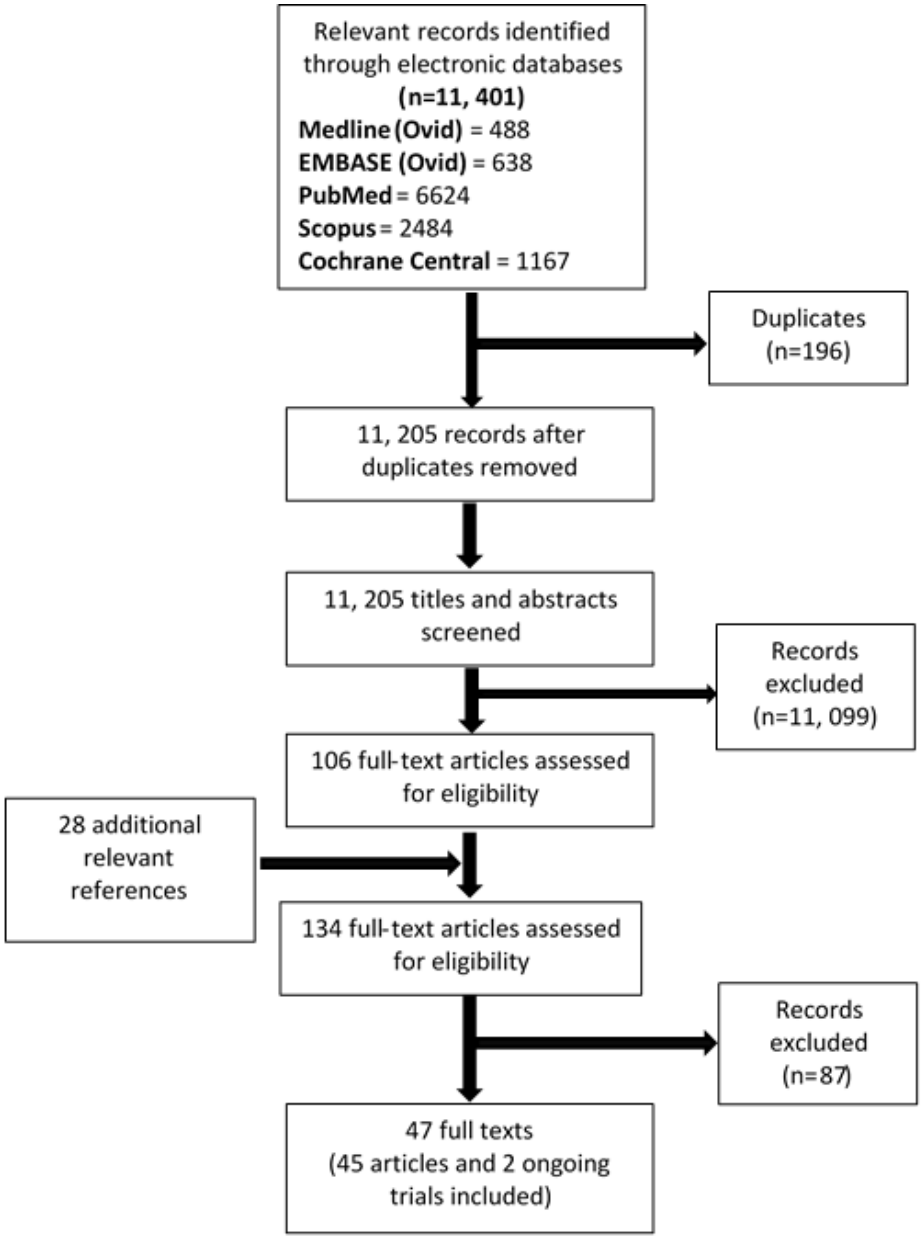
The need to present information on conversational agent personality was motivated by the concepts presented in the study by de Haan et al [43], which posits that personalities are not just limited to humans but can be extended to nonhuman artifacts to explain their actions and behavior [43]. Furthermore, it states that personality traits are especially important in the design of *socially interactive robots*, such as conversational agents. The 5 dimensions of personality presented in this paper were derived from the following: extraversion, agreeableness, conscientiousness, emotional stability, and culture. We have used these headings to guide our analysis of the conversational agents' personality traits in this review. We also aimed to identify and analyze the patterns in the description of conversational agents pertaining to personality traits. Multiple codes were sometimes assigned to the same agent where necessary, but this was limited to a maximum of 3 codes to maintain some degree of specificity.

Results

Search Findings

The initial database searches yielded 11,401 records, and another 28 records were retrieved through additional sources such as the gray literature sources and screening of reference lists of relevant studies. A total of 196 duplicates were identified and removed, leaving 11,233 titles and abstracts that needed to be screened. Title and abstract screening led to the exclusion of 11,099 records, resulting in 134 full texts that needed to be assessed for eligibility. Of these, 87 articles were excluded, resulting in a final pool of 47 reports comprising 45 studies and 2 ongoing trials (Figure 2).

Figure 2. PRISMA flow chart.

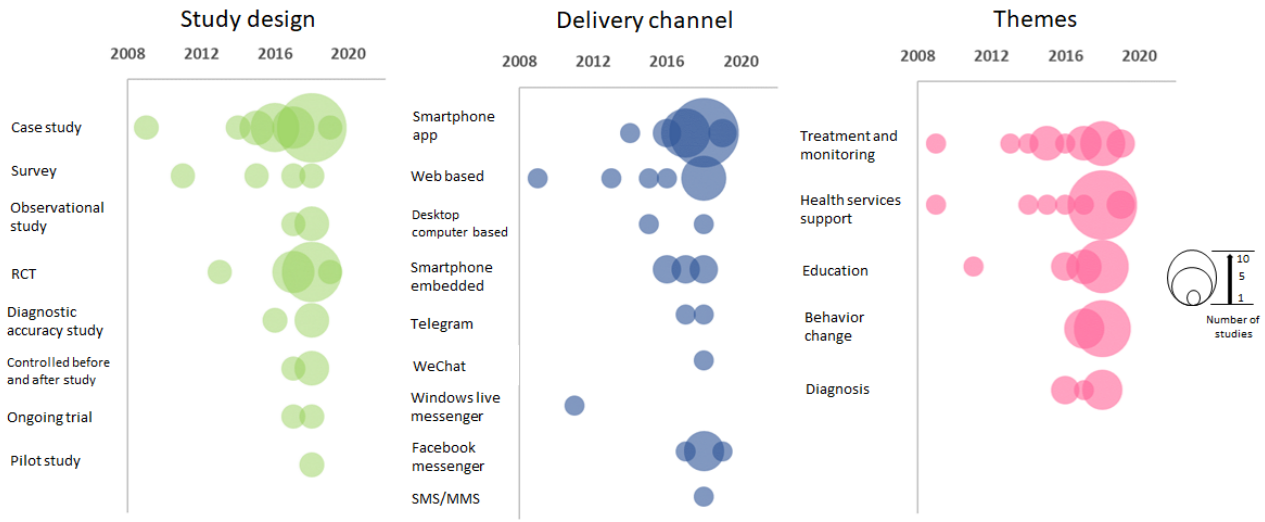


Characteristics of Included Studies

In this scoping review, 40 included studies were from high-income countries (HICs) and 6 were from low- and middle-income countries (LMICs). A total of 22 studies were from European countries, including Italy [44,45], Switzerland [30,46-52], France [53,54], Portugal [55], The Netherlands [56], the United Kingdom [57-61], Spain [62,63], and Sweden [64]. Moreover, 8 studies originated from Asian countries: Philippines [65], China [66], Japan [67,68], Pakistan [69], India [70,71], and Hong Kong [72]. Other geographic regions acknowledged in the studies of this review were Australia [73,74], Canada [75], New Zealand [76,77], South Africa [78], and the United States of America [79-89].

A variety of study designs were used in the included studies, comprising 20 case studies [44,48,51,61-63,66,69,71,73-79,82,84,85,89], 4 surveys [55,56,59,65], 3 observational studies [53,86,87], 11 randomized controlled trials [46,49,50,57,64,67,72,80,81,83,88], 3 diagnostic accuracy studies [58,60,68], 3 controlled before and after studies [30,45,70], 2 ongoing trials [51,54], and 1 pilot study [47] (Figure 3).

Figure 3. Bubble plots showing the distribution of identified study designs, types of conversational agents and healthcare topics in the included articles, plotted against the year of the publication. The scale on the right indicates that the size of the bubble is associated with the number of studies whereby the smallest denotes 1 study and the largest, 10 studies.



The types of literature included 25 journal articles [44,48,55-57,61-65,67,69,72,74-76,80-87,89], 11 conference abstracts [45,47,49,50,52,59,70,71,73,78,79], 4 conference papers [30,46,66,77], 1 poster abstract [68], 4 electronic preprints [53,58,60,88], and 2 clinical trial protocols [51,54].

There was an increase in the number of publications each year, from 3 in 2015 to 5 in 2016, 10 in 2017, and 23 in 2018. Some author groups were highly productive and published at least two papers within 2 years. Kowatsch et al published 3 papers between 2017 and 2018 based on their open source behavioral intervention platform MobileCoach, which allows the authors to design a text-based health care conversational agent for obesity management and behavior change [30,46,90]. Griol et al published articles on conversational agent for chronic conditions, including chronic pulmonary disease [63] and Alzheimer disease [62] in 2015 and 2016, respectively. Such productive teams reiterate the research interest in this area of conversational agents. Furthermore, the high frequency of publication indicates the feasibility and support to conduct research successfully in this area.

Characteristics of Conversational Agents in the Included Studies

Conversational Agent Delivery Channel

Conversational agents were delivered through a variety of means in the included studies. Most (n=23) were smartphone apps [30,46-50,53,55,58-61,64,67,70,71,75,77,81,83,85,86,88]; web based (n=5) [57,66,73,74,82]; desktop computer based (n=2) [65,79]; used smartphone-embedded software (n=6; eg, Siri, Google Assistant, Alexa, etc) [44,51,62,76,84,87], Telegram [45,78], WeChat [72], SMS and multimedia messaging service [89], Windows live messenger [56], or Facebook Messenger [52,80]; and 4 were made available on more than 1 platform [53,59,68,83]. Three studies did not specify the method of conversational agent delivery [54,63,69].

Technical Development Approach

A total of 8 studies made a reference to the technical details of the conversational agent development process. Some mentioned specific tools such as C and MS Access [65]. Others discussed the application of well-known concepts, to conversational agent development such as using the Computers are Social Actors paradigm to develop a health advice conversational agent, or converting the structure association technique (SAT) into digital SAT for implementation on a LINE platform [67,83]. Some emphasized data set creation and sources for the knowledge base [44]. Four studies provided an in-depth workflow with a step-by-step explanation of the technical development of the conversational agent. Cheng et al [79] provided a very detailed technical explanation of the development process—broken down and explained in parts: program development on Google's home device, webhook and internal logic, and web interface. Galescu et al [82] described the CARDIAC system architecture including a knowledge base, task models, dialogue management, speech recognition, and language generation. Griol et al [63] presented a spoken dialogue system with specific details of the proposed emotion recognizer. For example, it considers pitch, frequency, energy, and rhythm of speech input from the user. Joerin et al [75] provided a less technically dense explanation for chatbot conversational agent development but mentioned technologies used in the process, such as emotion algorithms and machine learning techniques [75].

Input and Output Modalities

The conversational agents could be categorized according to whether the user input was fixed (ie, predetermined text) or unrestricted (ie, free text/speech). A total of 10 studies employed fixed text user inputs [30,46,47,49,50,52,54,58,83,88], with 2 additional studies enabling fixed text and image inputs [67,68]. Moreover, 19 studies allowed free text user inputs [45,48,51,56,57,60,61,66,69,70,72,74,77,78,80,81,85,86,89], and 4 studies used both fixed and free text user inputs

[53,64,65,73]. Speech was enabled in 8 studies [44,55,63,71,76,79,82,84], whereas free text and speech were employed in 3 studies [62,75,87]. The method of user input was unspecified in 1 study [59] (Multimedia Appendix 2).

Similarly, output modalities largely employed text alone (n=30) [45-47,49-51,53,54,56-58,60,61,64-66,68-70,72-74,77,78,80,81,83,85,88,89]; text and speech (n=5) [48,55,63,71,87]; speech alone (n=4) [44,79,82,84]; text and images (n=4) [30,67,75,86]; text, speech, and images [62]; or text, speech, images, and videos [52,76]. The input and output methods were not specified in 1 of the studies [59] (Multimedia Appendix 2).

Conversational Agent Personality

We condensed the descriptive terms used in individual studies to present the conversational agents into a list of 9 relevant personality traits as presented in Table 1.

The conversational agents in the included studies were health care professional like [57,58,62,66,71,73,74,86], informal [46,52,53,56,61,65,81,85], coach like [47,49,52,64,66,70,80], knowledgeable [56,60,68,72,89], human like [48,78,79,88], culture specific [47,48,53], factual [68,76], gender specific [46,78], and some identified explicitly as a conversational agent [46,65].

One article [78] reported on a conversational agent personality that was criticized for being overly formal, and some articles did not report on the personality of the conversational agent at all [30,44,45,50,51,54,55,59,63,67,69,75,77,82-84,87].

Table 1. Personality codes derived for the conversational agents included in this review, adapted from Haan et al.

Personality codes	Descriptions
Coach like	Encouraging, motivating, and nurturing
Conversational agent identity	Explicitly identifies as a conversational agent
Culture specific	Speaks the native language or has native names
Factual	Nonjudgmental, no personal opinions, and responses based on facts or observations
Gender specific	Male and female versions available
Health care professional like	Designed to be a doctor or expert, that is, mimics a health care professional
Human like	Tries to emulate humans, for example, participants reported feeling like they were talking to another human or researchers used features like "typing" to make the conversation more human like
Informal	Informal, like talking to a friend. Uses exclamations, abbreviations, and emoticons
Knowledgeable	Content created or informed by medical experts

Human Involvement

A health care administrator or professional was available via the conversational agent for the user to communicate with in some studies. The role of the human varied from an administrator who could be contacted via a dedicated chat channel for the user to ask questions or an individual whose role was to monitor the user's activity on the conversational agent and provide personalized feedback to them. Seven studies [30,46,47,70,72,78,85] reported on human involvement in the conversation and the remaining articles did not.

Conversational Agent Goals

All the conversational agents in this review were identified as *goal oriented*. Goal-oriented conversational agents have a clearly defined end point and are employed to execute a specific function, unlike *chit chat* agents that have no specific end goal, do not delve into the details of any topic, and have a primary aim of merely keeping the conversation going [91]. Goal-oriented conversational agents were further divided into those that yielded long- or short-term outcomes. Of the included studies, 22 articles focused on conversational agents with long-term goals and 23 with short-term goals (Multimedia Appendix 3 [30,44-89]). Two studies reported on conversational agents with both short-term and long-term goals [45,56], for example, answering immediate queries (short) and providing education and increasing users' knowledge on the topic over time (long) [56]. Conversational agents with short-term scope provided users with a response or service almost instantaneously, such as answering health-related queries [84]. Conversely, those with long-term scope needed to build a relationship with the user, over time, to help them overcome health-related issues such as smoking cessation [72] or working through a mental health problem [80].

Conversational Agent Content Analysis

Five distinct themes were identified in terms of conversational agent content: treatment and monitoring (ie, treatment implementation, management, adherence, support, and monitoring), health service support (ie, connecting patients to health care services), education (ie, provision of health care–related information), lifestyle behavior change (ie, supporting users in tackling various modifiable health risk factors), and diagnosis (ie, identification of the nature of a disease or a condition). A number of included conversational

agents spanned several different themes (Multimedia Appendices 3 and 4 [30,44-89]).

Treatment and Monitoring

Overall, 17 articles reported on conversational agents that focused on treatment, monitoring, or rehabilitation of patients with specific conditions. One study reported on a conversational agent to help preserve cognitive abilities in those with Alzheimer disease [62]. Two other studies focused on conversational agents to provide support and treatment for metabolic conditions such as type 2 diabetes [70] and obesity [46]. Eight studies presented conversational agents for managing mental health using techniques such as counseling [67]; cognitive behavioral therapy (CBT) [64,80] method of levels therapy [57]; positive psychology [61]; provision of a virtual companion [66]; and a combination of modalities such as CBT with mindfulness-based therapy, emotionally focused therapy, and motivational interviewing [75,81]. One study each reported on the use of a conversational agent for monitoring patients with asthma [85], HIV [45], heart failure [82], and chronic respiratory disease management [63]. Non–disease-specific conversational agents were used as a health information advisor [83] and pediatric generic medicine consultant [65].

Health Care Services Support

Overall, 19 studies reported on conversational agents used to support or complement existing health care services. These tasks included remote delivery of health care services for mental health support [67,75,81], breast cancer [53,54], dysarthria [44], obesity [50], diabetes management [79], chronic respiratory diseases [63], asthma [85], heart failure [82], and HIV management [45]. Other studies discussed conversational agents automating health care services such as patient history taking [48,77], providing health advice [83], symptom checking [58], and triaging and diagnosis support [60,69,74].

Education

We found 13 articles in which conversational agents were used primarily for educating patients or users. Education focused on topics such as sexual health [59,76] including information on HIV [78], overcoming unhealthy habits such as alcohol misuse [73] and smoking cessation [72], improving well-being [88], diabetes management [79], breast cancer [53,54], and medication-related queries [55] as well as general health [56,84,87], which covered more than 1 topic of focus, for example, education on sex, drugs, and alcohol for adolescents.

Lifestyle Behavioral Changes

We identified 12 studies with conversational agents for healthy lifestyle behavior change in the general population as well as overweight and obese individuals. Two studies discussed conversational agents for the management of obesity in younger patients, including adolescents [46,50]. They largely employed a coach-like conversational agent to promote physical activity [51] and healthy eating [52], sometimes with incentive provision, and provided techniques on how to reverse obesity [30,47,49,71]. Other behavioral change interventions used a social media–driven conversational agent for smoking cessation [72], a health coach for diabetes prevention [86], a reflection companion to encourage physical activity in adults [89], and emotionally intelligent agents to improve mental health [61] and well-being [88].

Diagnosis

Seven articles presented health care conversational agents with a primary purpose of establishing a diagnosis. Three articles reported on conversational agents' triage, diagnosis, or a combination of both, mainly employing a symptom checker function [58,60,74]. Three more studies reported purely on the diagnostic accuracy of 2 conversational agents [69,71,77]. One article reported on a conversational agent for diagnosing sexually transmitted infections to overcome barriers such as social stigma, embarrassment, and discomfort associated with traditional diagnostic approaches that require a medical interview with a health care professional [68].

Conversational Agent Evaluation

Included studies that evaluated conversational agents reported on their accuracy (in terms of information retrieval, diagnosis, and triaging), user acceptability, and effectiveness. Some studies reported on more than 1 outcome, for example, acceptability and effectiveness. In general, evaluation data were mostly positive, with a few studies reporting the shortcomings of the conversational agent or technical issues experienced by users. Seventeen studies presented self-reported data from participants in the form of surveys, questionnaires, etc. In 16 studies, the data were objectively assessed in the form of changes in BMI, number of user interactions, etc. In 12 studies, there was a mixture of self-reported and objectively assessed outcomes and outcomes were not reported in the two ongoing trials (Multimedia Appendix 4).

Accuracy: Information, Diagnosis, and Triaging

Eleven studies reported on the accuracy of conversational agents [44,58,60,66,68,69,71,74,76,77,82] (Multimedia Appendix 4). Middleton et al [58] and Razzaki et al [60] evaluated 2 versions of the Babylon conversational agent, respectively: *Babylon check* and *Babylon chatbot for triage and diagnosis*. In both studies, the conversational agents were tested on their triage and diagnostic accuracy using clinical vignettes as in the Membership of the Royal College of General Practitioners exams, and their performance was compared with that of doctors. The conversational agents were found to be more accurate, faster, and provided safer triage and diagnosis compared with doctors and nurses. Similarly, Ghosh et al [74] and Danda et al [71] assessed conversational agents on their general diagnostic accuracy, and these had a precision rate of 82% and 86%, respectively. Ni et al [77] assessed Chatbot MANDY, designed to automate patient intake, on its ability to adequately diagnose the patient based on their symptoms. There was a prediction accuracy of 100%, 64%, 25%, and 14% for respiratory issues, chest pain, headache, and dizziness, respectively [77]. Furthermore, 2 studies tested the accuracy of conversational agents employed for sexual health purposes [68,76]. The conversational agent used by Kobori et al [68] diagnosed sexually transmitted infections with an accuracy of 77.7% and had high effectiveness (97.7%) in encouraging patients to visit the clinic earlier. In contrast, Wilson et al [76] compared smart assistants—Google Assistant, Siri, and Google

search—to determine their accuracy in responding to queries around sexual health. The Google search option was found to provide the best answers and also had the lowest failure rate [76]. Another study compared 3 known virtual assistants—Siri, Google Assistant, and Amazon Alexa—on their abilities to recognize speech from individuals with dysarthria [44]. They all performed similarly (50-60% recognition), with Siri being the only agent attempting to parse all the dialogue inputted [44]. Two studies discussed the accuracy of 2 conversational agents in making diagnoses in children and adolescents [66,69]. Teenchat had a 78.34% precision rate in diagnosing stress [66], whereas Aquabot had an accuracy of 85%, 86.64%, and 87.2% (3 groups aged 18-28 years) for achluophobia and 88%, 87.6%, and 87.53% (3 patient groups aged 1-7 years) for autism [69]. Finally, Galescu et al [82] discussed the accuracy of a conversational agent *CARDIAC* in speech recognition for heart failure patients. A significant number of errors were detected and attributed to insufficient vocabulary coverage in the language model as evidenced by an *out-of-vocab* rate of 3% [82].

Effectiveness

The effectiveness of health care conversational agents was assessed in 8 studies [47,52,57,61,70,75,81,84]. Furthermore, 10 studies reported on the effectiveness and acceptability, of which 5 are presented here [49,64,67,80,86] and the remainder are presented under *Acceptability* (Multimedia Appendix 4). Five studies described conversational agents targeting a healthy lifestyle change specifically for healthy eating [52], active lifestyle [49], obesity [47], and diabetes management [70,86]. Casas et al [52] reported improvements in food consumption, whereas Stasinaki [47] and Heldt et al [49] noted increases in physical activity performance with high compliance. Shaikh et al [70] reported successful reduction in HbA_{1c} (glycated hemoglobin) levels postengagement with *Wellthy diabetes*, whereas Stein et al [86] reported successful weight loss (2.38%) and satisfaction was high, rated at 87% for the diabetes prevention chatbot.

Eight studies noted the effectiveness of conversational agents for mental health applications [57,61,64,67,75,80,81,84]. The conversational agent *Tess* by Fulmer et al [81] initiated a statistically significant improvement in depression and anxiety compared with the control group. Two studies looked at the use of machine learning–based conversational agents for CBT in young adults [64,80]. The conversational agent was both effective (reduced levels of depression and perceived stress and improved psychological well-being) and well received (high engagement with the chat app and high levels of satisfaction) [64,80]. This positive effect was reproduced by Joerin et al [75], where emotional support from Tess decreased symptoms of anxiety and depression by 18% and 13%, respectively [75]. Inkster et al [61] employed the Patient Health Questionnaire-9 self-reported depression scale to note significant improvements in depression scores in the high user group compared with the low user group [61]. In addition, 67.7% of users found the app usage to be helpful and encouraging [61]. In the study by Kamita et al [67], the counseling bot encouraged significant improvements in users' self-esteem, anxiety, and depression compared with the control condition. Besides effectiveness, user ratings of acceptability, using the technology acceptance model, were higher in the conversational agent condition compared with the control [67]. Gaffney et al [57] proposed a conversational agent *MYLO* that was significantly better than the existing conversational agent ELIZA in problem solving and helpfulness, but both were equally effective in lowering distress. Miner et al [84] compared Apple's Siri, Microsoft's Cortana, Samsung's S Voice, and Google Now on their abilities to respond to questions about mental health, interpersonal violence, and physical health. Siri responded appropriately and empathetically to issues concerning depression and physical health, and Cortana responded appropriately and empathetically to matters involving interpersonal violence [84].

Acceptability

A total of 26 studies commented on the acceptability of conversational agents (Multimedia Appendix 4). Five studies commenting on acceptability and effectiveness were discussed above [49,64,67,80,86] (see the *Effectiveness* section), and the remaining 21 studies are presented here [30,45,46,48,50,53,55,56,59,62,63,65,72,73,78,79,83,85,87-89]. Several studies (n=6) were targeted at children or adolescents. Three studies discussed conversational agents for health education on medication, asthma management, drugs, sex, and alcohol [56,65,85]. Acceptability was generally denoted by high response rates and scores like *strongly agree* or *agree* for user-friendliness, appropriateness, consistency, and speed of response [65]. In addition, users in the study by Crutzen et al [56] favored the conversational agent over existing methods of information provision. In another 3 studies, conversational agents were employed for the management of obesity in adolescents [30,46,50]. Acceptability was high in all studies, as evidenced by enjoyment of the chats; bonding; formation of social and emotional relationships; and high perceived ease of use, usefulness, and intention to use [30,46,50]. In the study by L'Allemand et al [50], high compliance was attributed to the rewarding game system.

In 4 studies, health care conversational agents were targeted at chronic conditions [55,62,63,79]. The specific conditions addressed were Alzheimer disease, diabetes, heart failure, and chronic respiratory disease. In the study by Cheng et al [79], users responded positively, particularly to features of conversational agents that allowed for personalization and the conversational agent's ability to understand and respond to natural conversation flow. Some difficulties included learning commands, restricted answer options, slow processing speed, and some problematic responses [79]. Lobo et al [55] reported user acceptability in the form of usability, where the conversational agent had a system usability score of 88, which was considered *very good*. Griol et al [62] considered the Alzheimer patients' caregiver's perspective when judging the acceptability of the conversational agent. The global rate for the system (on a scale from 0 to 10) was 8.6, and the application was thought to be attractive, adequate, and appropriate for its purpose. In another study, Griol et al [63] employed an emotionally sensitive conversational agent for chronic respiratory disease patients who rated this agent significantly higher for interaction rate, frequency, and empathy than the baseline version.

A further 3 studies were concerned with sexual health and/or HIV management [45,59,78]. They indicated that in this field, conversational agents could be used for a variety of functions such as booking an appointment, getting test results, therapy, and event reminders [45]. In addition, the conversational agent in the study by van Heerden et al [78] was well received when used as a counseling tool because it was given an avatar-like profile image and the conversation was embedded in a familiar chat interface, which users associated with talking to another human being. In the study by Nadarzynski et al [59], users favored the conversational agent because of its ubiquity as a convenient smartphone app and its ability to perform remote services such as video consultation, potentially alleviating any inhibitions users may have around discussing sexual health in person.

Two studies employed an emotionally sensitive conversational agent for mental health counselling and general health information advice [83,88]. In the study by Liu et al [83], the sympathetic conversational agent was rated more positively than the advice-only condition. Another conversational agent for well-being improvement procured positive feedback from participants who thought it was *an interesting experience*, *pretty quick*, and *fun* [88].

In 3 studies, conversational agents were used for healthy behavior change, specifically targeting smoking cessation, alcohol misuse treatment, and physical activity promotion [72,73,89]. For smoking cessation, participants indicated enjoyment when conversing with the conversational agent, and effectiveness was also insinuated by 38.3% reporting not having smoked in the past week and 69.4% admitting to a reduction in smoking frequency [72]. In the study by Elmasri et al [73], the participants (young adults) reported a higher satisfaction rate with the use of the conversational agent to manage and treat alcohol misuse. For physical activity promotion through the use of a reflection companion, response rates were high (96% at baseline, 90% at follow-up), insinuating high engagement throughout the study. Furthermore, use of the system beyond the stipulated study period was an indicator of viability. Moreover, 16 of the 33 participants opted to continue without any reward, suggesting participants found some added value in using the conversational system [89].

Two studies examined the acceptability of conversational agents for health care service delivery [48,87]. Outcomes were reported qualitatively, including comments on ease of use, humanity of the chatbot, and users' comfort with the input functionalities available to them as well as criticisms on technical difficulties [48]. Bickmore et al [87] more specifically compared conversational assistants Siri, Alexa, and Google Assistant on their provision of health information and found satisfaction to be lowest with Alexa and highest with Siri. Overall, there was a neutral rating for satisfaction, with a median score of 4 (IQR 1-6) [87].

One study discussed a condition-specific conversational agent application targeted at improving the quality of life and medication adherence of breast cancer patients [53]. Participants implied a positive experience when interacting with the conversational agent, whereby 88% said it provided them with support in tracking their treatment and mentioned that they would recommend the conversational agent to their friends. There was an overall satisfaction of 94% [53].

Discussion

Principal Findings

Our scoping review identified 45 studies and 2 ongoing clinical trials. Although conversational agents have been widely employed in various fields, their use in health care is still in its infancy, as evidenced by the study findings that indicate much of the literature being published recently (2016-2018). Most conversational agents used text input and were machine learning based and mobile app delivered. The 3 most commonly reported themes in the health care conversational agent–related literature were treatment and monitoring, health services support, and patient education. Results from the studies evaluating conversational agents were generally positive, reporting effectiveness, accuracy, and acceptability of the conversational agent. However, there is currently a dearth of robust evaluations and a predominance of small case studies.

Our review shows that most of the health care conversational agents reported in the literature used machine learning and were long-term goal oriented. This suggests that conversational agents are evolving from conducting simple transactional tasks toward more involved end points such as long-term disease management [80] and behavior change [30]. The majority of the conversational agents identified in this review targeted patients, with only a few aimed at health care professionals, for example, by automating patient intake or aiding in patient triage and diagnosis. In addition, research into the use of conversational agents to support both formal and informal caregivers is limited and could be a productive area to explore, given that previous systematic reviews on the use of digital technology for caregivers of patients with psychosis [92] or dementia [93] have shown positive outcomes.

Our findings show a predominance of text-based conversational agents, with only a few apps using speech as the main mode of communication. Yet, certain populations, such as older people, may be more comfortable interacting via speech, as some individuals may find the dexterity involved with typing on small keypads on smartphones challenging and time consuming. Furthermore, most conversational agents included in our review were app based. Research shows that the use of apps (which need to be downloaded and regularly updated) is often associated with high dropout rates and low utilization [94]. Such disadvantages do not seem to apply to messaging apps such as Facebook Messenger, iMessage, Telegram, WeChat, or WhatsApp, which are already commonly used in the general population. Future research should aim to overcome this limitation brought on by smartphone apps by embedding future health care conversational agents in platforms, which the target population already uses regularly. The advantage of having numerous publishing platform options is the novelty of conversational agents over smartphone apps, and this should be further explored.

A recent systematic review on the effectiveness of ECAs and other conversational agents noted a lack of an established method for evaluating health care conversational agents in health care and a dearth of data on adverse effects [32]. This corresponds to our findings, with most studies being case studies and lacking information on potential adverse effects. Side effects to consider may relate to the content of the conversational agent conversations, which may not be accurate, evidence based, or suitable for the specific circumstance. For example, if a mental health conversational agent user has suicidal tendencies, the conversational agent may not be best equipped to handle such a situation and may provide inappropriate advice, leaving the user at fatal risk. Additional unwanted effects could arise from the black box effect associated with the use of machine learning–based conversational agents, whereby their suggestions are somewhat unpredictable [95]. Furthermore, conversational agents allowing for free text input may lead to significant privacy concerns, especially for vulnerable populations, as individuals can share private and sensitive data in conversations [96]. There is a need for stringent certification from a regulatory board in cases where conversational agents are given roles akin to health care professionals.

The health care sectors for conversational agent application identified in the review were generally very broad, with references to only a few specialties including mental health [97], neurodegeneration [62], metabolic medicine (obesity [47] and diabetes [70,79]), and sexual health [68]. Future applications could expand toward other health care fields where evidence has suggested potential for digital health interventions such as dermatology [98], primary care [99], geriatrics [100], and oncology [101].

There is also a need for more geographically diverse research. Although our review identified 12 articles with a geographical focus in Asia, the evidence stemming from middle-income countries was scarce, and there were no studies from a low-income country. However, digital health initiatives are becoming more common in developing countries, often with a different, context-specific scope, such as ensuring access to health care using social media [102]. To ensure safe and effective use of solutions developed in HIC settings, there is a need for more research to corroborate the safety, effectiveness, and acceptability of these agents in LMICs too. Furthermore, it is important to explore the integration of conversational agents into the existing health systems and services. A *hybrid system*, where digital technology supplements health care services, is increasingly seen as the optimal solution [103]. This mirrors our acknowledgment that conversational agents will be most advantageous in supporting rather than substituting health care professionals. In most studies, conversational agents were developed and presented independently, unsupported by humans, and separate from the existing health care delivery models, which may prove unsustainable in the long run. Future research should consider evaluating *hybrid systems* encompassing conversational agents in their health care delivery, as reported in some of the included studies where conversational agents were complemented by frequent meetings and phone calls with the physicians.

Although the studies reported accuracy, efficacy, effectiveness, and acceptability as outcomes, there were no measurements of cost, efficiency, or how the solution led to improved productivity when used instead of or to augment the work of a health professional. Therefore, it was not possible to ascertain whether the solutions developed were cost-effective compared with alternative approaches.

Strengths and Limitations

We conducted a comprehensive literature search of multiple databases, including gray literature sources. We prioritized sensitivity over specificity in our search strategy to capture a holistic representation of conversational agent usage uptake in health care. However, given the novelty of the field and the employed terminology, some unpublished studies discussed at niche conferences or meetings may have been omitted. Furthermore, although classification of the themes of our conversational agents was based on thorough analysis, team discussions, and consensus, it might not be all inclusive and may require further development with the advent of new conversational agents. In addition, although some conversational agents belong to more than 1 theme, we mostly classified them based on the dominant mode of application for the sake of clarity. Finally, we excluded articles with poorly reported data on chatbot assessments; therefore, we may have missed some health care conversational agents (Multimedia Appendix 5 [36,97,104-188]). We decided to exclude these because they did not appear to contribute anything additional or noteworthy to our review. The personality traits presented were guided by a reference paper on chatbot personality assignment [43] and also a condensation of descriptive terms from several articles. The lack of depth and breadth in the description of the content and development of many conversational agents led us to organically develop a framework for this paper. This framework is, therefore, still exploratory and adapted to suit the purposes of this review and may well be explored and further refined with more in-depth analysis such as previously published frameworks [189].

Conclusions

Conversational agents are an up-and-coming form of technology to be used in health care, which has yet to be robustly assessed. Most conversational agents reported in the literature to date are text based, machine learning driven, and mobile app delivered. Future research should focus on assessing the feasibility, acceptability, safety, and effectiveness of diverse conversational agent formats aligned with the target population's needs and preferences. There is also a need for clearer guidance on health care –related conversational agents' development and evaluation and further exploration on the role of conversational agents within existing health systems.

Acknowledgments

This research is supported by the Ageing Research Institute for Society and Education (ARISE), Nanyang Technological University, Singapore. This study is also supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) program.

Authors' Contributions

LTC conceived the idea for this study. DD, BK, and LC screened the articles. DD, BK, and LC extracted and analyzed the data. DD and LC wrote the manuscript. BK, TK, JR, RA, and YLT revised the manuscript critically.

Conflicts of Interest

TK is affiliated with the Center for Digital Health Interventions, a joint initiative of the Department of Management, Technology, and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. TK is also a cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. Other authors declare that they have no competing interests.

Multimedia Appendix 1
Search strategy.
[DOCX File , 18 KB - jmir_v22i8e17158_app1.docx]

Multimedia Appendix 2
Types of user input (blue) and output (green) in the conversational agents.
[DOCX File , 32 KB - jmir_v22i8e17158_app2.docx]

Multimedia Appendix 3
Characteristics of conversational agents reported in the included studies.
[DOCX File , 44 KB - jmir_v22i8e17158_app3.docx]

Multimedia Appendix 4
Characteristics of included studies.
[DOCX File , 45 KB - jmir_v22i8e17158_app4.docx]

Multimedia Appendix 5
List of excluded studies and reasons for exclusion.
[DOCX File , 30 KB - jmir_v22i8e17158_app5.docx]

References

1. Chatbot. Oxford Living Dictionaries. 1990. URL: https://en.oxforddictionaries.com/definition/chatbot [accessed 2020-07-18]
2. Veretskaya O. What is a Chatbot and How to Use It for Your Business. Medium. 2017. URL: https://medium.com/swlh/what-is-a-chatbot-and-how-to-use-it-for-your-business-976ec2e0a99f [accessed 2020-07-18]
3. Smart Speaker Sales More Than Tripled in 2017. Billboard. 2017. URL: https://www.billboard.com/articles/business/8085524/smart-speaker-sales-tripled-25-million-year-2017 [accessed 2020-07-18]
4. Perez S. 39 Million Americans Now Own a Smart Speaker, Report Claims. Tech Crunch. 2018. URL: https://techcrunch.com/2018/01/12/39-million-americans-now-own-a-smart-speaker-report-claims/ [accessed 2020-07-18]
5. Smart Speakers 2018: World Market Forecast to 2023 - Key Vendors Covered include Alphabet, Amazon, Harman Intl, Alibaba and Sonos. Cision PR Newswire. 2018. URL: https://www.prnewswire.com/news-releases/smart-speakers-2018-world-market-forecast-to-2023—key-vendors-covered-include-alphabet-amazon-harman-intl-alibaba-and-sonos-300676460.html [accessed 2020-07-18]
6. Saeed H. Developing a Chatbot? Learn the Difference between AI, Machine Learning, and NLP. Chatbots Life. 2016. URL: https://chatbotslife.com/developing-a-chatbot-learn-the-difference-between-ai-machine-learning-and-nlp-40a3f745aec4 [accessed 2020-07-18]
7. Seeman P. Natural language generation: an overview. J Comput Sci Res 2012;1(3):50-57 [FREE Full text] [doi: 10.1093/oxfordhb/9780199276349.013.0015]
8. What is a Chatbot? All You Need to Know About Chatbots!. Botpress: Open-Source Conversational AI Platform. 2018. URL: https://botpress.io/learn/what-and-why/ [accessed 2020-07-18] [WebCite Cache ID 746pXZ3d7]
9. Understanding the 'Black Box' of Artificial Intelligence. Reddit. 2018. URL: https://www.reddit.com/r/artificial/comments/9iapbc/understanding_the_black_box_of_artificial/ [accessed 2020-07-18] [WebCite Cache ID 746puKdDT]

10. Lewis C, Monnet D. AI and Machine Learning Black Boxes: The Need for Transparency and Accountability. KDnuggets News. 2017. URL: https://www.kdnuggets.com/2017/04/ai-machine-learning-black-boxes-transparency-accountability.html [accessed 2020-07-18]
11. Holm EA. In defense of the black box. Science 2019 Apr 5;364(6435):26-27. [doi: 10.1126/science.aax0162] [Medline: 30948538]
12. Weizenbaum J. ELIZA---a computer program for the study of natural language communication between man and machine. Commun ACM 2012;9(1):36-45. [doi: 10.1145/365153.365168]
13. Deryugina OV. Chatterbots. Sci Tech Inf Proc 2010 Sep 5;37(2):143-147. [doi: 10.3103/s0147688210020097]
14. The History of Chatbots. Futurism: Science and Technology News and Videos. 2016. URL: https://futurism.com/images/the-history-of-chatbots-infographic/ [accessed 2020-07-21]
15. Colby K. Artificial Paranoia: A Computer Simulation of Paranoid Processes. New York, USA: Elsevier; 2013.
16. McTear M, Callejas Z, Griol D. The Conversational Interface: Talking to Smart Devices. Switzerland: Springer International Publishing; 2016.
17. Rose Medical Center. Letting fingers do the talking. Computer makes patient satisfaction surveys a snap. Rose Medical Center, Denver, CO. Profiles Healthc Mark 1992(48):40-44. [Medline: 10120010]
18. Delichatsios HK, Friedman RH, Glanz K, Tennstedt S, Smigelski C, Pinto BM, et al. Randomized trial of a 'talking computer' to improve adults' eating habits. Am J Health Promot 2001;15(4):215-224. [doi: 10.4278/0890-1171-15.4.215] [Medline: 11349340]
19. Friedman E. Your friendly neighborhood diagnosis-aiding talking computer. Hospitals 1981 Oct 16;55(20):105-6, 108, 113. [Medline: 7275074]
20. Friedman RB, Newsom RS, Entine SM, Cheung S, Schultz JV. A simulated patient-physician encounter using a talking computer. J Am Med Assoc 1977 Oct 31;238(18):1927-1929. [Medline: 578551]
21. Migneault JP, Farzanfar R, Wright JA, Friedman RH. How to write health dialog for a talking computer. J Biomed Inform 2006 Oct;39(5):468-481 [FREE Full text] [doi: 10.1016/j.jbi.2006.02.009] [Medline: 16564749]
22. Mayo J. 2016: The Year When Chatbots Were Hot. Chatbots Life. 2016. URL: https://chatbotslife.com/2016-the-year-when-chatbots-were-hot-3d61046527f9 [accessed 2020-07-18]
23. Bruner J. Why 2016 is Shaping Up to Be the Year of the Bot. O'Reilly Media Inc. 2016. URL: https://www.oreilly.com/ideas/why-2016-is-shaping-up-to-be-the-year-of-the-bot [accessed 2020-07-18]
24. Goebel T. 2016 – The Year of the Chatbot. Aspect Blog. 2016. URL: https://blogs.aspect.com/2016-the-year-of-the-chatbot/ [accessed 2020-07-18]
25. Stormon A. The Uncomfortable Truth About Bots: 3 Reasons Why They're Failing Miserably. Chatbots Magazine. 2017. URL: https://chatbotsmagazine.com/3-uncomfortable-reasons-why-chatbots-are-failing-8913901a29e5 [accessed 2020-07-18]
26. Prize L. Mitsuku Wins 2019 Loebner Prize and Best Overall Chatbot at AISB X. AISB – The Society for the Study of Artificial Intelligence and Simulation of Behaviour. 2019. URL: https://aisb.org.uk/new_site/?p=350 [accessed 2020-07-18]
27. Neville R, Greene A, McLeod J, Tracey A, Tracy A, Surie J. Mobile phone text messaging can help young people manage asthma. Br Med J 2002 Sep 14;325(7364):600 [FREE Full text] [doi: 10.1136/bmj.325.7364.600/a] [Medline: 12228151]
28. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
29. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
30. Kowatsch T, Volland D, Shih I, Rüegger D, Künzler F, Barata F. Design and Evaluation of a Mobile Chat App for the Open Source Behavioral Health Intervention Platform MobileCoach. In: Chatbots International Conference on Design Science Research in Information System and Technology. 2017 Presented at: DESRIST 2017; May 30-June 1, 2017; Karlsruhe, Germany. [doi: 10.1007/978-3-319-59144-5_36]
31. Tropea P, Schlieter H, Sterpi I, Judica E, Gand K, Caprino M, et al. Rehabilitation, the great absentee of virtual coaching in medical care: scoping review. J Med Internet Res 2019 Oct 1;21(10):e12805 [FREE Full text] [doi: 10.2196/12805] [Medline: 31573902]
32. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
33. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 7;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
34. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 May 9;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
35. Pereira J, Díaz O. Using health chatbots for behavior change: a mapping study. J Med Syst 2019 Apr 4;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]

36. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017 Jul 21;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]
37. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
38. Montenegro J, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Appl 2019 Sep 7;129(11):56-67. [doi: 10.1016/j.eswa.2019.03.054]
39. Xing Z, Yu F, Du J, Walker JS, Paulson CB, Mani NS, et al. Conversational interfaces for health: bibliometric analysis of grants, publications, and patents. J Med Internet Res 2019 Nov 18;21(11):e14672 [FREE Full text] [doi: 10.2196/14672] [Medline: 31738171]
40. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
41. Embodied Conversational Agent. Chatbots. URL: https://www.chatbots.org/embodied_conversational_agent/ [accessed 2020-07-18]
42. Bickmore T. Relational Agents. Northeastern University. URL: http://www.ccs.neu.edu/home/bickmore/agents/ [accessed 2020-07-18]
43. de Haan H, Snijder J, van Nimwegen C, Beun R. Chatbot Personality and Customer Satisfaction. Info Support Research. 2018. URL: https://research.infosupport.com/wp-content/uploads/Chatbot-Personality-and-Customer-Satisfaction-Bachelor-Thesis-Information-Sciences-Hayco-de-Haan.pdf [accessed 2020-07-18]
44. Ballati F, Corno F, de Russis L. 'Hey Siri, do you understand me?': Virtual Assistants and Dysarthria. In: 7th International Workshop on the Reliability of Intelligent Environments. 2018 Presented at: WoRIE'18; June 25-28, 2018; Rome, Italy URL: https://www.researchgate.net/publication/325466714_Hey_Siri_do_you_understand_me_Virtual_Assistants_and_Dysarthria
45. Vita S, Marocco R, Pozzetto I, Morlino G, Vigilante E, Palmacci V. The 'doctor apollo' chatbot: a digital health tool to improve engagement of people living with HIV. J Int AIDS Soc 2018 Oct;21(Suppl 8):e25187 [FREE Full text] [doi: 10.1002/jia2.25187] [Medline: 30362663]
46. Kowatsch T, Nißen M, Shih C, Rüegger D, Volland D, Filler A. Text-Based Healthcare Chatbots Supporting Patient and Health Professional Teams: Preliminary Results of a Randomized Controlled Trial on Childhood Obesity. In: 17th International Conference on Intelligent Virtual Agents. 2017 Presented at: IVA'17; August 27-30, 2017; Stockholm, Sweden URL: https://tinyurl.com/yxpmarws
47. Stasinaki A, Brogle B, Buchter D, Shih CHI, Heldt K, White C, et al. A novel digital health intervention improves physical performance in obese youth. Swiss Medical Weekly 2018 May;148:10s-10s [FREE Full text]
48. Denecke K, Hochreutener SL, Pöpel A, May R. Self-anamnesis with a conversational user interface: concept and usability study. Methods Inf Med 2018 Nov;57(5-06):243-252. [doi: 10.1055/s-0038-1675822] [Medline: 30875703]
49. Heldt K, Buchter D, Brogle B, Shih C, Ruegger D, Filler A. Telemedicine therapy for overweight adolescents: first results of a novel smartphone app intervention using a behavioural health platform. Obesity Facts 2018 May 11(Suppl 1):214-215 [FREE Full text]
50. L'Allemand D, Shih C, Heldt K, Buchter D, Brogle B, Ruegger D. Design and interim evaluation of a smartphone app for overweight adolescents using a behavioural health intervention platform. Obesity Reviews 2018 Dec 19(Suppl 1):102.
51. Can a smartphone app that includes a chatbot-based coaching and incentives increase physical activity in healthy adults? Clinical Trials. 2017. URL: https://clinicaltrials.gov/ct2/show/NCT03384550 [accessed 2020-07-18]
52. Casas J, Mugellini E, Abou Khaled O. Food Diary Coaching Chatbot. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers. 2018 Presented at: UbiComp'18; October 8-12, 2018; Singapore p. 1676-1680. [doi: 10.1145/3267305.3274191]
53. Chaix B, Bibault JE, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. JMIR Cancer 2019 May 2;5(1):e12856 [FREE Full text] [doi: 10.2196/12856] [Medline: 31045505]
54. Artificial Intelligence vs Physicians for Breast Cancer Patients' Information. Clinical Trials. 2018. URL: https://clinicaltrials.gov/ct2/show/NCT03556813 [accessed 2018-07-18]
55. Lobo J, Ferreira L, Ferreira A. CARMIE: a conversational medication assistant for heart failure. Int J E-Health Med Commun 2017;8(4):21-37. [doi: 10.4018/ijehmc.2017100102]
56. Crutzen R, Peters GJ, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health 2011 May;48(5):514-519. [doi: 10.1016/j.jadohealth.2010.09.002] [Medline: 21501812]
57. Gaffney H, Mansell W, Edwards R, Wright J. Manage Your Life Online (MYLO): a pilot trial of a conversational computer-based intervention for problem solving in a student sample. Behav Cogn Psychother 2014 Nov;42(6):731-746. [doi: 10.1017/S135246581300060X] [Medline: 23899405]

58. Middleton K, Butt M, Hammerla N, Hamblin S, Mehta K, Parsa A. Sorting out symptoms: design and evaluation of the 'babylon check' automated triage system. Cornell University. URL: https://arxiv.org/abs/1606.02041 [accessed 2020-07-22]
59. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
60. Razzaki S, Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, et al. A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. Cornell University. URL: https://arxiv.org/abs/1806.10698 [accessed 2020-07-22]
61. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
62. Griol D, Callejas Z. Mobile conversational agents for context-aware care applications. Cogn Comput 2015 Aug 21;8(2):336-356. [doi: 10.1007/s12559-015-9352-x]
63. Griol D, Molina JM, Callejas Z. Towards Emotionally Sensitive Conversational Interfaces for E-therapy. In: Artificial Computation in Biology and Medicine: International Work-Conference on the Interplay Between Natural andArtificial Computation, IWINAC 2015, Elche, Spain, June 1-5, 2015, Proceedings, Part I. Berlin: Springer; 2015.
64. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
65. Comendador BE, Francisco BB, Medenilla JS, Mae S. Pharmabot: A Pediatric Generic Medicine Consultant Chatbot. Journal of Automation and Control Engineering. 2015. URL: http://www.joace.org/index.php?m=content&c=index&a=show&catid=42&id=218 [accessed 2020-07-22]
66. Huang J, Li Q, Xue Y, Cheng T, Xu S, Jia J, et al. TeenChat: A Chatterbot System for Sensing and Releasing Adolescents' Stress. In: International Conference on Health Information Science. Heidelberg: Springer; May 2015:133-145.
67. Kamita T, Ito T, Matsumoto A, Munakata T, Inoue T. A Chatbot System for Mental Healthcare Based on SAT Counseling Method. Mobile Information Systems 2019 Mar 03;2019(2):1-11. [doi: 10.1155/2019/9517321]
68. Kobori Y, Osaka A, Soh S, Okada H. MP15-03 Novel application for sexual transmitted infection screening with an AI chatbot. J Urol 2018 Apr 03;199(4S):1-11. [doi: 10.1016/j.juro.2018.02.516]
69. Mujeeb S, Hafeez M, Arshad T. Aquabot: a diagnostic chatbot for achluophobia and autism. Int J Adv Comput Sci Appl 2017 Dec;8(9):39-46. [doi: 10.14569/IJACSA.2017.080930]
70. Sosale A, Shaikh M, Shah A, Chawla R, Makkar B, Kesavadev J, et al. Real-world effectiveness of a digital therapeutic in improving glycaemic control in south asians living with type 2 diabetes. Diabetes 2018 May;67(Supplement 1):866-P-86640. [doi: 10.2337/db18-866-P]
71. Danda P, Srivastava BML, Shrivastava M. Vaidya: A Spoken Dialog System for Health Domain. In: Proceedings of the 13th International Conference on Natural Language Processing. 2015 Presented at: International Conference on Natural Language Processing; 2016; Varanasi, India.
72. Wang H, Zhang Q, Ip M, Fai Lau J. Social media–based conversational agents for health management and interventions. Computer 2018 Aug 03;51(8):26-33. [doi: 10.1109/MC.2018.3191249]
73. Elmasri D, Maeder A. A Conversational Agent for an Online Mental Health Intervention. In: Brain Informatics and Health. Cham: Springer; Sep 23, 2016.
74. Ghosh S, Bhatia S, Bhatia A. Quro: facilitating user symptom check using a personalised chatbot-oriented dialogue system. Stud Health Technol Inform 2018;252:51-56. [Medline: 30040682]
75. Joerin A, Rauws M, Ackerman ML. Psychological artificial intelligence service, Tess: delivering on-demand support to patients and their caregivers: technical report. Cureus 2019 Jan 28;11(1):e3972 [FREE Full text] [doi: 10.7759/cureus.3972] [Medline: 30956924]
76. Wilson N, MacDonald EJ, Mansoor OD, Morgan J. In bed with Siri and Google Assistant: a comparison of sexual health advice. BMJ 2017 Dec 13;359:j5635. [doi: 10.1136/bmj.j5635] [Medline: 29237603]
77. Ni L, Lu C, Liu N, Liu J. MANDY: towards a smart primary care chatbot application. In: Knowledge and Systems Sciences. Heidelberg: Springer; 2017.
78. van Heerden A, Ntinga X, Vilakazi K. The potential of conversational agents to provide a rapid HIV counseling and testing services. In: International Conference on the Frontiers and Advances in Data Science (FADS). 2016 Presented at: 2017 International Conference on the Frontiers and Advances in Data Science (FADS); 2017; Xi'an, China. [doi: 10.1109/FADS.2017.8253198]
79. Cheng A, Raghavaraju V, Kanugo J, Handrianto YP. Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes. 2018 Presented at: 15th IEEE Annual Consumer Communications & Networking Conference; Jan, 12-15; Las Vegas, NV, USA. [doi: 10.1109/CCNC.2018.8319283]
80. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]

81. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
82. Galescu L, Allen J, Ferguson G, Quinn J, Swift M. Speech recognition in a dialog system for patient health monitoring. 2009 Presented at: 2009 IEEE International Conference on Bioinformatics and Biomedicine Workshops; November 1-4, 2009; Washington, DC, USA. [doi: 10.1109/BIBMW.2009.5332111]
83. Liu B, Sundar SS. Should machines express sympathy and empathy? Experiments with a health advice chatbot. Cyberpsychol Behav Soc Netw 2018 Oct;21(10):625-636. [doi: 10.1089/cyber.2018.0110] [Medline: 30334655]
84. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med 2016 May 01;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
85. Rhee H, Allen J, Mammen J, Swift M. Mobile phone-based asthma self-management aid for adolescents (mASMAA): a feasibility study. Patient Prefer Adherence 2014;8:63-72 [FREE Full text] [doi: 10.2147/PPA.S53504] [Medline: 24470755]
86. Stein N, Brooks K. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. JMIR Diabetes 2017 Nov 01;2(2):e28 [FREE Full text] [doi: 10.2196/diabetes.8590] [Medline: 30291087]
87. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and google assistant. J Med Internet Res 2018 Sep 04;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
88. Ghandeharioun A, McDuff D, Czerwinski M, Rowan K. EMMA: An Emotionally Intelligent Personal Assistant for Improving Wellbeing. Researchgate. URL: https://www.researchgate.net/publication/330035444_EMMA_An_Emotionally_Intelligent_Personal_Assistant_for_Improving_Wellbeing [accessed 2020-07-22]
89. Kocielnik R, Xiao L, Avrahami D, Hsieh G. Reflection Companion: a conversational system for engaging users in reflection on physical activity. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Jul 05;2(2):1-26. [doi: 10.1145/3214273]
90. Kowatsch T, Nißen M, Rüegger D, Stieger M, Flückiger C, Allemand M. The impact of interpersonal closeness cues in text-based healthcare chatbots on attachment bond and the desire to continue interacting: an experimental design. University of Zurich. URL: https://www.zora.uzh.ch/id/eprint/158352/1/Kowatsch%2520et%2520al%25202018%2520InterPersCloseness-of-THCB.pdf [accessed 2020-07-22]
91. ReDial: Recommendation dialogs for bridging the gap between chit-chat and goal-oriented chatbots. Microsoft. URL: https://www.microsoft.com/en-us/research/blog/redial-recommendation-dialogs-for-bridging-the-gap-between-chit-chat-and-goal-oriented-chatbots/ [accessed 2020-07-22]
92. Onwumere J, Amaral F, Valmaggia LR. Digital technology for caregivers of people with psychosis: systematic review. JMIR Ment Health 2018 Sep 05;5(3):e55 [FREE Full text] [doi: 10.2196/mental.9857] [Medline: 30185402]
93. Ruggiano N, Brown EL, Li J, Scaccianoce M. Rural Dementia Caregivers and Technology: What Is the Evidence? Res Gerontol Nurs 2018 Jul 01;11(4):216-224. [doi: 10.3928/19404921-20180628-04] [Medline: 30036405]
94. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]
95. Coiera E. Paper Review: the Babylon Chatbot. Wordpress. URL: https://coiera.com/2018/06/29/paper-review-the-babylon-chatbot/ [accessed 2020-07-22]
96. Thompson D, Baranowski T. Chatbots as extenders of pediatric obesity intervention: an invited commentary on 'Feasibility of Pediatric Obesity & Pre-Diabetes Treatment Support through Tess, the AI Behavioral Coaching Chatbot'. Transl Behav Med 2019 May 16;9(3):448-450. [doi: 10.1093/tbm/ibz065] [Medline: 31094432]
97. D'Alfonso S, Santesteban-Echarri O, Rice S, Wadley G, Lederman R, Miles C, et al. Artificial intelligence-assisted online social therapy for youth mental health. Front Psychol 2017;8:796 [FREE Full text] [doi: 10.3389/fpsyg.2017.00796] [Medline: 28626431]
98. Spinazze P, Bottle A, Car J. Digital health sensing for personalized dermatology. Sensors (Basel) 2019 Aug 5;19(15):3426. [doi: 10.3390/s19153426] [Medline: 31387237]
99. Pal K, Dack C, Ross J, Michie S, May C, Stevenson F, et al. Digital health interventions for adults with type 2 diabetes: qualitative study of patient perspectives on diabetes self-management education and support. J Med Internet Res 2018 Jan 29;20(2):e40 [FREE Full text] [doi: 10.2196/jmir.8439]
100. Bhattarai P, Phillips JL. The role of digital health technologies in management of pain in older people: an integrative review. Arch Gerontol Geriatr 2017;68:14-24. [doi: 10.1016/j.archger.2016.08.008]
101. Devine KA, Viola AS, Coups EJ, Wu YP. Digital health interventions for adolescent and young adult cancer survivors. JCO Clin Cancer Inform 2018;2:1-15. [doi: 10.1200/CCI.17.00138]
102. Amrita DB. Health care social media: expectations of users in a developing country. Med 2 0 2013;2(2):e4 [FREE Full text] [doi: 10.2196/med20.2720] [Medline: 25075239]
103. Kerr D, Axelrod C, Hoppe C, Klonoff DC. Diabetes and technology in 2030: a utopian or dystopian future? Diabet Med 2018 Apr;35(4):498-503. [doi: 10.1111/dme.13586] [Medline: 29356078]

104. Mascitti I, Feituri M, Funghi F, Correnti S. COACH BOT: Modular e-course with virtual coach tool support. In: Proceedings of the International Conference on Agents and Artificial Intelligence. 2010 Jul 01 Presented at: International Conference on Agents and Artificial Intelligence; 2010; January, 22-24. [doi: 10.5220/0002589901150120]
105. Abashev A, Grigoryev R, Grigorian K, Boyko V. Programming tools for messenger-based chatbot system organization: implication for outpatient and translational medicines. BioNanoSci 2016 Nov 22;7(2):403-407. [doi: 10.1007/S12668-016-0376-9]
106. Ahmad NS, Sanusi MH, Abd Wahab MH, Mustapha A, Sayadi ZA, Saringat MZ. Conversational bot for pharmacy: a natural language approach. 2018 Presented at: 2018 IEEE Conference on Open Systems (ICOS); 21-22 Nov, 2018; Langkawi Island, Malaysia, Malaysia. [doi: 10.1109/ICOS.2018.8632700]
107. Alexander JA. Computer assisted optometry--a tutorial with examples. Am J Optom Arch Am Acad Optom 1973 Sep;50(9):730-736. [doi: 10.1097/00006324-197309000-00007] [Medline: 4584475]
108. Chatbots Meet eHealth: Automatizing Healthcare. University of Naples. URL: http://ceur-ws.org/Vol-1982/paper6.pdf [accessed 2020-07-22]
109. Atay C, Ireland D, Liddle J, Wiles J, Vogel A, Angus D, et al. P3-404: can a smartphone-based chatbot engage older community group members? The impact of specialised content. Alzheimers Demen 2016 Jul 01;12:P1005-P1006. [doi: 10.1016/j.jalz.2016.06.2070]
110. Improving adherence in automated e-coaching. In: Persuasive Strategies To Improve Driving Behaviour Of Elderly Drivers By A Feedback Approach. Cham: Springer; 2016.
111. Brixey J, Hoegen R, Lan W, Rusow J, Singla K, Yin X. Shihbot: A facebook chatbot for sexual health information on hiv/aids. In: Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue. 2017 Presented at: 18th Annual SIGdial Meeting on Discourse and Dialogue Cite this publication; 2017; Saarbrücken, Germany. [doi: 10.18653/v1/W17-5544]
112. Callejas Z, Griol D, McTear F, López-Cózar R. A virtual coach for active ageing based on sentient computing and m-health. Ambient Assisted Living and Daily Activities 2014:-. [doi: 10.1007/978-3-319-13105-4_10]
113. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S. Towards a chatbot for digital counselling. In: Proceedings of the 31st British Computer Society Human Computer Interaction Conference. 2017 Presented at: 31st British Computer Society Human Computer Interaction Conference; 3-6 July, 2017; University of Sunderland, UK. [doi: 10.14236/ewic/HCI2017.24]
114. Cameron G, Cameron D, Megaw G, Bond R. Best Practices for Designing Chatbots in Mental Healthcare–A Case Study on iHelpr. In: Proceedings of the 32nd International BCS Human Computer Interaction Conference. 2018 Presented at: British HCI Conference 2018; 2018; Belfast. [doi: 10.14236/ewic/HCI2018.129]
115. Chung K, Park RC. Chatbot-based heathcare service with a knowledge base for cloud computing. Cluster Comput 2018 Mar 16;22(S1):1925-1937. [doi: 10.1007/s10586-018-2334-5]
116. Cooper A, Ireland D. Designing a chat-bot for non-verbal children on the autism spectrum. Stud Health Technol Inform 2018;252:63-68. [Medline: 30040684]
117. Denecke K, Lutz Hochreutener S, Poepel A, May R. Talking to Ana: A Mobile Self-Anamnesis Application with Conversational User Interface. Cluster Comput 2019 Jan;22(1):-. [doi: 10.1145/3194658.3194670]
118. Dharwadkar R, Deshpande NA. A Medical ChatBot. Int J Comp Trends Technol 2018;60(1):- [FREE Full text]
119. Divya S, Indumathi V, Ishwarya S, Priyasankari M, Devi SK. A self-diagnosis medical chatbot using artificial intelligence. Journal of Web Development and Web Designing 2018;3(1):-.
120. Do HI, Fu WT. Empathic Virual Assistant for Healthcare Information with Positive Emotional Experience. 2016 Presented at: 2016 IEEE International Conference on Healthcare Informatics; 2016; United States.
121. Dubosson F, Schaer R, Savioz R, Schumacher M. Going beyond the relapse peak on social network smoking cessation programmes: ChatBot opportunities. Swiss Med Informatics 2017 Sep 20:-. [doi: 10.4414/smi.33.00397]
122. Fadhil A, Gabrielli S. Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: International Conference on Pervasive Computing Technologies for Healthcare; May 23-26, 2017; Barcelona, Spain. [doi: 10.1145/3154862.3154914]
123. Fadhil A, Villafiorita A. An adaptive learning with gamification & conversational UIs: The rise of CiboPoliBot. 2017 Presented at: 25th Conference on User Modeling, Adaptation and Personalization; 9-12th July, 2017; Bratislava, Slovakia. [doi: 10.1145/3099023.3099112]
124. Fadhil A, Diaconu M, Gabrielli S, Villafiorita A. CoachAI: A conversational UI assisted e-coaching platform. Cornell Univeristy. 2017. URL: https://arxiv.org/abs/1904.11961 [accessed 2020-07-22]
125. Beyond Patient Monitoring: Conversational Agents Role in Telemedicine & Healthcare Support For Home-Living Elderly Individuals. Cornell University. URL: https://arxiv.org/abs/1803.06000 [accessed 2020-07-22]
126. Can a Chatbot Determine My Diet?: Addressing Challenges of Chatbot Application for Meal Recommendation. Cornell Univeristy. URL: https://arxiv.org/abs/1802.09100 [accessed 2020-07-22]
127. A Conversational Interface to Improve Medication Adherence: Towards AI Support in Patient's Treatment. Cornell University. URL: https://arxiv.org/abs/1803.09844 [accessed 2020-07-22]

128. Fadhil A, Wang Y, Reiterer H. Assistive conversational agent for health coaching: a validation study. Methods Inf Med 2019 Jun;58(1):9-23. [doi: 10.1055/s-0039-1688757] [Medline: 31117129]
129. CARDIAC: An intelligent conversational assistant for chronic heart failure patient heath monitoring. Researchgate. URL: https://www.researchgate.net/publication/228374471_CARDIAC_An_intelligent_conversational_assistant_for_chronic_heart_failure_patient_heath_monitoring [accessed 2020-07-22]
130. Ferguson G, Quinn J, Horwitz C, Swift M, Allen J, Galescu L. Towards a personal health management assistant. J Biomed Inform 2010 Oct;43(5 Suppl):S13-S16 [FREE Full text] [doi: 10.1016/j.jbi.2010.05.014] [Medline: 20937478]
131. Implementation and feasibility study of a tailored health education bot in Telegram for mothers of children with obesity and overweight. Researchgate. URL: https://tinyurl.com/y49e6xsj [accessed 2020-07-22]
132. From Books to Bots: Using Medical Literature to Create a Chat Bot. Researchgate. URL: https://www.researchgate.net/publication/304358640_From_Books_to_Bots_Using_Medical_Literature_to_Create_a_Chat_Bot [accessed 2020-07-22]
133. SLOWBot (chatbot) Lifestyle Assistant. ACM Digital Library. URL: https://dl.acm.org/doi/10.1145/3240925.3240953 [accessed 2020-07-22]
134. Conversational System to Assist the User when Accessing Web Sources in the Medical Domain. ResearchGate. URL: https://www.researchgate.net/publication/261031810_Conversational_System_to_Assist_the_User_when_Accessing_Web_Sources_in_the_Medical_Domain [accessed 2020-07-22]
135. Hassoon A, Schrack J, Naiman D, Lansey D, Baig Y, Stearns V, et al. Increasing physical activity amongst overweight and obese cancer survivors using an alexa-based intelligent agent for patient coaching: protocol for the physical activity by technology help (PATH) trial. JMIR Res Protoc 2018 Feb 12;7(2):e27 [FREE Full text] [doi: 10.2196/resprot.9096] [Medline: 29434016]
136. Allergybot: A chatbot technology intervention for young adults with food allergies dining out. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2017. URL: https://www.semanticscholar.org/paper/A-Conversational-System-to-Assist-the-User-when-Web-Gatius-Namsrai/b6e0f6874204f77fea2b57005470f17e1d5f8d3f [accessed 2020-07-22]
137. Cooper A, Ireland D. Designing a chat-bot for non-verbal children on the autism spectrum. Stud Health Technol Inform 2018;252:63-68. [Medline: 30040684]
138. Chat-Bots for People with Parkinson's Disease: Science Fiction or Reality? Studies in health technology and informatics. URL: https://www.researchgate.net/publication/280726980_Chat-Bots_for_People_with_Parkinson's_Disease_Science_Fiction_or_Reality [accessed 2020-07-22]
139. Hello Harlie: Enabling Speech Monitoring Through Chat-Bot Conversations. Studies in health technology and informatics. URL: https://www.researchgate.net/publication/323104602_Hello_Harlie_Enabling_Speech_Monitoring_Through_Chat-Bot_Conversations [accessed 2020-07-22]
140. Kanagarajan K, Saradha A. An intelligent conversation agent for health care domain. IJSC 2014 Apr 01;4(3):772-776. [doi: 10.21917/ijsc.2014.0110]
141. Kramer JN, Künzler F, Mishra V, Presset B, Kotz D, Smith S, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540 [FREE Full text] [doi: 10.2196/11540] [Medline: 30702430]
142. Oh KJ, Lee DK, Ko BS, Hyeon J, Choi HJ. Empathy bot: conversational service for psychiatric counseling with chat assistant. Stud Health Technol Inform 2017;245:1235. [Medline: 29295322]
143. Lee D, Oh KJ, Choi HJ. The chatbot feels you - a counseling service using emotional response generation. 2017 Presented at: IEEE International Conference on Big Data and Smart Computing (BigComp); February 13-16, 2017; Jeju Island, Korea. [doi: 10.1109/BIGCOMP.2017.7881752]
144. Lokman AS, Zain JM. One-match and all-match categories for keywords matching in chatbot. Am J Appl Sci 2010 Oct 01;7(10):1406-1411. [doi: 10.3844/ajassp.2010.1406.1411]
145. An architectural design of Virtual Dietitian (ViDi) for diabetic patients. IEEE Explore. 2009. URL: https://ieeexplore.ieee.org/document/5234671?reload=true&arnumber=5234671 [accessed 2020-07-22]
146. Designing a Chatbot for diabetic patients. International Conference on Software Engineering & Computer Systems. URL: https://www.researchgate.net/publication/266872926_Designing_a_Chatbot_for_Diabetic_Patients [accessed 2020-07-22]
147. A novel approach for medical assistance using trained chatbot. ResearchGate. URL: https://www.researchgate.net/publication/318474956_A_novel_approach_for_medical_assistance_using_trained_chatbot [accessed 2020-07-22]
148. Marciel KK, Saiman L, Quittell LM, Dawkins K, Quittner AL. Cell phone intervention to improve adherence: cystic fibrosis care team, patient, and parent perspectives. Pediatr Pulmonol 2010 Feb;45(2):157-164 [FREE Full text] [doi: 10.1002/ppul.21164] [Medline: 20054860]
149. The use of a chatbot in radiology education. European Society of Radiology. URL: https://epos.myesr.org/poster/esr/ranzcr2018/R-0095 [accessed 2020-07-22]

150. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J Med Internet Res 2018 Jun 26;20(6):e10148 [FREE Full text] [doi: 10.2196/10148] [Medline: 29945856]
151. A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation. IEEExplore. URL: https://ieeexplore.ieee.org/document/7962482 [accessed 2020-07-22]
152. Likita: A Medical Chatbot To Improve HealthCare Delivery In Africa. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. URL: https://www.researchgate.net/publication/330522151_Likita_A_Medical_Chatbot_To_Improve_HealthCare_Delivery_In_Africa [accessed 2020-07-22]
153. Chatbot Dimensions that Matter: Lessons from the Trenches. Web Engineering. URL: https://link.springer.com/chapter/10.1007/978-3-319-91662-0_9 [accessed 2020-07-22]
154. Automated Medical Chatbot. SSRN. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3090881 [accessed 2020-07-22]
155. Proposal for the development of a mobile virtual assistant for treatment of tuberculosis (2018). Repositório da Produção USP. URL: https://repositorio.usp.br/item/002939130 [accessed 2020-07-22]
156. SleepBot: encouraging sleep hygiene using an intelligent chatbot. ResearchGate. URL: https://www.researchgate.net/publication/331428159_SleepBot_encouraging_sleep_hygiene_using_an_intelligent_chatbot [accessed 2020-07-22]
157. HomeNL: Homecare Assistance in Natural Language. An Intelligent Conversational Agent for Hypertensive Patients Management. Centre pour la Communication Scientifique Directe. URL: https://hal.archives-ouvertes.fr/inria-00519752/ [accessed 2020-07-22]
158. Chatbot Utilization for Medical Consultant System. IEEEXplore. URL: https://ieeexplore.ieee.org/document/8621678 [accessed 2020-07-22]
159. Sanative Chatbot For Health Seekers. International Journal of Engineering and Computer Science. URL: http://www.ijecs.in/index.php/ijecs/article/view/720 [accessed 2020-07-22]
160. MamaBot: a System based on ML and NLP for supporting Women and Families during Pregnancy. Semantic Scholar. URL: https://www.semanticscholar.org/paper/MamaBot%3A-a-System-based-on-ML-and-NLP-for-Women-and-Vaira-Bochicchio/c6ac5cdb449e1e08bc0321a675e88d95b1dc88b6 [accessed 2020-07-22]
161. Chatbots and Conversational Interfaces: Three Domains of Use. CEUR Workshop Proceedings. URL: http://ceur-ws.org/Vol-2101/paper8.pdf [accessed 2020-07-22]
162. Blanson Henkemans OA, van der Boog PJ, Lindenberg J, van der Mast CA, Neerincx MA, Zwetsloot-Schonk BJ. An online lifestyle diary with a persuasive computer assistant providing feedback on self-management. Technol Health Care 2009;17(3):253-267. [doi: 10.3233/THC-2009-0545] [Medline: 19641261]
163. Towards Fully Automated Psychotherapy for Adults - BAS - Behavioral Activation Scheduling Via Web and Mobile Phone. Semantic Scholar. URL: https://www.semanticscholar.org/paper/Towards-Fully-Automated-Psychotherapy-for-Adults-Griffioen-Both-Cuijpers/43b3c78317a273c7ec46c1696370bf68e42b30fb [accessed 2020-07-22]
164. Allen J, Ferguson G, Blaylock N, Byron D, Chambers N, Dzikovska M, et al. Chester: towards a personal medication advisor. J Biomed Inform 2006 Oct;39(5):500-513 [FREE Full text] [doi: 10.1016/j.jbi.2006.02.004] [Medline: 16545620]
165. Bickmore TW, Schulman D, Sidner C. Automated interventions for multiple health behaviors using conversational agents. Patient Educ Couns 2013 Aug;92(2):142-148 [FREE Full text] [doi: 10.1016/j.pec.2013.05.011] [Medline: 23763983]
166. Evaluating Quality of Chatbots and Intelligent Conversational Agents. arXiv. URL: https://arxiv.org/ftp/arxiv/papers/1704/1704.04579.pdf [accessed 2020-07-22]
167. Conversational Agents and Mental Health: Theory-Informed Assessment of Language and Affect. Stanford Univeristy. URL: http://ilpubs.stanford.edu:8090/1141/1/healthDialog.pdf [accessed 2020-07-22]
168. Rizzo AA, Lange B, Buckwalter JG, Forbell E, Kim J, Sagae K, et al. An intelligent virtual human system for providing healthcare information and support. Stud Health Technol Inform 2011;163:503-509. [Medline: 21335847]
169. Different measurements metrics to evaluate a chatbot system. ACM Digital Library. URL: https://dl.acm.org/doi/10.5555/1556328.1556341 [accessed 2020-07-22]
170. Dr. Vdoc: A Medical Chatbot that Acts as a Virtual Doctor. Research & Reviews: Journal of Medical Science and Technology. URL: http://medicaljournals.stmjournals.in/index.php/RRJoMST/article/view/30 [accessed 2020-07-22]
171. Kazi H, Chowdhry B, Memon Z. MedChatBot: an UMLS based chatbot for medical students. IJCA 2012 Oct 20;55(17):1-5. [doi: 10.5120/8844-2886]
172. Coach Me: A Platform For Promoting Healthy Lifestyle. ResearchGate. URL: https://www.researchgate.net/publication/307573372_Coach_Me_A_Platform_For_Promoting_Healthy_Lifestyle [accessed 2020-07-22]
173. Bickmore TW, Pfeifer LM, Byron D, Forsythe S, Henault LE, Jack BW, et al. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. J Health Commun 2010;15 Suppl 2:197-210. [doi: 10.1080/10810730.2010.499991] [Medline: 20845204]

174. Lindenberg K, Moessner M, Harney J, McLaughlin O, Bauer S. E-health for individualized prevention of eating disorders. Clin Pract Epidemiol Ment Health 2011;7:74-83 [FREE Full text] [doi: 10.2174/1745017901107010074] [Medline: 21687562]
175. Dowling M, Rickwood D. Exploring hope and expectations in the youth mental health online counselling environment. Comp Hum Behav 2016 Feb;55:62-68. [doi: 10.1016/j.chb.2015.08.009]
176. Dowling M, Rickwood D. Investigating individual online synchronous chat counselling processes and treatment outcomes for young people. Adv Ment Health 2015 Jan 30;12(3):216-224. [doi: 10.1080/18374905.2014.11081899]
177. Dowling M, Rickwood D. A naturalistic study of the effects of synchronous online chat counselling on young people's psychological distress, life satisfaction and hope. Couns Psychother Res 2015 Jul 14;15(4):274-283. [doi: 10.1002/capr.12037]
178. Azevedo RF, Morrow D, Graumlich J, Willemsen-Dunlap A, Hasegawa-Johnson M, Huang TS, et al. Using conversational agents to explain medication instructions to older adults. AMIA Annu Symp Proc 2018;2018:185-194 [FREE Full text] [Medline: 30815056]
179. Brown RL, McDermott RJ, Marty PJ. A conversational information computer system for health and safety operation: the occupational surveillance interactive system (OSIS). Am Ind Hyg Assoc J 1981 Nov;42(11):824-830. [doi: 10.1080/15298668191420756] [Medline: 7315741]
180. Crutzen R, Bosma H, Havas J, Feron F. What can we learn from a failed trial: insight into non-participation in a chat-based intervention trial for adolescents with psychosocial problems. BMC Res Notes 2014 Nov 20;7:824 [FREE Full text] [doi: 10.1186/1756-0500-7-824] [Medline: 25409911]
181. Denecke K, Tschanz M, Dorner TL, May R. Intelligent conversational agents in healthcare: hype or hope? Stud Health Technol Inform 2019;259:77-84. [Medline: 30923277]
182. Designing for Health Chatbots. Cornell University. URL: https://arxiv.org/abs/1902.09022 [accessed 2020-07-22]
183. Mindbot: A Social-Based Medical Virtual Assistanta. IEEEXplore. URL: https://ieeexplore.ieee.org/document/7776377?reload=true [accessed 2020-07-22]
184. Stieger M, Nißen M, Rüegger D, Kowatsch T, Flückiger C, Allemand M. PEACH, a smartphone- and conversational agent-based coaching intervention for intentional personality change: study protocol of a randomized, wait-list controlled trial. BMC Psychol 2018 Sep 04;6(1):43 [FREE Full text] [doi: 10.1186/s40359-018-0257-9] [Medline: 30180880]
185. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. J Med Internet Res 2019 Apr 05;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]
186. Intelligent chatbot for analysis and diagnosis of the psychiatric disorders. ResearchGate. URL: https://www.researchgate.net/publication/329416982_INTELLIGENT_CHATBOT_FOR_ANALYSIS_AND_DIAGNOSIS_OF_THE_PSYCHIATRIC_DISORDERS [accessed 2020-07-22]
187. Designing Just-in-time Adaptive Interventions and Healthcare Chatbots with the Open Source Platform MobileCoach. University of St. Gallen. URL: https://www.alexandria.unisg.ch/255053/ [accessed 2020-07-22]
188. Chatbots and the new world of HCI. Interactions. URL: https://interactions.acm.org/archive/view/july-august-2017/chatbots-and-the-new-world-of-hci [accessed 2020-07-22]
189. Garcia DM, Lopez SS, Donis H. Voice activated virtual assistants personality perceptions and desires- comparing personality evaluation frameworks. 2018 Presented at: British Human Computer Interaction Conference; July 2018; Belfast. [doi: 10.14236/ewic/HCI2018.40]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
ECA: embodied conversational agent
EMBASE: Excerpta Medica database
HIC: high-income country
LMIC: low- and middle-income country
MEDLINE: Medical Literature Analysis and Retrieval System Online
NLP: natural language processing
OCLC: Online Computer Library Center
SAT: structure association technique

Edited by G Eysenbach; submitted 22.11.19; peer-reviewed by A Kocaballi, E Judica; comments to author 09.12.19; revised version received 11.04.20; accepted 13.06.20; published 07.08.20.

Please cite as:
Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, Atun R
Conversational Agents in Health Care: Scoping Review and Conceptual Analysis
J Med Internet Res 2020;22(8):e17158
URL: http://www.jmir.org/2020/8/e17158/
doi:10.2196/17158
PMID:32763886

Review

Blended Learning Compared to Traditional Learning in Medical Education: Systematic Review and Meta-Analysis

Alexandre Vallée[1], MD, PhD; Jacques Blacher[1], MD, PhD; Alain Cariou[2], MD, PhD; Emmanuel Sorbets[1], MD, PhD

[1]Diagnosis and Therapeutic Center, Hypertension and Cardiovascular Prevention Unit, Hôtel-Dieu Hospital, Assistance Publique–Hôpitaux de Paris, Paris-Descartes University, Paris, France

[2]Cochin University Hospital, Paris, France

Corresponding Author:
Alexandre Vallée, MD, PhD
Diagnosis and Therapeutic Center, Hypertension and Cardiovascular Prevention Unit
Hôtel-Dieu Hospital, Assistance Publique–Hôpitaux de Paris
Paris-Descartes University
Paris,
France
Phone: 33 142348587
Email: alexandre.g.vallee@gmail.com

Abstract

Background: Blended learning, which combines face-to-face learning and e-learning, has grown rapidly to be commonly used in education. Nevertheless, the effectiveness of this learning approach has not been completely quantitatively synthesized and evaluated using knowledge outcomes in health education.

Objective: The aim of this study was to assess the effectiveness of blended learning compared to that of traditional learning in health education.

Methods: We performed a systematic review of blended learning in health education in MEDLINE from January 1990 to July 2019. We independently selected studies, extracted data, assessed risk of bias, and compared overall blended learning versus traditional learning, offline blended learning versus traditional learning, online blended learning versus traditional learning, digital blended learning versus traditional learning, computer-aided instruction blended learning versus traditional learning, and virtual patient blended learning versus traditional learning. All pooled analyses were based on random-effect models, and the I^2 statistic was used to quantify heterogeneity across studies.

Results: A total of 56 studies (N=9943 participants) assessing several types of learning support in blended learning met our inclusion criteria; 3 studies investigated offline support, 7 studies investigated digital support, 34 studies investigated online support, 8 studies investigated computer-assisted instruction support, and 5 studies used virtual patient support for blended learning. The pooled analysis comparing all blended learning to traditional learning showed significantly better knowledge outcomes for blended learning (standardized mean difference 1.07, 95% CI 0.85 to 1.28, I^2=94.3%). Similar results were observed for online (standardized mean difference 0.73, 95% CI 0.60 to 0.86, I^2=94.9%), computer-assisted instruction (standardized mean difference 1.13, 95% CI 0.47 to 1.79, I^2=78.0%), and virtual patient (standardized mean difference 0.62, 95% CI 0.18 to 1.06, I^2=78.4%) learning support, but results for offline learning support (standardized mean difference 0.08, 95% CI –0.63 to 0.79, I^2=87.9%) and digital learning support (standardized mean difference 0.04, 95% CI –0.45 to 0.52, I^2=93.4%) were not significant.

Conclusions: From this review, blended learning demonstrated consistently better effects on knowledge outcomes when compared with traditional learning in health education. Further studies are needed to confirm these results and to explore the utility of different design variants of blended learning.

(J Med Internet Res 2020;22(8):e16504) doi:10.2196/16504

KEYWORDS

blended learning; virtual patients; online learning; computer-aided instruction; traditional learning; meta-analysis

Introduction

Background

New types of learning, such as e-learning, have become popular in medical education [1,2] since the emergence of the internet [2]. These new models allow learning to transcend boundaries of space and time; they improve collaborative and individualized learning effectiveness and are more convenient [3-5]. Nevertheless, e-learning presents some disadvantages, including high cost multimedia materials, high costs for platform maintenance, and often user training is required. In parallel, traditional learning presents several limitations, including requiring the physical presence of students and teachers at a specific time and place [6].

Blended learning is characterized by the combination of traditional face-to-face learning and asynchronous or synchronous e-learning [7]. Blended learning is a promising alternative for medical education because of its advantages over traditional learning. In academia, this learning format has had a rapid growth and is now widely used [8].

Increased research on blended learning has been reported since the 1990s [9-11]. Synthesis of these studies may inform students and teachers on the effectiveness of blended learning [12]. Previous systematic reviews have reported that blended learning has the potential to improve clinical training among medical students [13] and undergraduate nursing education [14]. In parallel, many reviews have summarized the potential of blended learning in medical education [15,16]. A meta-analysis [12] showed that blended learning was more effective than nonblended learning but with a high level of heterogeneity.

Nevertheless, these reviews were limited to only some areas of health education, and few have used quantitative synthesis in the evaluation of the effectiveness of blended learning; therefore, the purpose of this study was to quantitatively synthesize the studies that evaluated the efficacy (using knowledge outcomes) of blended learning for health education (with students, postgraduate trainees, or practitioners).

Objective

The objective of this review was to evaluate the effectiveness of blended learning for health education on knowledge outcomes assessed with subjective (eg, learner self-report) or objective evaluations (eg, multiple-choice question knowledge test) of learners' factual or conceptual understanding of the course in studies where blended learning was compared with traditional learning.

Methods

Comparison Categories and Definitions

Blended learning was compared with traditional learning, overall, and after stratification by type of learning support, the following comparisons were made: offline blended learning versus traditional learning, online blended learning versus traditional learning, digital blended learning versus traditional learning, computer-aided instruction blended learning versus traditional learning, and virtual patient blended learning versus traditional learning.

Offline learning was defined as the use of personal computers or laptops to assist in delivering stand-alone multimedia materials without the need for internet or local area network connections [17]. These could be supplemented by videoconferences, emails, and audio-visual learning materials kept in either magnetic storage (CD-ROM, floppy disk, flash memory, multimedia cards, external hard disks) as long as the learning activities did not rely on this connection [18].

Online support was defined as all online materials used in learning courses.

Digital education was a broad construct describing a wide range of teaching and learning strategies that were exclusively based on the use of electronic media and devices as training, communication, and interactions tools [19]. These aspects could pertain to educational approaches, concepts, methods, or technologies. Moreover, these concepts facilitated remote learning, which could help address the shortage of health professionals in settings with limited resources by reducing the time constraints and geographic barriers to training.

Computer-assisted instruction was defined as the use of interactive CD-ROM, multimedia software, or audio-visual material to augment instruction including multimedia presentations, live synchronous virtual sessions offered via a web-based learning platform, presentations with audio-visuals, and synchronous or asynchronous discussion forums to enhance participation and increase engagement [20,21].

Virtual patients were defined as interactive computers simulations of real-life clinical scenarios for health professional training, education, or assessment. This broad definition encompassed a variety of systems that used different technologies and addressed various learning needs [22].

Traditional learning, in this paper, was used to describe all nonblended learning such as nondigital and not online, but also only online, only e-learning, or other single support educational methods (lectures, face-to-face, reading exercises, group discussion in classroom).

Reporting Standards

We conducted and reported our study according to PRISMA guidelines [23] and Cochrane systematic review guidelines [24].

Eligibility Criteria

Inclusion criteria for studies were based on the PICOS (population, intervention, comparison, outcome, and study design) framework.

Studies were included if they were conducted among health learners, used a blended learning intervention in the experimental group, involved a comparison of blended learning with traditional learning, included quantitative outcomes with respect to knowledge assessed with either subjective or objective evaluations, and were randomized controlled trials or nonrandomized studies (which are widely used in health education). Only studies published in English were included.

Data Sources

To identify relevant studies, we conducted a search of citations published in MEDLINE between January 1990 and July 2019. Key search terms included delivery concepts (*blended, hybrid, integrated, computer-aided, computer assisted, virtual patient, learning, training, education, instruction, teaching, course*), participant characteristics (*physician, medic*, nurs*, pharmac*, dent*, health**), and study design concepts (*compar*, trial*, evaluat*, assess*, effect*, pretest*, pre-test, posttest*, post-test, preintervention, pre-intervention, postintervention, post-intervention*). Asterisks were used as a truncation symbol for searching. Multimedia Appendix 1 describes the complete research strategy.

Study Selection

Using the eligibility criteria, AV and ES independently screened all articles and abstracts and reviewed the full text of potentially eligible abstracts.

Data Extraction

AV and ES independently extracted relevant characteristics related to participants, intervention, comparators, outcome measures, and results from the studies that were found to be eligible using a standard data collection form. Any disagreements were resolved through discussion with a third research team member until agreement was reached.

Risk of Bias Assessment

During the data extraction process, researchers independently assessed the risk of bias for each study using the Cochrane Collaboration's risk of bias tool [25]. Evaluation criteria included the following: random sequence generation, allocation concealment, blinding of students and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, or other which included publication bias. Funnel plots were used to evaluate publication bias. Risk of bias for each criterion was rate as low, high, or unclear according to the Cochrane risk of bias instructions.

Data Synthesis

Analyses were performed for knowledge outcomes using SAS software (version 9.4; SAS Institute). The standardized mean difference (standard mean difference; Hedges g effect size), converted from means and standard deviation from each study, was used [15]. When the mean was available, but the standard deviation was not, we used the mean standard deviation of all other studies. Since the overall scores of included studies were not the same and standard mean difference could eliminate the effects of absolute values, we adjusted the mean and standard deviation so that the average standard deviation could replace the missing value of standard deviation. We employed a random-effects model for the meta-analysis (statistically significant if $P<.05$).

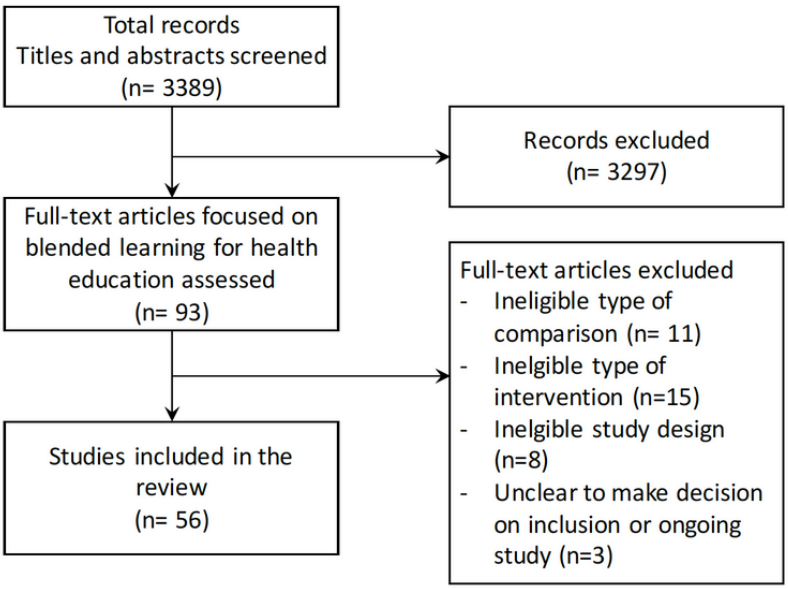
The I^2 statistic was used to quantify heterogeneity across studies [26]. When the estimated I^2 was equal to or greater than 50%, this indicated a large amount of heterogeneity. As the studies were functionally different and involved different study designs, participants, interventions, and settings; a random-effects model that allowed more heterogeneity was used. Forest plots were created to display the meta-analysis findings. To explore publication bias, funnel plots were created and Begg tests were performed (statistically significant if $P<.05$). To explore potential sources of heterogeneity, multiple meta-regression and subgroup analyses based on the study design were performed. Sensitivity analyses to test the robustness of findings were also performed.

Results

Study Selection

The search strategy identified 3389 articles from MEDLINE. After scanning the titles and abstracts, 93 articles were found to be potentially eligible, and their full texts were read for further assessment. Of these, 56 articles were included [9-11,22,27-78] (Figure 1). All articles that were included had been published in peer-reviewed journal.

Figure 1. Study flow diagram.



Type of Participants

In the 56 articles, 9943 participants were included. In 30 out of 56 participant subgroups, participants were from the field of medicine [11,31-34,36,38,39,41,43,44,46-50,53,64-67,69-74,76,78,79]. The participant subgroups from fields other than medicine were as follows: 16 studies in nursing [9,10,12,27,29,35,37,40,51,52,57,58,61-63,75], 1 in pharmacy [37], 3 in physiotherapy [12,30,45], 5 in dentistry [10,42,54,55,59], and 4 interprofessional education [56,60,68,77].

Of the 56 studies, 47 were conducted in high-income countries: 14 were from the United States [9,10,29,31,36,38,43,50,53-55,59,61,73], 2 from Canada [47,58], 5 from Germany [39,41,46,57,76], 3 from the United Kingdom [42,56,75], 3 from Spain [30,45,48], 1 from France [74], 1 from Greece [34], 1 from Sweden [67], 1 from the Netherlands [37], 1 from Korea [40], 1 from Poland [79], 1 from Serbia [70], 1 from Croatia [64], 1 from Turkey [32], 2 from Taiwan [28,51], 1 from Japan [69], and 7 from Australia [44,49,63,65,66,68,78]. Of the 56 studies, 9 studies were conducted in low- or middle-income countries: 2 from Thailand [52,62], 1 from China [77], 1 from Malaysia [72], 2 from Iran [27,71], 1 from Jordan [35], 1 from South Africa [11], and 1 from Uruguay [60].

The technical characteristics of the blended learning systems, topics of educational content, applied design methods, and other information on the validity of outcome measurements can be found in Multimedia Appendix 1.

Effects of Interventions

Blended Learning Versus Traditional Learning

The pooled effect size reflected a significantly large effect on knowledge outcome (standard mean difference 1.07, 95% CI 0.85 to 1.28, z=9.72, n=9943, P<.001). A significant heterogeneity was observed among studies (I^2=94.3%). Figure 2 shows details of the main analysis. The test of asymmetry funnel plot (Figure 3) indicated publication bias among studies (Begg test P=.01). The trim and fill method indicated that the effect size changed to 0.41 (95% CI 0.16 to 0.66, P<.001) after adjusting for publication bias, which suggested that blended learning was more effective than traditional learning.

Figure 2. Forest plot of blended learning to traditional learning comparison for knowledge outcomes. df: degree of freedom; CI: confidence interval; SMD: standard mean difference.

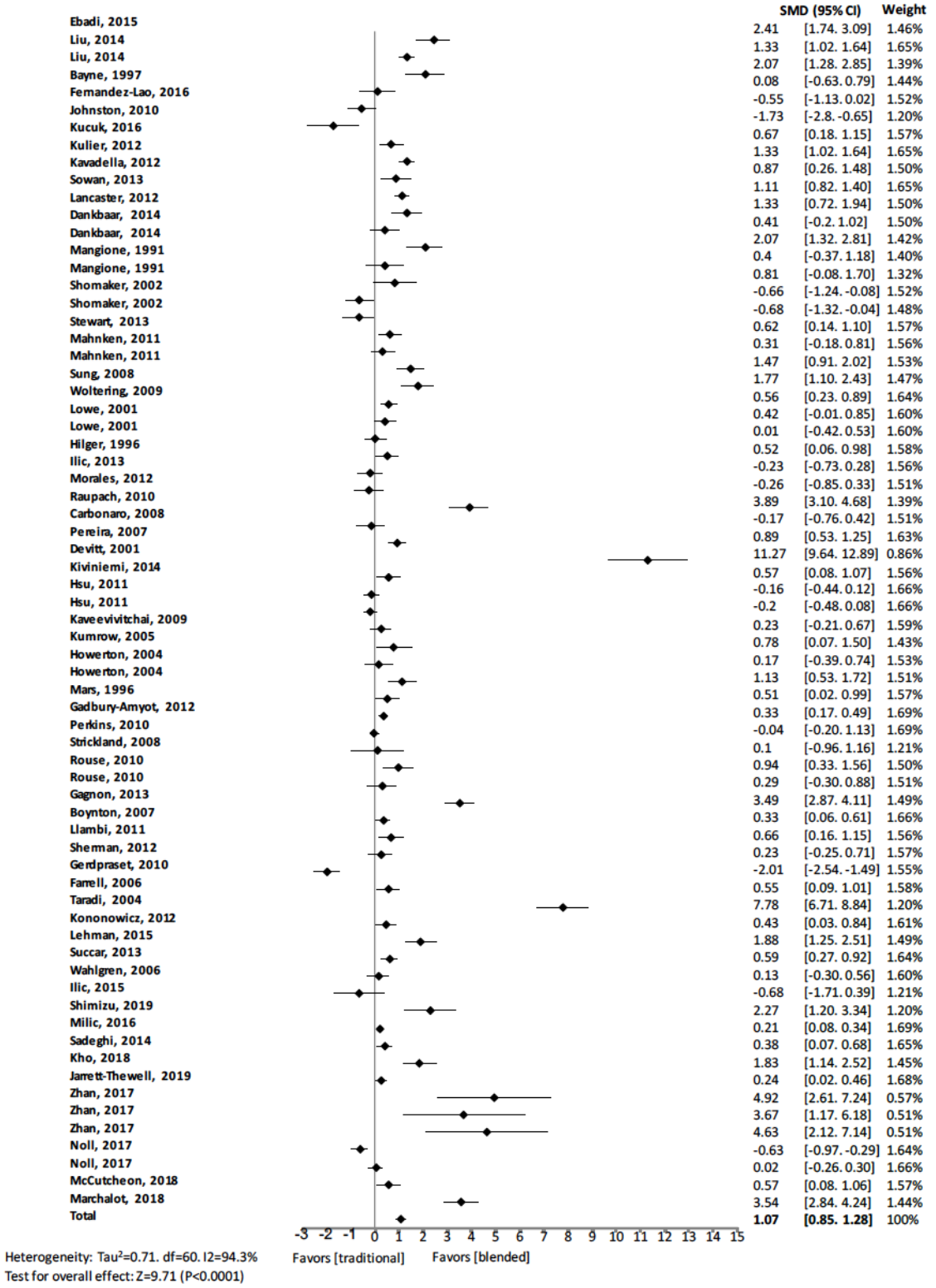
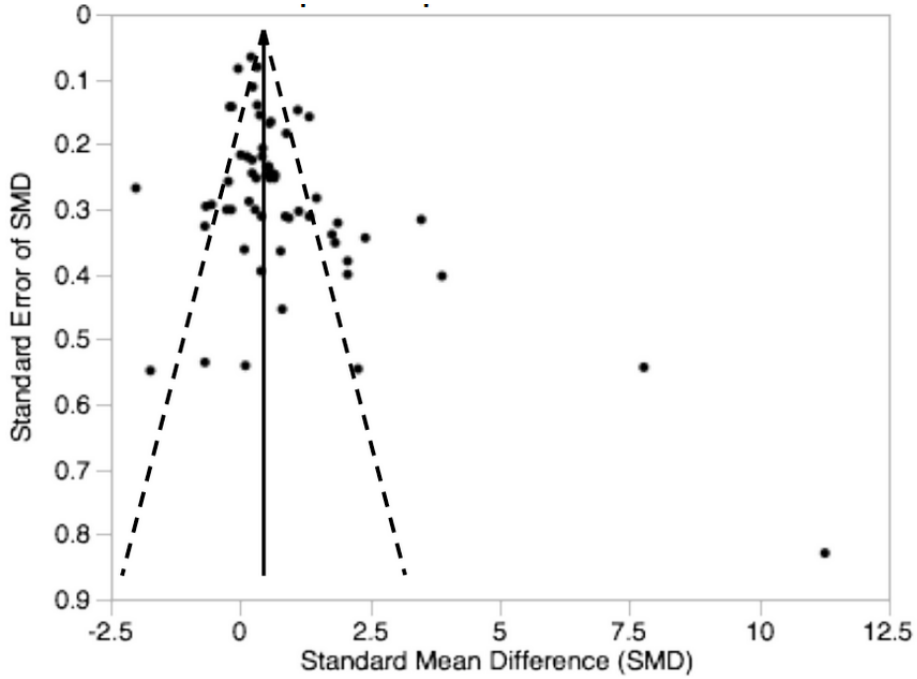


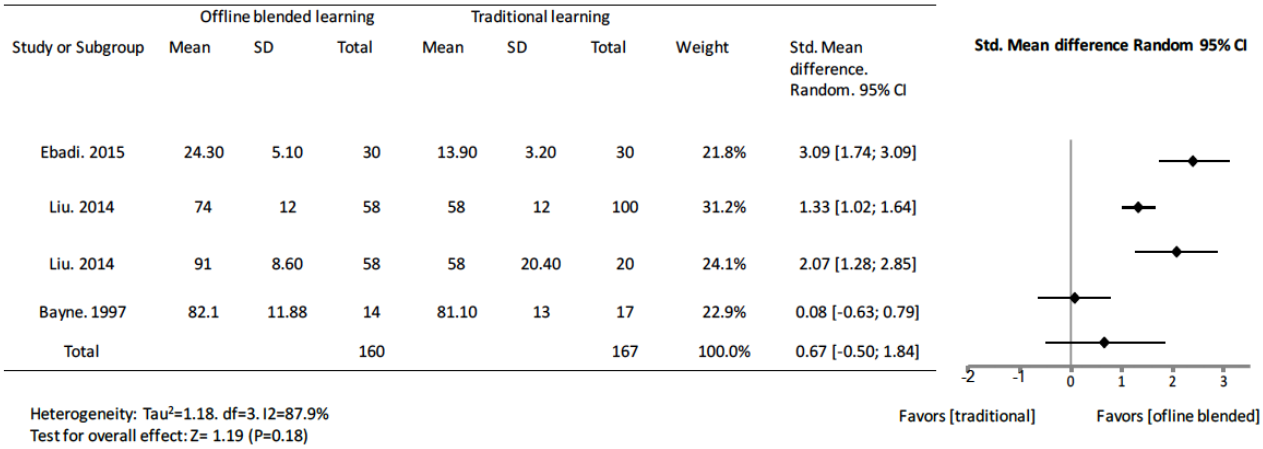
Figure 3. Funnel plot of blended learning versus traditional learning.



Offline Blended Learning Versus Traditional Learning

Of the 3 studies [27-29] comparing offline blended learning to traditional learning, in 2 studies [27,28], the groups with blended resources scored better than their corresponding control groups with significant positive standard mean differences. The other study did not show a statistically significant difference in knowledge outcome (standard mean difference 0.08, 95% CI –0.63 to 0.79) [29]. The pooled effect for knowledge outcomes suggested no significant effects from offline blended learning over traditional education alone (standard mean difference 0.67, 95% CI –0.50 to 1.84, I^2=87.9%, n=327) (Figure 4).

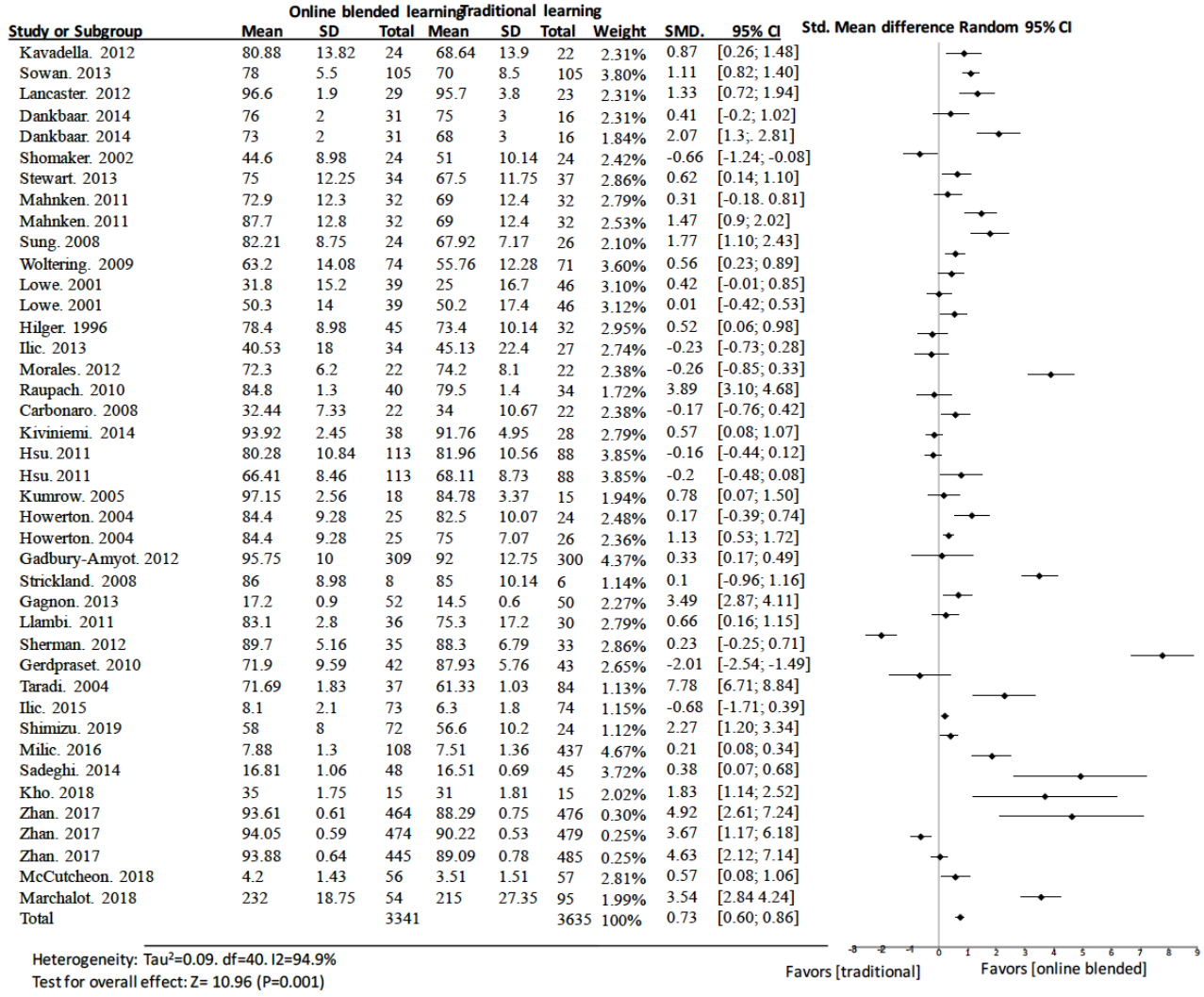
Figure 4. Forest plot of offline blended learning to traditional learning comparison for knowledge outcomes. df: degree of freedom; CI: confidence interval.



Online Blended Learning Versus Traditional Learning

In studies comparing online blended learning to traditional learning, 26 [34-37,39-41,43,46,50,53-55,58,60,64,69-72,74,75,77,78] of the 41 studies [34-47,50,51,53-55,57,58,60,62,64,68-72,74,75,77,78] showed that groups with blended learning had better scores than those of their corresponding control groups. The pooled effect for knowledge outcomes was a standard mean difference of 0.73 (95% CI 0.60 to 0.86, n=6976) (Figure 5). There was a substantial amount of heterogeneity in the pooled analysis (I^2=94.9%).

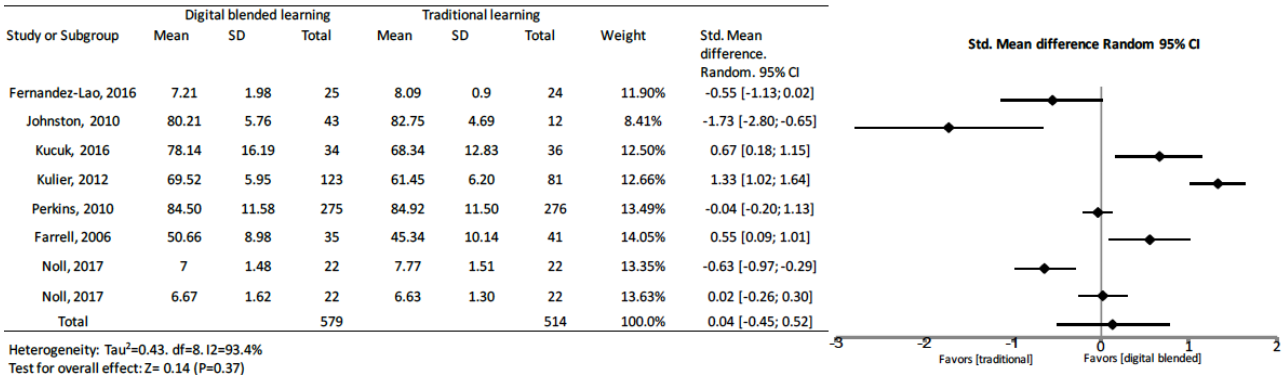
Figure 5. Forest plot of online blended learning to traditional learning comparison for knowledge outcomes. df: degree of freedom; CI: confidence interval; SMD: standard mean difference.



Digital Learning Versus Traditional Learning

Only 3 [32,33,63] of 7 studies [30-33,56,63,76] comparing digital learning to traditional learning presented a better score than the control group. The pooled effect for knowledge outcomes suggested no significant effects between blended and traditional learning (standard mean difference 0.04, 95% CI –0.45 to 0.52, I^2=93.4%, n=1093) (Figure 6).

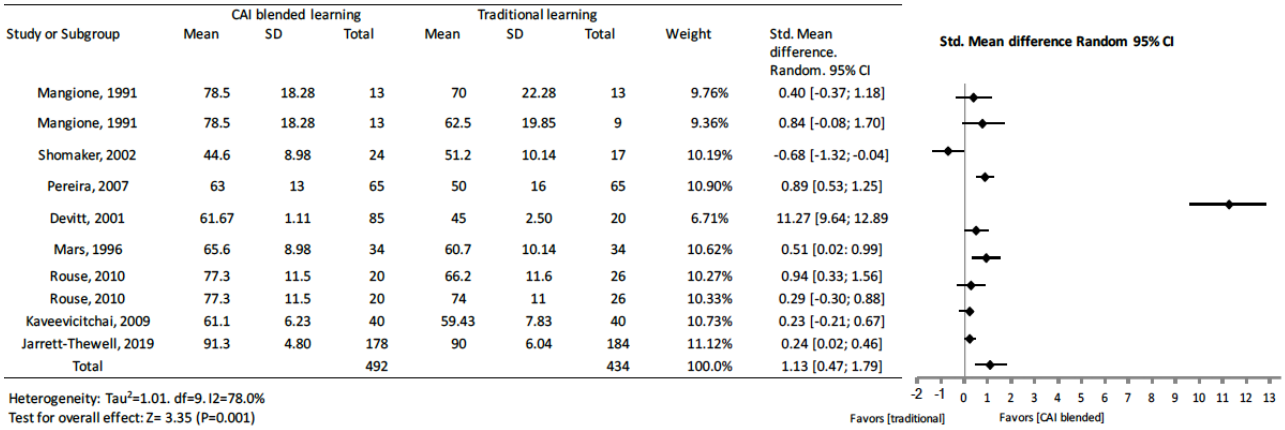
Figure 6. Forest plot of digital blended learning to traditional learning comparison for knowledge outcomes. df: degree of freedom; CI: confidence interval.



Computer-Assisted Instruction Blended Learning Versus Traditional Learning

Of the studies focusing on computer-assisted instruction blended learning, 5 [10,11,48,49,73] of the 8 studies [9-11,38,48,49,52,73] showed significantly higher scores than those of traditional learning. Only 1 study [38] showed a significant negative effect compared to traditional learning (standard mean difference –0.68, 95% CI –1.32 to –0.04). The other studies showed no significant difference [9,52]. The pooled effect for knowledge outcomes suggested a significant improvement of computer-assisted instruction blended with traditional education over traditional education alone (standard mean difference 1.13, 95% CI 0.47 to 1.79, I^2=78.0%, n=926) (Figure 7).

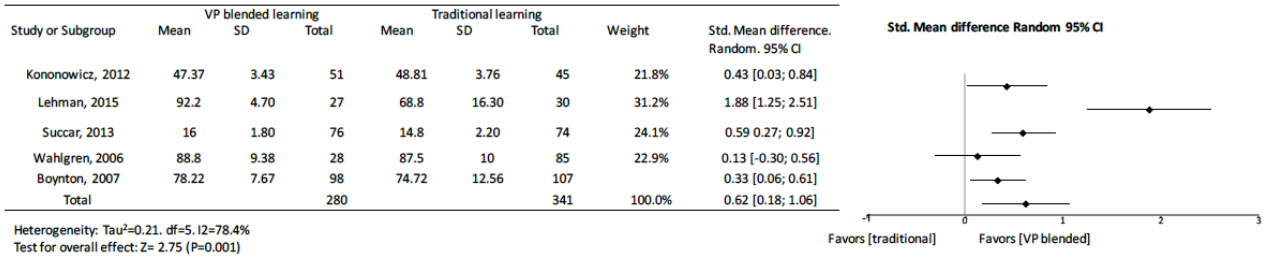
Figure 7. Forest plot of computer-assisted instruction blended learning to traditional learning comparison for knowledge outcomes. df: degree of freedom; CI: confidence interval.



Virtual Patient Blended Learning Versus Traditional Learning

In 4 [59,65,66,79] of the 5 studies [59,65-67,79] on knowledge outcomes when using virtual patients as a supplement to traditional learning, the groups with supplementary virtual patient learning support scored better than their corresponding control groups. Only 1 study with virtual patients did not show a statistically significant difference in knowledge outcomes (standard mean difference 0.13, 95% CI –0.30 to 0.56) [67]. The pooled effect for knowledge outcomes suggested significant effects for virtual patient blended learning (standard mean difference 0.62, 95% CI 0.18 to 1.06, I^2=78.4%, n=621) (Figure 8).

Figure 8. Forest plot of virtual patient blended learning to traditional learning comparison for knowledge outcomes. df: degree of freedom; CI: confidence interval.



Sensitivity Analyses

None of the subgroup analyses that were initially planned explained the heterogeneity of the results. Among many analyzed aspects, we considered the differences regarding the efficiency of learning with blended learning between the health professions disciplines. Most of the studies involved students of medicine as participants (30/56, 54%).

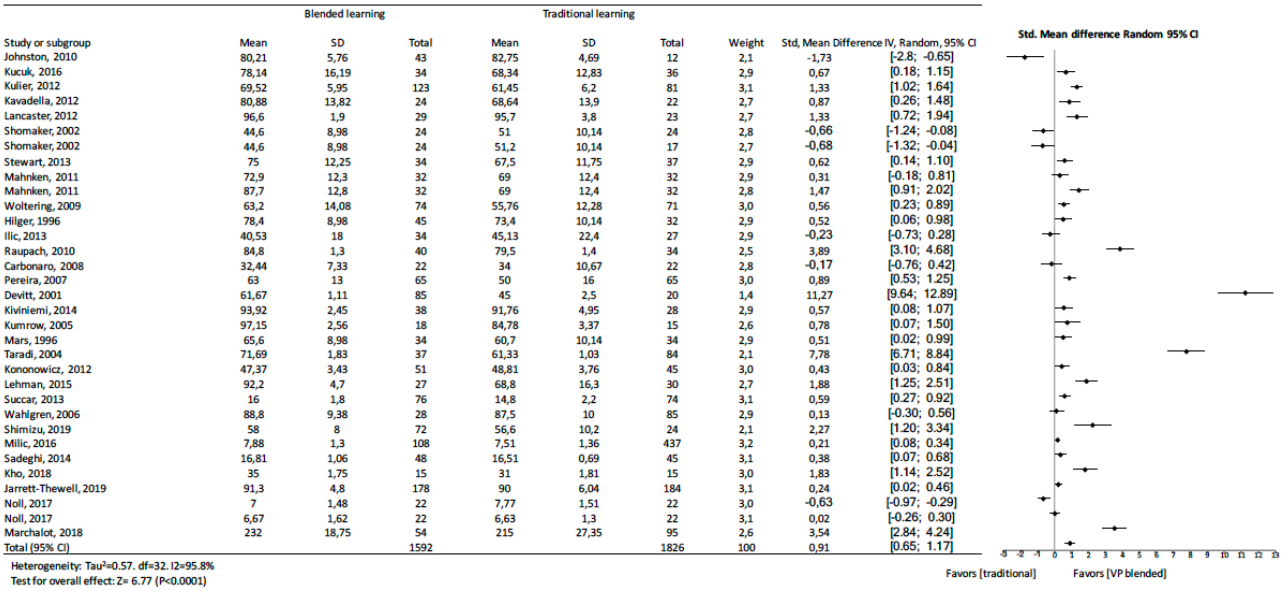
When analyzing knowledge outcomes in medicine, nursing, and dentistry, some differences were apparent. The pooled effect of medicine studies showed a standard mean difference of 0.91 (95% CI 0.65 to 1.17, z= 6.77, I^2=95.8%, n=3418, P<.001) (Figure 9), nursing studies showed a standard mean difference of 0.75 (95% CI 0.26 to 1.24, z=2.99, I^2=94.9%, n=1590, P=.008) (Figure 10), and dentistry studies showed a standard mean difference of 0.35 (95% CI 0.17 to 0.53, z=3.78, I^2=37.6%, n=1130, P=<.001) (Figure 11). Dentistry studies included 3 online blended learning studies (standard mean difference 0.37, 95% CI 0.14 to 0.64, z=2.63, I^2=58.3%, n=879), 1 virtual patient learning study, and 1 computer-assisted instruction learning study.

Additional interest was observed for offline blended learning in nursing compared to traditional learning (standard mean difference 1.28, 95% CI 0.25 to 2.31, z=2.43, I^2=86.2%, n=249), and in computer-assisted instruction (standard mean difference 0.53, 95% CI 0.17 to 0.90, z=2.84, I^2=23.9%, n=174), but not for online blended learning (standard mean difference 0.68, 95% CI –0.07 to 1.45, z=1.76, I^2=96.7%, n=1091).

Additional interest was observed for digital blended learning compared to traditional learning in medicine (standard mean difference 0.26, 95% CI 0.07 to 0.45, z=2.71, I^2=95.6%, n=417)

[31-33,76], in virtual patient (standard mean difference 0.71, 95% CI 0.14 to 1.28, z=2.45, I^2=85.8%, n=416) [65-67,79], in online (standard mean difference 1.26, 95% CI 0.81 to 1.71, z=5.49, I^2=96.1%, n=1879) [34,36,38,39,41,43,44,46,47,50,53,64,69-72,74,78], and in computer-aided-instruction (standard mean difference 2.1, 95% CI 0.68 to 3.44, z=2.91, I^2=97.9%, n=706) [11,38,48,49,73] suggesting more positive effects of blended learning over traditional learning alone for learning in medicine.

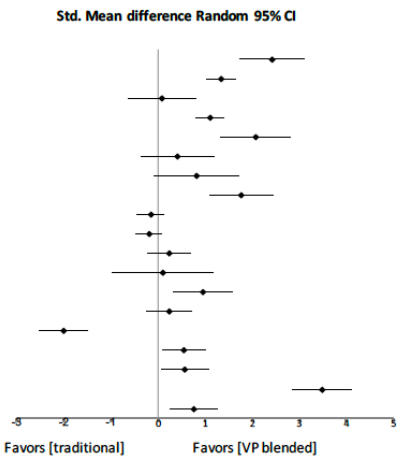
Figure 9. Forest plot of blended learning to traditional learning comparison for knowledge outcomes for medical students. df: degree of freedom; CI: confidence interval.



Study or subgroup	Mean	SD	Total	Mean	SD	Total	Weight	Std, Mean Difference IV, Random, 95% CI
Johnston, 2010	80,21	5,76	43	82,75	4,69	12	2,1	-1,73 [-2.8; -0.65]
Kucuk, 2016	78,14	16,19	34	68,34	12,83	36	2,9	0,67 [0.18; 1.15]
Kulier, 2012	69,52	5,95	123	61,45	6,2	81	3,1	1,33 [1.02; 1.64]
Kavadella, 2012	80,88	13,82	24	68,64	13,9	22	2,7	0,87 [0.26; 1.48]
Lancaster, 2012	96,6	1,9	29	95,7	3,8	23	2,7	1,33 [0.72; 1.94]
Shomaker, 2002	44,6	8,98	24	51	10,14	24	2,8	-0,66 [-1.24; -0.08]
Shomaker, 2002	44,6	8,98	24	51,2	10,14	17	2,7	-0,68 [-1.32; -0.04]
Stewart, 2013	75	12,25	34	67,5	11,75	37	2,9	0,62 [0.14; 1.10]
Mahnken, 2011	72,9	12,3	32	69	12,4	32	2,9	0,31 [-0.18; 0.81]
Mahnken, 2011	87,7	12,8	32	69	12,4	32	2,8	1,47 [0.91; 2.02]
Woltering, 2009	63,2	14,08	74	55,76	12,28	71	3,0	0,56 [0.23; 0.89]
Hilger, 1996	78,4	8,98	45	73,4	10,14	32	2,9	0,52 [0.06; 0.98]
Ilic, 2013	40,53	18	34	45,13	22,4	27	2,9	-0,23 [-0.73; 0.28]
Raupach, 2010	84,8	1,3	40	79,5	1,4	34	2,5	3,89 [3.10; 4.68]
Carbonaro, 2008	32,44	7,33	22	34	10,67	22	2,8	-0,17 [-0.76; 0.42]
Pereira, 2007	63	13	65	50	16	65	3,0	0,89 [0.53; 1.25]
Devitt, 2001	61,67	1,11	85	45	2,5	20	1,4	11,27 [9.64; 12.89]
Kiviniemi, 2014	93,92	2,45	38	91,76	4,95	28	2,9	0,57 [0.08; 1.07]
Kumrow, 2005	97,15	2,56	18	84,78	3,37	15	2,6	0,78 [0.07; 1.50]
Mars, 1996	65,6	8,98	34	60,7	10,14	34	2,9	0,51 [0.02; 0.99]
Taradi, 2004	71,69	1,83	37	61,33	1,03	84	2,1	7,78 [6.71; 8.84]
Kononowicz, 2012	47,37	3,43	51	48,81	3,76	45	3,0	0,43 [0.03; 0.84]
Lehman, 2015	92,2	4,7	27	68,8	16,3	30	2,7	1,88 [1.25; 2.51]
Succar, 2013	16	1,8	76	14,8	2,2	74	3,1	0,59 [0.27; 0.92]
Wahlgren, 2006	88,8	9,38	28	87,5	10	85	2,9	0,13 [-0.30; 0.56]
Shimizu, 2019	58	8	72	56,6	10,2	24	2,1	2,27 [1.20; 3.34]
Milic, 2016	7,88	1,3	108	7,51	1,36	437	3,2	0,21 [0.08; 0.34]
Sadeghi, 2014	16,81	1,06	48	16,51	0,69	45	3,1	0,38 [0.07; 0.68]
Kho, 2018	35	1,75	15	31	1,81	15	3,0	1,83 [1.14; 2.52]
Jarrett-Thewell, 2019	91,3	4,8	178	90	6,04	184	3,1	0,24 [0.02; 0.46]
Noll, 2017	7	1,48	22	7,77	1,51	22	3,0	-0,63 [-0.97; -0.29]
Noll, 2017	6,67	1,62	22	6,63	1,3	22	3,1	0,02 [-0.26; 0.30]
Marchalot, 2018	232	18,75	54	215	27,35	95	2,6	3,54 [2.84; 4.24]
Total (95% CI)			1592			1826	100	0,91 [0.65; 1.17]

Heterogeneity: Tau²=0.57. df=32. I2=95.8%
Test for overall effect: Z= 6.77 (P<0.0001)

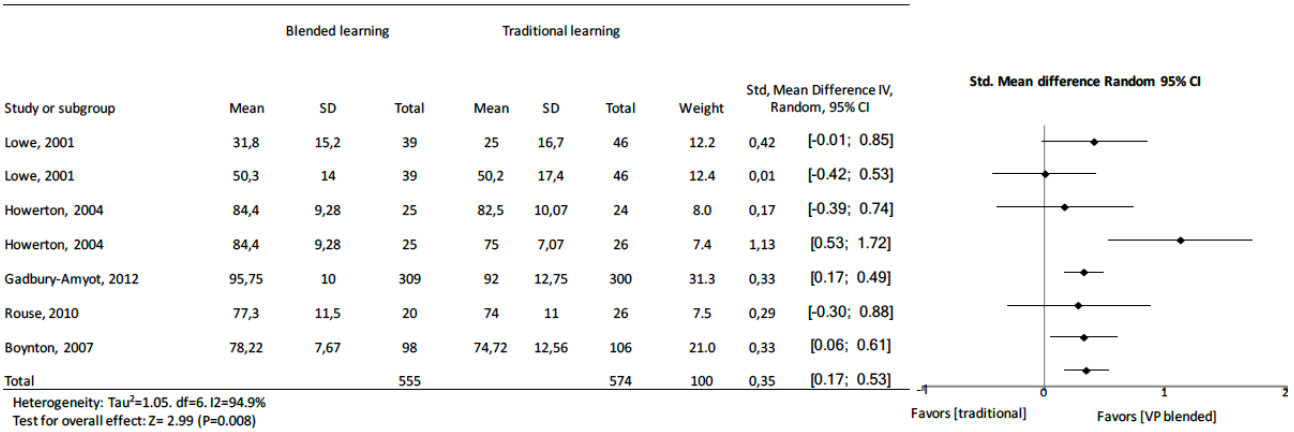
Figure 10. Forest plot of blended learning to traditional learning comparison for knowledge outcomes for nurses as students. df: degree of freedom; CI: confidence interval.



Study or subgroup	Blended learning Mean	SD	Total	Traditional learning Mean	SD	Total	Weight	Std, Mean Difference IV, Random, 95% CI
Ebadi, 2015	24,3	5,1	30	13,9	3,2	30	5,40	2,41 [1.74; 3.09]
Liu, 2014	74	12	58	58	12	100	5,90	1,33 [1.02; 1.64]
Bayne, 1997	82,1	11,88	14	81,1	13	17	5,40	0,08 [-0.63; 0.79]
Sowan, 2013	78	5,5	105	70	8,5	105	5,90	1,11 [0.82; 1.40]
Dankbaar, 2014	73	2	31	68	3	16	5,30	2,07 [1.32; 2.81]
Mangione, 1991	78,5	18,28	13	70	22,28	13	5,30	0,4 [-0.37; 1.18]
Mangione, 1991	78,5	18,28	13	62,5	19,85	9	5,10	0,81 [-0.08; 1.70]
Sung, 2008	82,21	8,75	24	67,92	7,17	26	5,50	1,77 [1.10; 2.43]
Hsu, 2011	80,28	10,84	113	81,96	10,56	88	5,90	-0,16 [-0.44; 0.12]
Hsu, 2011	66,41	8,46	113	68,11	8,73	88	5,90	-0,2 [-0.48; 0.08]
Kaveevivitchai, 2009	61,1	6,23	40	59,43	7,83	40	5,80	0,23 [-0.21; 0.67]
Strickland, 2008	86	8,98	8	85	10,14	6	4,70	0,1 [-0.96; 1.16]
Rouse, 2010	77,3	11,5	20	66,2	11,6	26	5,50	0,94 [0.33; 1.56]
Sherman, 2012	89,7	5,16	35	88,3	6,79	33	5,50	0,23 [-0.25; 0.71]
Gerdpraset, 2010	71,9	9,59	42	87,93	5,76	43	5,70	-2,01 [-2.54; -1.49]
Farrell, 2006	50,66	8,98	35	45,34	10,14	41	5,70	0,55 [0.09; 1.01]
McCutcheon, 2018	4,2	1,43	56	3,51	1,51	57	5,70	0,57 [0.08; 1.06]
Gagon, 2013	17,2	0,9	52	14,5	0,6	50	5,70	3,49 [2.87; 4.11]
Total (95% CI)			802			788	100	0,75 [0.26; 1,24]

Heterogeneity: Tau²=1.05. df=17. I2=94.9%
Test for overall effect: Z= 2.99 (P=0.008)

Figure 11. Forest plot of blended learning to traditional learning comparison for knowledge outcomes for dentistry students. df: degree of freedom; CI: confidence interval.



Study or subgroup	Mean	SD	Total	Mean	SD	Total	Weight	Std, Mean Difference IV, Random, 95% CI
Lowe, 2001	31,8	15,2	39	25	16,7	46	12.2	0,42 [-0.01; 0.85]
Lowe, 2001	50,3	14	39	50,2	17,4	46	12.4	0,01 [-0.42; 0.53]
Howerton, 2004	84,4	9,28	25	82,5	10,07	24	8.0	0,17 [-0.39; 0.74]
Howerton, 2004	84,4	9,28	25	75	7,07	26	7.4	1,13 [0.53; 1.72]
Gadbury-Amyot, 2012	95,75	10	309	92	12,75	300	31.3	0,33 [0.17; 0.49]
Rouse, 2010	77,3	11,5	20	74	11	26	7.5	0,29 [-0.30; 0.88]
Boynton, 2007	78,22	7,67	98	74,72	12,56	106	21.0	0,33 [0.06; 0.61]
Total			555			574	100	0,35 [0.17; 0.53]

Heterogeneity: Tau^2=1.05. df=6. I2=94.9%
Test for overall effect: Z= 2.99 (P=0.008)

Risk of Bias

Risk of bias is shown in Figure 12. The risk of bias of evaluators was avoided in several studies by using automated assessment instruments. Thus, we rated the risk as low in 50 of the 56 studies. Nevertheless, it was still unclear whether the instruments had been correctly validated. Attribution bias was within acceptable levels in some studies (low risk in 24 of the 56 studies), but this did not exclude voluntary bias and its influence on the estimated effect. Reporting bias was considered low in 28 of the 56 studies.

We cannot consider allocation bias as a significant problem in this review because, if studies described an adequate randomization method or an unclear description, it was assumed that randomization was unlikely be defective. Performance bias on traditional learning may be a problem, but it is impossible to avoid in this type of research. It is possible to blind participants in blended learning design comparisons, but these studies are still rare in the literature. We cannot reliably estimate publication bias given the high degree of heterogeneity of the included studies.

Figure 12. Risk of bias summary (+ low risk of bias; - high risk of bias; ? unclear risk of bias).

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attribution bias)	Selective reporting (reporting data)	Others
Bayne 1997	?	?	?	-	+	+
Liu 2014	+	?	-	-	?	+
Liu 2014a	?	?	?	-	+	+
Ebadi 2015	+	?	?	+	+	+
Fernandez-lao 2016	+	?	+	+	+	?
Johnston 2010	+	?	+	+	+	?
Kucuk 2016	+	?	+	+	+	+
Kononowicz 2012	?	?	+	-	+	+
Lehman 2015	?	+	+	+	+	?
Succar 2013	+	?	?	?	+	+
Wahlgren 2006	?	?	+	+	+	+
Kulier 2012	+	+	-	+	-	-
Kavadella 2012	+	?	?	+	+	+
Mangione 1991	-	?	?	+	+	?
Rouse 2000	?	?	?	-	+	+
Mars 1996	?	?	?	?	+	+
Sowan 2013	-	-	?	+	+	?
Lancaster 2012	-	?	-	+	+	+
Dankbaar 2014	-	?	-	-	+	+
Shomaker 2002	+	+	?	+	+	+
Stewart 2013	+	+	+	+	+	+
Mahnken 2011	+	+	?	+	?	?
Sung 2008	?	?	?	+	+	?
Woltering 2008	?	+	-	+	+	?
Lowe 2001	+	+	?	+	+	+
Hilger 1996	+	+	?	?	+	+
Ilic 2013	-	-	-	+	+	?
Arroyo-Morales 2012	+	+	?	+	+	+
Raupach 2010	?	?	?	+	+	+
Carbonaro 2008	+	+	?	+	+	?
Pereira 2008	?	?	?	+	+	?
Devitt 2001	-	-	-	+	+	?
Kiviniemi 2014	-	-	?	+	+	+
Hsu 2011	+	+	?	+	+	+
Kaveevivitchai 2009	+	+	?	+	+	?
Kumrow 2007	-	-	-	+	+	?
Howerton 2004	+	+	?	+	+	?
Gadbury-Amyot 2013	?	?	-	+	+	?
Perkins 2010	+	+	?	+	+	+
Strickland 2009	?	?	?	+	+	?
Gagnon 2013	+	+	?	+	+	+
Boynton 2007	-	-	-	+	-	?
Llambi 2011	?	?	?	+	?	?
Sherman 2012	+	+	+	+	+	+
Gerdprasert 2010	+	+	?	+	+	+
Farrell 2008	-	-	?	+	?	?
Taradi 2005	?	?	?	+	+	+
Ilic 2015	+	+	+	+	+	?
Shimizu 2019	?	+	?	+	?	?
Milic 2016	-	-	+	+	+	?
Sadeghi 2014	+	+	?	?	+	?
Kho 2018	+	+	+	+	+	+
Jarrett-Thelwell 2019	-	-	?	+	+	-
Marchalot 2018	-	-	?	+	+	?
McCutcheon 2018	+	+	-	+	+	+
Noll 2017	+	+	+	+	+	?
Zhan 2017	+	+	?	+	+	+

Discussion

Principal Findings

This meta-analysis provided several findings. First, blended learning had a large consistent positive effect (standard mean difference 1.07, 95% CI 0.85 to 1.28) on knowledge acquisition in comparison to traditional learning in health professions. A possible explanation could be that, compared to traditional learning, blended learning allowed students to review electronic materials as often as necessary and at their own pace, and this likely enhanced learning performance [80]. The trim and fill method showed that the pooled effect size changed to 0.41 (95% CI 0.16 to 0.66), meaning that blended learning remained more effective than traditional learning. The strength of this meta-analysis was that it reinforced previous results [12]; however, a large heterogeneity was observed across the studies.

The participant subgroup analyses partially explained these differences.

The effectiveness of blended learning is complex and dependent on how well the evaluation fits, since it occurs before the implementation of any innovation as well as allowing planners to determine the needs, considering participant characteristics, analyzing contextual matters, and gathering baseline information [81]. Some interventional studies have highlighted the potential of blended learning to increase course completion rates, improve retention, and increase student satisfaction [82]. Nevertheless, comparisons between blended learning environments and the disparity between academic achievement or grade dispersions have been studied; no significant differences were observed [83]. The effectiveness of blended learning may be dependent on student characteristics, design features, and learning outcomes. Learner success is dependent on the ability to cope with technical difficulty, technical skills, and knowledge in computer operations and internet navigation. Thus, the success of blended learning is highly dependent on experience with the internet and computer apps. Some studies have observed that the success of blended learning was largely associated on the capability to participate in blended course. A previous study [84] showed that high motivation among blended learning led to persistence in their courses. Moreover, time management is a crucial effectiveness factor for successful online learning [85].

Second, offline blended learning did not show a positive pooled effect compared to traditional learning; however, 2 of the 3 studies were in nursing. These results were consistent with a previous meta-analysis on offline digital education [86]. Nevertheless, potential benefits of offline education such as unrestrained knowledge transfer and enriched accessibility of health education have previously been suggested [87]. These interventions could be focused on an interactive, associative, and perceptual learning experience by text, images, audio-video, or other components [88,89].

Third, the effect of digital learning on knowledge outcomes presented inconsistent effects according to the group or subgroup analysis. Overall, the 8 digital blended learning studies showed a nonsignificant effect compared to traditional learning whereas in the medicine subgroup, digital learning had a positive effect (standard mean difference 0.26, 95% CI 0.07 to 0.45). Previous studies [18,90] have shown similar results. Nevertheless, George et al [18] showed the effectiveness of digital learning for undergraduate health professionals compared to traditional learning.

Fourth, in the 10 studies related to computer-assisted instruction, we observed a significant difference in knowledge acquisition outcomes. Furthermore, the difference was higher in the medicine subgroup. This finding must be interpreted with caution because of the high level of heterogeneity (all computer-assisted instruction: I^2=78.0%; medicine computer-assisted instruction: I^2=97.9%). Previous studies showed that computer-assisted instruction was equally as effective as traditional learning [91]. Nevertheless, the results of these studies also had high levels of heterogeneity and require cautious interpretation. We believe that a comparative approach focusing on the differences in intervention design, sample characteristics, and context of learning is needed to better understand the effectiveness of computer-assisted instruction. Computer-assisted instruction could be perceived negatively by some students and impact outcomes.

Fifth, the participants in Al-Riyami et al's study [92] reported difficulties accessing the course because of network difficulties with university's server and internet; therefore, the asynchronous features of the discussion boards were not used to their full potential in this study. Both problems could have emerged regardless of the online course. In traditional learning, students may choose not to engage in discussions, and internet connectivity issues can happen anywhere. This supports the contention above that local conditions, rather than a general effect, may render one or the other mode of instruction preferable to the other.

Sixth, virtual patient blended learning had an overall positive pooled effect when compared to traditional learning on knowledge outcomes; this was also found in a similar meta-analysis [93]. Our observations also supplement the evidence in previous reviews [94,95] which included studies since 2010. Nevertheless, virtual patient simulations predominantly affect skill rather than knowledge outcomes. This could explain the low number of studies and the low added value of virtual patient in comparison to traditional learning. Virtual patients have greater impact in skills training, in applying problem solving, and when direct patient contact is not possible [93]. As proposed by Cook and Triola [96], virtual patients can be said to be a modality for learning in which learners actively use and train their clinical reasoning and critical thinking abilities before bedside learning [96]. Nevertheless, some exceptions can be noted. A need for more guidance within virtual patient simulations may appear in studies with different instructional methods where narrative virtual patient design was better than more autonomous problem-oriented designs [97]. Feedback given by humans in a virtual patient system was better than an animated backstory in increasing empathy [98], but no feedback had no more positive result on the outcomes than learning from a virtual patient scenario [99]. This reminds us that presenting realistic patient scenarios with a great degree of freedom may not be an excuse for neglecting guidance in relation to learning objectives [100].

Strengths and Limitations

This meta-analysis had numerous strengths. An evaluation of the effectiveness of blended learning for health professions is timely and very important for both health educators and learners. We intentionally kept our scope broad in terms of learning topic and included all studies with learners from health professions.

The samples used in this study consisted in various health professionals (in medicine, nursing, dentistry, and others) across a wide variety of health care disciplines. Although, these observations could explain the high level of heterogeneity, we found moderate or large effects for the pooled effects sizes of almost all subgroup analyses exploring variations in participant types. Thus, these results could suggest that health care learning should use blended learning in several and various disciplines of health learning.

However, some limitations must be considered. The systematic literature search encompassed one database (MEDLINE) with few exclusion criteria. The quality of the analyses was dependent on the quality of data from the included studies. Although the standard deviation of some interventions was not available due to poor reporting, we used the average standard deviation of the other studies and imputed effect sizes with concomitant potential for error. Results of subgroup analyses should be interpreted with caution because of the absence of a priori hypotheses in some cases, such as study design, country socioeconomic status, and outcome assessment. In addition, since variability of study interventions, assessment instruments, circumstances were not assessed and could be potential sources of heterogeneity, this should also be cause for cautious interpretation of results. Finally, publication bias was also found.

Conclusions

This study has implications for research on blended learning in health professions. Even though conclusions could be weakened by heterogeneity across studies, the results of this synthesis reinforced that blended learning may have a positive effect on knowledge acquisition related to health professions. Blended learning could be promising and worthwhile for further application in health professions. The difference in effects across subgroup analyses of health population indicated that different methods of conducting blended courses could demonstrate differing effectiveness. Therefore, researchers and educators should pay attention to how to implement a blended course effectively. This question may be answered successfully through studies directly comparing different blended instructional methods.

Authors' Contributions

AV performed the design and conceived the original idea. AV and ES read the articles and selected the articles. AV performed the statistical analyses. AV wrote the manuscript. ES, AC, and JB participated in the writing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental appendix.
[DOCX File , 82 KB - jmir_v22i8e16504_app1.docx]

References

1. Bediang G, Stoll B, Geissbuhler A, Klohn AM, Stuckelberger A, Nko'o S, et al. Computer literacy and e-learning perception in Cameroon: the case of Yaounde Faculty of Medicine and Biomedical Sciences. BMC Med Educ 2013;13:57 [FREE Full text] [doi: 10.1186/1472-6920-13-57] [Medline: 23601853]
2. Choules AP. The use of elearning in medical education: a review of the current situation. Postgrad Med J 2007 Apr;83(978):212-216 [FREE Full text] [doi: 10.1136/pgmj.2006.054189] [Medline: 17403945]
3. Yu S, Yang K. Attitudes toward web-based distance learning among public health nurses in Taiwan: a questionnaire survey. Int J Nurs Stud 2006 Aug;43(6):767-774. [doi: 10.1016/j.ijnurstu.2005.09.005] [Medline: 16253261]
4. Peng Y, Wu X, Atkins S, Zwarentein M, Zhu M, Zhan XX, et al. Internet-based health education in China: a content analysis of websites. BMC Med Educ 2014;14:16 [FREE Full text] [doi: 10.1186/1472-6920-14-16] [Medline: 24467710]
5. Moreira IC, Ventura SR, Ramos I, Rodrigues PP. Development and assessment of an e-learning course on breast imaging for radiographers: a stratified randomized controlled trial. J Med Internet Res 2015;17(1):e3 [FREE Full text] [doi: 10.2196/jmir.3344] [Medline: 25560547]
6. Kemp N, Grieve R. Face-to-face or face-to-screen? Undergraduates' opinions and test performance in classroom vs. online learning. Front Psychol 2014;5:1278 [FREE Full text] [doi: 10.3389/fpsyg.2014.01278] [Medline: 25429276]
7. Kim K, Bonk CJ, Oh E. The present and future state of blended learning in workplace learning settings in the United States. Perf. Improv 2008 Sep;47(8):5-16. [doi: 10.1002/pfi.20018]
8. Dziuban C, Graham CR, Moskal PD, Norberg A, Sicilia N. Blended learning: the new normal and emerging technologies. Int J Educ Technol High Educ 2018 Feb 15;15(1):15-13. [doi: 10.1186/s41239-017-0087-5]
9. Mangione S, Nieman LZ, Greenspon LW, Margulies H. A comparison of computer-assisted instruction and small-group teaching of cardiac auscultation to medical students. Med Educ 1991 Sep;25(5):389-395. [Medline: 1758315]
10. Rouse DP. The effectiveness of computer-assisted instruction in teaching nursing students about congenital heart disease. Comput Nurs 2000;18(6):282-287. [Medline: 11105402]
11. Mars M, McLean M. Students' perceptions of a multimedia computer-aided instruction resource in histology. S Afr Med J 1996 Sep;86(9):1098-1102. [Medline: 8888778]
12. Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. The effectiveness of blended learning in health professions: systematic review and meta-analysis. J Med Internet Res 2016;18(1):e2 [FREE Full text] [doi: 10.2196/jmir.4807] [Medline: 26729058]
13. Rowe M, Frantz J, Bozalek V. The role of blended learning in the clinical education of healthcare students: a systematic review. Med Teach 2012;34(4):e216-e221. [doi: 10.3109/0142159X.2012.642831] [Medline: 22455712]

14. McCutcheon K, Lohan M, Traynor M, Martin D. A systematic review evaluating the impact of online or blended learning vs. face-to-face learning of clinical skills in undergraduate nurse education. J Adv Nurs 2015 Feb;71(2):255-270. [doi: 10.1111/jan.12509] [Medline: 25134985]
15. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
16. Jwayyed S, Stiffler KA, Wilber ST, Southern A, Weigand J, Bare R, et al. Technology-assisted education in graduate medical education: a review of the literature. Int J Emerg Med 2011;4:51 [FREE Full text] [doi: 10.1186/1865-1380-4-51] [Medline: 21824405]
17. Rasmussen K, Belisario JM, Wark PA, Molina JA, Loong SL, Cotic Z, et al. Offline eLearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014 Jun;4(1):010405 [FREE Full text] [doi: 10.7189/jogh.04.010405] [Medline: 24976964]
18. George PP, Papachristou N, Belisario JM, Wang W, Wark PA, Cotic Z, et al. Online eLearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014 Jun;4(1):010406 [FREE Full text] [doi: 10.7189/jogh.04.010406] [Medline: 24976965]
19. Martinengo L, Yeo NJY, Tang ZQ, Markandran KD, Kyaw BM, Tudor Car L. Digital education for the management of chronic wounds in health care professionals: protocol for a systematic review by the Digital Health Education Collaboration. JMIR Res Protoc 2019 Mar 25;8(3):e12488 [FREE Full text] [doi: 10.2196/12488] [Medline: 30907743]
20. Bandla H, Franco RA, Simpson D, Brennan K, McKanry J, Bragg D. Assessing learning outcomes and cost effectiveness of an online sleep curriculum for medical students. J Clin Sleep Med 2012 Aug 15;8(4):439-443 [FREE Full text] [doi: 10.5664/jcsm.2042] [Medline: 22893775]
21. Viljoen CA, Scott Millar R, Engel ME, Shelton M, Burch V. Is computer-assisted instruction more effective than other educational methods in achieving ECG competence among medical students and residents? Protocol for a systematic review and meta-analysis. BMJ Open 2017 Dec 26;7(12):e018811 [FREE Full text] [doi: 10.1136/bmjopen-2017-018811] [Medline: 29282268]
22. Kononowicz AA, Zary N, Edelbring S, Corral J, Hege I. Virtual patients - what are we talking about? a framework to classify the meanings of the term in healthcare education. BMC Med Educ 2015;15:11 [FREE Full text] [doi: 10.1186/s12909-015-0296-3] [Medline: 25638167]
23. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
24. Higgins J, Green S. editors. Cochrane Handbook for Systematic Reviews of Interventions Internet. Chichester, UK: John Wiley & Sons, Ltd; 2008.
25. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [Medline: 22008217]
26. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
27. Ebadi A, Yousefi S, Khaghanizade M, Saeid Y. Assessment competency of nurses in biological incidents. Trauma Mon 2015 Nov;20(4):e25607 [FREE Full text] [doi: 10.5812/traumamon.25607] [Medline: 26839862]
28. Liu W, Chu K, Chen S. The development and preliminary effectiveness of a nursing case management e-learning program. Comput Inform Nurs 2014 Jul;32(7):343-352. [doi: 10.1097/CIN.0000000000000050] [Medline: 24633234]
29. Bayne T, Bindler R. Effectiveness of medication calculation enhancement methods with nurses. J Nurs Staff Dev 1997;13(6):293-301. [Medline: 9429370]
30. Fernández-Lao C, Cantarero-Villanueva I, Galiano-Castillo N, Caro-Morán E, Díaz-Rodríguez L, Arroyo-Morales M. The effectiveness of a mobile application for the development of palpation and ultrasound imaging skills to supplement the traditional learning of physiotherapy students. BMC Med Educ 2016 Oct 19;16(1):274 [FREE Full text] [doi: 10.1186/s12909-016-0775-1] [Medline: 27756288]
31. Johnston R, Hepworth J, Goldsmith M, Lacasse C. Use of iPod™ technology in medical-surgical nursing courses: effect on grades. Int J Nurs Educ Scholarsh 2010;7:Article43. [doi: 10.2202/1548-923X.2092] [Medline: 21126230]
32. Küçük S, Kapakin S, Göktaş Y. Learning anatomy via mobile augmented reality: effects on achievement and cognitive load. Anat Sci Educ 2016 Oct;9(5):411-421. [doi: 10.1002/ase.1603] [Medline: 26950521]
33. Kulier R, Gülmezoglu AM, Zamora J, Plana MN, Carroli G, Cecatti JG, et al. Effectiveness of a clinically integrated e-learning course in evidence-based medicine for reproductive health training: a randomized trial. JAMA 2012 Dec 5;308(21):2218-2225. [doi: 10.1001/jama.2012.33640] [Medline: 23212499]
34. Kavadella A, Tsiklakis K, Vougiouklakis G, Lionarakis A. Evaluation of a blended learning course for teaching oral radiology to undergraduate dental students. Eur J Dent Educ 2012 Feb;16(1):e88-e95. [doi: 10.1111/j.1600-0579.2011.00680.x] [Medline: 22251359]
35. Sowan AK, Jenkins LS. Use of the seven principles of effective teaching to design and deliver an interactive hybrid nursing research course. Nurs Educ Perspect 2013;34(5):315-322. [Medline: 24245382]

36. Lancaster JW, Wong A, Roberts SJ. 'Tech' versus 'talk': a comparison study of two different lecture styles within a Master of Science nurse practitioner course. Nurse Educ Today 2012 Jul;32(5):e14-e18. [doi: 10.1016/j.nedt.2011.09.018] [Medline: 22071277]
37. Dankbaar MEW, Storm DJ, Teeuwen IC, Schuit SCE. A blended design in acute care training: similar learning results, less training costs compared with a traditional format. Perspect Med Educ 2014 Sep;3(4):289-299 [FREE Full text] [doi: 10.1007/s40037-014-0109-0] [Medline: 24477971]
38. Shomaker TS, Ricks DJ, Hale DC. A prospective, randomized controlled study of computer-assisted learning in parasitology. Acad Med 2002 May;77(5):446-449. [Medline: 12010707]
39. Mahnken AH, Baumann M, Meister M, Schmitt V, Fischer MR. Blended learning in radiology: is self-determined learning really more effective? Eur J Radiol 2011 Jun;78(3):384-387. [doi: 10.1016/j.ejrad.2010.12.059] [Medline: 21288674]
40. Sung YH, Kwon IG, Ryu E. Blended learning on medication administration for new nurses: integration of e-learning and face-to-face instruction in the classroom. Nurse Educ Today 2008 Nov;28(8):943-952. [doi: 10.1016/j.nedt.2008.05.007] [Medline: 18599162]
41. Woltering V, Herrler A, Spitzer K, Spreckelsen C. Blended learning positively affects students' satisfaction and the role of the tutor in the problem-based learning process: results of a mixed-method evaluation. Adv Health Sci Educ Theory Pract 2009 Dec;14(5):725-738. [doi: 10.1007/s10459-009-9154-6] [Medline: 19184497]
42. Lowe CI, Wright JL, Bearn DR. Computer-aided Learning (CAL): an effective way to teach the Index of Orthodontic Treatment Need (IOTN)? J Orthod 2001 Dec;28(4):307-311. [doi: 10.1093/ortho/28.4.307] [Medline: 11709597]
43. Hilger AE, Hamrick HJ, Denny FW. Computer instruction in learning concepts of streptococcal pharyngitis. Arch Pediatr Adolesc Med 1996 Jun;150(6):629-631. [Medline: 8646314]
44. Ilic D, Hart W, Fiddes P, Misso M, Villanueva E. Adopting a blended learning approach to teaching evidence based medicine: a mixed methods study. BMC Med Educ 2013;13:169 [FREE Full text] [doi: 10.1186/1472-6920-13-169] [Medline: 24341502]
45. Arroyo-Morales M, Cantarero-Villanueva I, Fernández-Lao C, Guirao-Piñeyro M, Castro-Martín E, Díaz-Rodríguez L. A blended learning approach to palpation and ultrasound imaging skills through supplementation of traditional classroom teaching with an e-learning package. Man Ther 2012 Oct;17(5):474-478. [doi: 10.1016/j.math.2012.04.002] [Medline: 22579034]
46. Raupach T, Münscher C, Pukrop T, Anders S, Harendza S. Significant increase in factual knowledge with web-assisted problem-based learning as part of an undergraduate cardio-respiratory curriculum. Adv Health Sci Educ Theory Pract 2010 Aug;15(3):349-356 [FREE Full text] [doi: 10.1007/s10459-009-9201-3] [Medline: 19774475]
47. Carbonaro M, King S, Taylor E, Satzinger F, Snart F, Drummond J. Integration of e-learning technologies in an interprofessional health science course. Med Teach 2008 Feb;30(1):25-33. [doi: 10.1080/01421590701753450] [Medline: 18278648]
48. Pereira J, Palacios M, Collin T, Wedel R, Galloway L, Murray A, et al. The impact of a hybrid online and classroom-based course on palliative care competencies of family medicine residents. Palliat Med 2008 Dec;22(8):929-937. [doi: 10.1177/0269216308094561] [Medline: 18772211]
49. Devitt P, Smith JR, Palmer E. Improved student learning in ophthalmology with computer-aided instruction. Eye (Lond) 2001 Oct;15(Pt 5):635-639. [doi: 10.1038/eye.2001.199] [Medline: 11702976]
50. Kiviniemi MT. Effects of a blended learning approach on student outcomes in a graduate-level public health course. BMC Med Educ 2014;14:47 [FREE Full text] [doi: 10.1186/1472-6920-14-47] [Medline: 24612923]
51. Hsu L, Hsieh S. Effects of a blended learning module on self-reported learning performances in baccalaureate nursing students. J Adv Nurs 2011 Nov;67(11):2435-2444. [doi: 10.1111/j.1365-2648.2011.05684.x] [Medline: 21592188]
52. Kaveevivitchai C, Chuengkriankrai B, Luecha Y, Thanooruk R, Panijpan B, Ruenwongsa P. Enhancing nursing students' skills in vital signs assessment by using multimedia computer-assisted learning with integrated content of anatomy and physiology. Nurse Educ Today 2009 Jan;29(1):65-72. [doi: 10.1016/j.nedt.2008.06.010] [Medline: 18703258]
53. Kumrow DE. Evidence-based strategies of graduate students to achieve success in a hybrid web-based course. J Nurs Educ 2007 Mar;46(3):140-145. [Medline: 17396555]
54. Howerton WB, Enrique PRT, Ludlow JB, Tyndall DA. Interactive computer-assisted instruction vs. lecture format in dental education. J Dent Hyg 2004;78(4):10. [Medline: 16197750]
55. Gadbury-Amyot CC, Singh AH, Overman PR. Teaching with technology: learning outcomes for a combined dental and dental hygiene online hybrid oral histology course. J Dent Educ 2013 Jun;77(6):732-743 [FREE Full text] [Medline: 23740910]
56. Perkins GD, Fullerton JN, Davis-Gomez N, Davies RP, Baldock C, Stevens H, et al. The effect of pre-course e-learning prior to advanced life support training: a randomised controlled trial. Resuscitation 2010 Jul;81(7):877-881. [doi: 10.1016/j.resuscitation.2010.03.019] [Medline: 20409627]
57. Strickland S. The effectiveness of blended learning environments for the delivery of respiratory care education. J Allied Health 2009;38(1):E11-E16. [Medline: 19753407]

58. Gagnon M, Gagnon J, Desmartis M, Njoya M. The impact of blended teaching on knowledge, satisfaction, and self-directed learning in nursing undergraduates: a randomized, controlled trial. Nurs Educ Perspect 2013;34(6):377-382. [Medline: 24475598]
59. Boynton JR, Green TG, Johnson LA, Nainar SMH, Straffon LH. The virtual child: evaluation of an internet-based pediatric behavior management simulation. J Dent Educ 2007 Sep;71(9):1187-1193 [FREE Full text] [Medline: 17761625]
60. Llambí L, Esteves E, Martinez E, Forster T, García S, Miranda N, et al. Teaching tobacco cessation skills to Uruguayan physicians using information and communication technologies. J Contin Educ Health Prof 2011;31(1):43-48. [doi: 10.1002/chp.20100] [Medline: 21425359]
61. Sherman H, Comer L, Putnam L, Freeman H. Blended versus lecture learning: outcomes for staff development. J Nurses Staff Dev 2012 Jul;28(4):186-190. [doi: 10.1097/NND.0b013e31825dfb71] [Medline: 22821022]
62. Gerdprasert S, Pruksacheva T, Panijpan B, Ruenwongsa P. Development of a web-based learning medium on mechanism of labour for nursing students. Nurse Educ Today 2010 Jul;30(5):464-469. [doi: 10.1016/j.nedt.2009.10.007] [Medline: 19931950]
63. Farrell MJ, Rose L. Use of mobile handheld computers in clinical nursing education. J Nurs Educ 2008 Jan;47(1):13-19. [Medline: 18232610]
64. Taradi SK, Taradi M, Radic K, Pokrajac N. Blending problem-based learning with web technology positively impacts student learning outcomes in acid-base physiology. Adv Physiol Educ 2005 Mar;29(1):35-39 [FREE Full text] [doi: 10.1152/advan.00026.2004] [Medline: 15718381]
65. Lehmann R, Thiessen C, Frick B, Bosse HM, Nikendei C, Hoffmann GF, et al. Improving pediatric basic life support performance through blended learning with web-based virtual patients: randomized controlled trial. J Med Internet Res 2015 Jul 02;17(7):e162 [FREE Full text] [doi: 10.2196/jmir.4141] [Medline: 26139388]
66. Succar T, Zebington G, Billson F, Byth K, Barrie S, McCluskey P, et al. The impact of the Virtual Ophthalmology Clinic on medical students' learning: a randomised controlled trial. Eye (Lond) 2013 Oct;27(10):1151-1157. [doi: 10.1038/eye.2013.143] [Medline: 23867718]
67. Wahlgren C, Edelbring S, Fors U, Hindbeck H, Ståhle M. Evaluation of an interactive case simulation system in dermatology and venereology for medical students. BMC Med Educ 2006;6:40 [FREE Full text] [doi: 10.1186/1472-6920-6-40] [Medline: 16907972]
68. Ilic D, Nordin RB, Glasziou P, Tilson JK, Villanueva E. A randomised controlled trial of a blended learning education intervention for teaching evidence-based medicine. BMC Med Educ 2015 Mar 10;15:39 [FREE Full text] [doi: 10.1186/s12909-015-0321-6] [Medline: 25884717]
69. Shimizu I, Nakazawa H, Sato Y, Wolfhagen IHAP, Könings KD. Does blended problem-based learning make Asian medical students active learners?: a prospective comparative study. BMC Med Educ 2019 May 15;19(1):147 [FREE Full text] [doi: 10.1186/s12909-019-1575-1] [Medline: 31092243]
70. Milic NM, Trajkovic GZ, Bukumiric ZM, Cirkovic A, Nikolic IM, Milin JS, et al. Improving education in medical statistics: implementing a blended learning model in the existing curriculum. PLoS One 2016;11(2):e0148882 [FREE Full text] [doi: 10.1371/journal.pone.0148882] [Medline: 26859832]
71. Sadeghi R, Sedaghat MM, Sha Ahmadi F. Comparison of the effect of lecture and blended teaching methods on students' learning and satisfaction. J Adv Med Educ Prof 2014 Oct;2(4):146-150 [FREE Full text] [Medline: 25512938]
72. Kho MHT, Chew KS, Azhar MN, Hamzah ML, Chuah KM, Bustam A, et al. Implementing blended learning in emergency airway management training: a randomized controlled trial. BMC Emerg Med 2018 Jan 15;18(1):1 [FREE Full text] [doi: 10.1186/s12873-018-0152-y] [Medline: 29334925]
73. Jarrett-Thelwell FD, Burke JR, Poirier J, Petrocco-Napuli K. A comparison of student performance and satisfaction between a traditional and integrative approach to teaching an introductory radiology course on the extremities. J Chiropr Educ 2019 Mar;33(1):21-29 [FREE Full text] [doi: 10.7899/JCE-17-26] [Medline: 30444635]
74. Marchalot A, Dureuil B, Veber B, Fellahi J, Hanouz J, Dupont H, et al. Effectiveness of a blended learning course and flipped classroom in first year anaesthesia training. Anaesth Crit Care Pain Med 2018 Oct;37(5):411-415. [doi: 10.1016/j.accpm.2017.10.008] [Medline: 29175318]
75. McCutcheon K, O'Halloran P, Lohan M. Online learning versus blended learning of clinical supervisee skills with pre-registration nursing students: A randomised controlled trial. Int J Nurs Stud 2018 Jun;82:30-39 [FREE Full text] [doi: 10.1016/j.ijnurstu.2018.02.005] [Medline: 29574394]
76. Noll C, von Jan U, Raap U, Albrecht U. Mobile augmented reality as a feature for self-oriented, blended learning in medicine: randomized controlled trial. JMIR Mhealth Uhealth 2017 Sep 14;5(9):e139 [FREE Full text] [doi: 10.2196/mhealth.7943] [Medline: 28912113]
77. Zhan X, Zhang Z, Sun F, Liu Q, Peng W, Zhang H, et al. Effects of improving primary health care workers' knowledge about public health services in rural China: a comparative study of blended learning and pure e-learning. J Med Internet Res 2017 May 01;19(5):e116 [FREE Full text] [doi: 10.2196/jmir.6453] [Medline: 28461286]
78. Stewart A, Inglis G, Jardine L, Koorts P, Davies MW. A randomised controlled trial of blended learning to improve the newborn examination skills of medical students. Arch Dis Child Fetal Neonatal Ed 2013 Mar;98(2):F141-F144. [doi: 10.1136/archdischild-2011-301252] [Medline: 22684155]

79. Kononowicz AA, Krawczyk P, Cebula G, Dembkowska M, Drab E, Frączek B, et al. Effects of introducing a voluntary virtual patient module to a basic life support with an automated external defibrillator course: a randomised trial. BMC Med Educ 2012 Jun 18;12:41 [FREE Full text] [doi: 10.1186/1472-6920-12-41] [Medline: 22709278]
80. Wu J, Tennyson RD, Hsia T. A study of student satisfaction in a blended e-learning system environment. Computers & Education 2010 Aug;55(1):155-164. [doi: 10.1016/j.compedu.2009.12.012]
81. Kintu MJ, Zhu C, Kagambe E. Blended learning effectiveness: the relationship between student characteristics, design features and outcomes. Int J Educ Technol High Educ 2017 Feb 6;14(1):1-20. [doi: 10.1186/s41239-017-0043-4]
82. Garrison D, Kanuka H. Blended learning: Uncovering its transformative potential in higher education. The Internet and Higher Education 2004 Apr;7(2):95-105. [doi: 10.1016/j.iheduc.2004.02.001]
83. Kazu I, Demirkol M. Effect of blended learning environment model on high school students' academic achievement. Turk Online J Educ Technol 2014 Jan 01;13(1):78-87 [FREE Full text]
84. Lim DH, Kim H. Motivation and learner characteristics affecting online learning and learning application. Journal of Educational Technology Systems 2016 Jul 21;31(4):423-439. [doi: 10.2190/0LW0-KE8X-MDYH-X27F]
85. Song L, Singleton ES, Hill JR, Koh MH. Improving online learning: student perceptions of useful and challenging characteristics. The Internet and Higher Education 2004 Jan;7(1):59-70. [doi: 10.1016/j.iheduc.2003.11.003]
86. Posadzki P, Bala MM, Kyaw BM, Semwal M, Divakar U, Koperny M, et al. Offline digital education for postregistration health professions: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Apr 24;21(4):e12968 [FREE Full text] [doi: 10.2196/12968] [Medline: 31017584]
87. Greenhalgh T. Computer assisted learning in undergraduate medical education. BMJ 2001 Jan 6;322(7277):40-44 [FREE Full text] [Medline: 11141156]
88. Adams AM. Pedagogical underpinnings of computer-based learning. J Adv Nurs 2004 Apr;46(1):5-12. [doi: 10.1111/j.1365-2648.2003.02960.x] [Medline: 15030437]
89. Brandt MG, Davies ET. Visual-spatial ability, learning modality and surgical knot tying. Can J Surg 2006 Dec;49(6):412-416 [FREE Full text] [Medline: 17234070]
90. Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GCY, Tudor Car L. Mobile digital education for health professions: systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 12;21(2):e12937 [FREE Full text] [doi: 10.2196/12937] [Medline: 30747711]
91. Tomesko J, Touger-Decker R, Dreker M, Zelig R, Parrott JS. The effectiveness of computer-assisted instruction to teach physical examination to students and trainees in the health sciences professions: a systematic review and meta-analysis. J Med Educ Curric Dev 2017;4:2382120517720428 [FREE Full text] [doi: 10.1177/2382120517720428] [Medline: 29349338]
92. Al-Riyami S, Moles DR, Leeson R, Cunningham SJ. Comparison of the instructional efficacy of an internet-based temporomandibular joint (TMJ) tutorial with a traditional seminar. Br Dent J 2010 Dec 11;209(11):571-576. [doi: 10.1038/sj.bdj.2010.1085] [Medline: 21151070]
93. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Jul 02;21(7):e14676 [FREE Full text] [doi: 10.2196/14676] [Medline: 31267981]
94. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med 2010 Oct;85(10):1589-1602. [doi: 10.1097/ACM.0b013e3181edfe13] [Medline: 20703150]
95. Xiao Z, Wang C, Li L, Tang X, Li N, Li J, et al. Clinical efficacy and safety of aidi injection plus docetaxel-based chemotherapy in advanced nonsmall cell lung cancer: a meta-analysis of 36 randomized controlled trials. Evid Based Complement Alternat Med 2018;2018:7918258 [FREE Full text] [doi: 10.1155/2018/7918258] [Medline: 29991956]
96. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009 Apr;43(4):303-311. [doi: 10.1111/j.1365-2923.2008.03286.x] [Medline: 19335571]
97. Bearman M, Cesnik B, Liddell M. Random comparison of 'virtual patient' models in the context of teaching clinical communication skills. Med Educ 2001 Sep;35(9):824-832. [Medline: 11555219]
98. Foster A, Chaudhary N, Kim T, Waller JL, Wong J, Borish M, et al. Using virtual patients to teach empathy: a randomized controlled study to enhance medical students' empathic communication. Simul Healthc 2016 Jun;11(3):181-189. [doi: 10.1097/SIH.0000000000000142] [Medline: 26841278]
99. Tolsgaard MG, Jepsen RMHG, Rasmussen MB, Kayser L, Fors U, Laursen LC, et al. The effect of constructing versus solving virtual patient cases on transfer of learning: a randomized trial. Perspect Med Educ 2016 Feb;5(1):33-38 [FREE Full text] [doi: 10.1007/s40037-015-0242-4] [Medline: 26754313]
100. Edelbring S, Wahlström R. Dynamics of study strategies and teacher regulation in virtual patient learning activities: a cross sectional survey. BMC Med Educ 2016 Apr 23;16:122 [FREE Full text] [doi: 10.1186/s12909-016-0644-y] [Medline: 27108089]

Edited by G Eysenbach; submitted 04.10.19; peer-reviewed by Y Lecarpentier, M Grasl; comments to author 30.10.19; revised version received 07.11.19; accepted 13.11.19; published 10.08.20.

Please cite as:
Vallée A, Blacher J, Cariou A, Sorbets E
Blended Learning Compared to Traditional Learning in Medical Education: Systematic Review and Meta-Analysis
J Med Internet Res 2020;22(8):e16504
URL: https://www.jmir.org/2020/8/e16504
doi:10.2196/16504
PMID:32773378

Review

Review of Evaluation Metrics Used in Digital and Traditional Tobacco Control Campaigns

Lilian Chan[1], MPH, MBBS; Blythe O'Hara[1], BA, BSW, MAppSci, PhD; Philayrath Phongsavan[1], BA(Hons), MPH, PhD; Adrian Bauman[1], PhD, FAHMS, FAFHM; Becky Freeman[1], BSc, MSc, PhD

Sydney School of Public Health and Charles Perkins Centre, Prevention Research Collaboration, The University of Sydney, Camperdown, Australia

Corresponding Author:
Lilian Chan, MPH, MBBS
Sydney School of Public Health and Charles Perkins Centre
Prevention Research Collaboration
The University of Sydney
Johns Hopkins Dr
Camperdown, NSW 2006
Australia
Phone: 61 2 8627 7554
Email: lilian.chan@sydney.edu.au

Abstract

Background: Mass media campaigns for public health are increasingly using digital media platforms, such as web-based advertising and social media; however, there is a lack of evidence on how to best use these digital platforms for public health campaigns. To generate this evidence, appropriate campaign evaluations are needed, but with the proliferation of digital media–related metrics, there is no clear consensus on which evaluation metrics should be used. Public health campaigns are diverse in nature, so to facilitate analysis, this review has selected tobacco control campaigns as the scope of the study.

Objective: This literature review aimed to examine how tobacco control campaigns that use traditional and digital media platforms have been evaluated.

Methods: Medicine and science databases (Medical Literature Analysis and Retrieval System Online [MEDLINE], EMBASE, PsycINFO, Cumulative Index to Nursing and Allied Health Literature [CINAHL], and Scopus), and a marketing case study database (World Advertising Research Center) were searched for articles published between 2013 and 2018. Two authors established the eligibility criteria and reviewed articles for inclusion. Individual campaigns were identified from the articles, and information on campaigns and their evaluations were supplemented with searches on Google, Google Scholar, and social media platforms. Data about campaign evaluations were tabulated and mapped to a conceptual framework.

Results: In total, 17 campaigns were included in this review, with evaluations reported on by 51 articles, 17 marketing reports, and 4 grey literature reports. Most campaigns were from English-speaking countries, with behavioral change as the primary objective. In the process evaluations, a wide range of metrics were used to assess the reach of digital campaign activities, making comparison between campaigns difficult. Every campaign in the review, except one, reported some type of engagement impact measure, with website visits being the most commonly reported metric (11 of the 17 campaigns). Other commonly reported evaluation measures identified in this review include engagement on social media, changes in attitudes, and number of people contacting smoking cessation services. Of note, only 7 of the 17 campaigns attempted to measure media platform attribution, for example, by asking participants where they recalled seeing the campaign or using unique website tracking codes for ads on different media platforms.

Conclusions: One of the key findings of this review is the numerous and diverse range of measures and metrics used in tobacco control campaign evaluations. To address this issue, we propose principles to guide the selection of digital media–related metrics for campaign evaluations, and also outline a conceptual framework to provide a coherent organization to the diverse range of metrics. Future research is needed to specifically investigate whether engagement metrics are associated with desired campaign outcomes, to determine whether reporting of engagement metrics is meaningful in campaign evaluations.

(J Med Internet Res 2020;22(8):e17432) doi:10.2196/17432

KEYWORDS

mass media; internet; evaluation studies as topic; smoking cessation; public health

Introduction

Background

By 2019, advertising on the internet made up over half of all media spending in 8 countries, including the United Kingdom, China, the United States, and Australia [1]. The growing trend toward digital advertising has extended into public health mass media campaigns, with the majority of these campaigns now using digital media platforms, such as web advertising and social media, in addition to traditional media platforms [2].

Despite the increasing popularity of digital media use, there is a lack of robust evidence on how best to use digital platforms for public health campaigns, including questions around which platforms, or combinations of platforms, are most effective for driving behavioral change [3]. Developing a body of evidence in this area is vital to ensure public health campaigns are effective, that they reach intended audiences, and that there is appropriate investment of resources.

To generate this evidence, appropriate evaluations of campaigns are needed. With the proliferation of digital media platforms, metrics such as likes, engagements, impressions, and click-through rates have become commonplace in evaluations [3-8]. Despite the prevalence of their use, their meaning in public health is not completely understood, and there are currently no clear guidelines on which, if any, of these metrics are relevant for public health campaign evaluations. This situation will continue to become a greater challenge, as the continual emergence of new platforms, such as the recent popularity of Tik Tok (ByteDance) [9], leads to an ever-increasing number of digital evaluation metrics. In addition, the growing number of digital media platforms means that campaigns can use multiple media platforms, creating the additional challenge for practitioners to understand which platform, or combination of platforms, should be used for public health campaigns.

Given varied objectives, strategies, and activities of public health campaigns, this review focuses on campaigns relating only to tobacco control to facilitate comparison. Today, some tobacco control campaigns are among the most advanced public health campaigns in terms of funding, strategy, and evaluation, and have a large underpinning evidence base that describes effective campaigns [10]. Despite this, there is limited evidence on what constitutes effective digital media use in tobacco control campaigns, with the US Center for Disease Control and Prevention's Best Practices for Comprehensive Tobacco Control Programs acknowledging that there is insufficient evidence to make any recommendations on how to best use digital media channels [11]. This gap in knowledge is the background for this review.

Objectives

This paper examines how tobacco control campaigns that use traditional and digital media platforms have been evaluated in the published literature. A better understanding of how to evaluate these campaigns will enable practitioners and researchers to develop greater insight into how to effectively use digital media platforms for tobacco control campaigns, and more widely, for public health campaigns.

Methods

Data Collection

Data were collected through 3 search approaches: (1) in medicine and science journal databases, (2) in a marketing case studies database, and (3) through internet searches for grey literature, campaign websites, and social media sites.

For medicine and science journals, a search was conducted using the Medical Literature Analysis and Retrieval System Online (MEDLINE) via OvidSP (Wolters Kluwer Health), EMBASE via OvidSP, PsycINFO via OvidSP, and Cumulative Index to Nursing and Allied Health Literature (CINAHL; EBSCO) and Scopus (Elsevier). The search strategy used the following terms: (smok*.mp OR tobacco/) AND (campaign.mp OR mass media.mp) AND (digital.mp OR online.mp). Search results were limited to articles in English and published in the last 5 years (2013-2018). This timeframe was selected to ensure the relevance of this review because of the fast-changing nature of digital platforms and their usage patterns.

The review was supplemented with a search of the marketing database WARC (World Advertising Research Center). For this search, the keyword terms were smoking OR tobacco, with results limited to the last 5 years, within the *Non-profit, public sector, and education* database category.

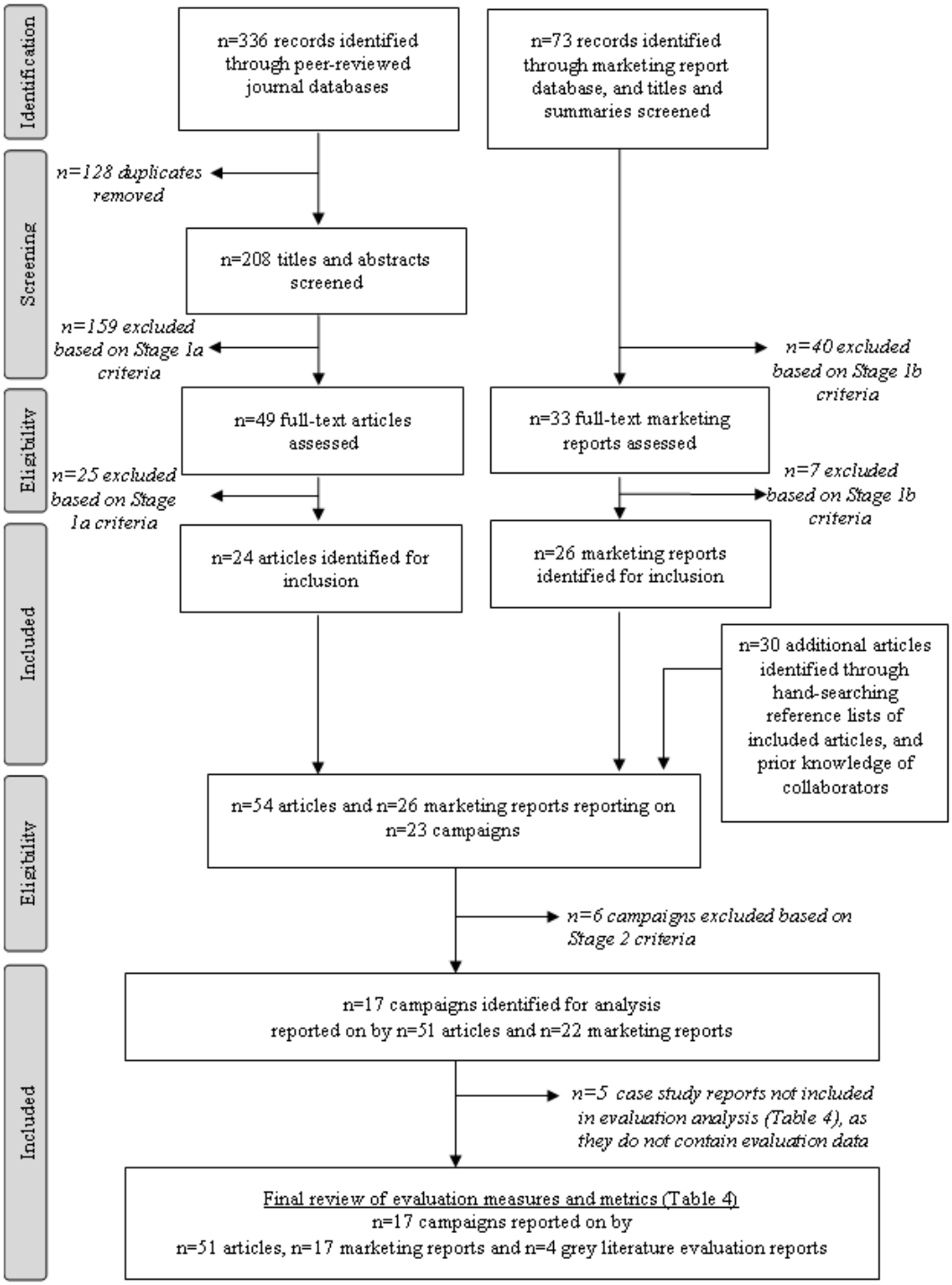
Subsequently, the reference lists of included articles and systematic reviews identified in the literature search were reviewed for additional relevant references.

The first stage of this review involved 2 authors (LC and BH) independently reviewing the same subset (25%) of all identified database search results to establish and test the eligibility criteria (see Multimedia Appendix 1). One author (LC) then reviewed the remaining search results against the criteria to identify literature that warranted full-text review. The same 2 authors then independently reviewed all full-text articles against the eligibility criteria.

Campaign Identification

The second stage of the literature review involved the identification of individual campaigns from the included articles (see Figure 1). Each identified campaign was searched on both Google Scholar and Google for evaluation reports, press releases, or other evaluation materials. Campaign websites and social media pages were also searched and examined. Based on these multiple sources, campaigns were assessed for inclusion in the review against the eligibility criteria (see Multimedia Appendix 1). One author (LC) conducted the additional searches and performed the initial assessment against the eligibility criteria. Two authors (BH and BF) independently reviewed any unresolved campaigns.

Figure 1. Flowchart of search strategy and campaign selection.



Data Extraction

All articles identified throughout the data collection process were recorded using Endnote (Version X8, Clarivate Analytics). Information from multiple sources was then tabulated by campaign to provide a complete picture of the evaluation measures and methods used by each campaign. To provide context for the evaluations, data on each campaign's objectives, target audience, and details of media usage (both paid and unpaid) were also collected.

Data Analysis

To summarize evaluation measures used by different campaigns, data were mapped to a conceptual framework (Table 1). This framework includes evaluation metrics that were commonly reported for the digital components of campaigns, alongside measures that have conventionally been used in campaign

evaluations [12,13]. The conceptual framework is based on the different levels of evaluation—process, impact, and outcome. Building on other campaign evaluation models [12,13], this framework incorporates several levels of impact evaluation: measures of campaign awareness, engagement, priming steps, and trialing behaviors (Table 1). Actions within each level of evaluation are not necessarily equal in value to the overall campaign outcome but are grouped together based the nature of the action. Information on whether and how campaigns measured which media platforms contributed to outcomes was also collected. Formative, precampaign, and message development evaluations were not included in this review.

Table 1. Conceptual framework of campaign evaluation metrics and measures

Process evaluation	Impact evaluation				Outcome evaluation
	Awareness	Proximal impact I: Engagement	Proximal impact II: Priming steps	Distal impact: Trialing behaviors	
Delivery of campaign	Seen the campaign and perception of the campaign	Showing interest in the campaign or message by taking an action	Priming steps of behavioral change	Initial trialing behaviors and antecedents of behaviors	Desired behavioral change
Delivery of • Television ads (Target Audience Rating Points [TARPs] or Gross Rating Points [GRPs]) • *Digital video ads (digital GRPs or impressions or video views)*[a] • *Digital banner ads (impressions or exposures)* • Other	• Campaign recall (including frequency) • *Media channel attribution (where campaign was viewed)b* • *Campaign response (eg, relevance, perceived effectiveness, believability)b*	• *Campaign website visits* • *Engagement on social media (eg, likes, comments, shares, follows)* • *Click through rates (on digital ads or social media posts)* • *Information-seeking action on the internet (web search)* • *Other action (eg, download mobile app, sign up to campaign)*	• Knowledge and beliefs • Attitudes: about smoking, tobacco industry, etc • Attitude: intention to quit • Information-seeking action offline (spoke with health care provider)	• Contact smoking cessation service or registrations to service • Quit attempts	• Sustained quit attempts • Population smoking prevalence rates • (For nonsmokers): Conversation with family or friend about smoking cessation

[a]All italics indicate metrics and measures that relate to digital media platforms.

[b]In this review, media channel attribution and campaign responses were measured through both digital platform evaluation methods and traditional evaluation methods.

Results

Study Selection

The medicine or science database searches identified 336 articles. After removal of duplicates, 208 articles were screened. This identified 49 articles for full-text review, and subsequently 24 articles were included in this review. The marketing database search identified 73 reports, and after review, 26 were included. From hand-searching references of the included articles, 30 additional articles were identified for this review (see Figure 1).

Campaign Selection

After further searches for more information about the identified campaigns in grey literature reports, campaign websites and social media pages, 6 campaigns were excluded for the following reasons: insufficient information about the campaign, insufficient information about the digital aspects of the campaign, lack of evaluation data, campaign related to e-cigarettes, and intervention assessed as not primarily a campaign. As a result, 17 campaigns were included in this review, reported on by 51 peer-reviewed articles and 22 marketing reports. However, 5 of the marketing reports provided contextual campaign information but did not contain unique evaluation data. Therefore, the analysis of evaluations of the 17 campaigns was based on 51 peer-reviewed articles, 17 marketing reports, and 4 grey literature evaluation reports.

Of the 17 identified campaigns, 7 were only located in marketing reports and grey literature, highlighting the benefit of using these additional sources of information for this review. Of the 51 peer-reviewed articles included in this review, 29 reported on the *Tips from Former Smokers* campaign, 7 reported on the *Truth FinishIt* campaign, and 7 reported on *The Real Cost* campaign.

Campaign Characteristics

Most campaigns were from high-income, English-speaking countries, with 6 from the United States, 4 from Canada, 3 from Australia, and 2 from the United Kingdom. In all, 13 of the 17 campaigns had a primary objective of behavioral change, 2 were

awareness-raising campaigns, and 2 were campaigns aimed at changing social norms.

Campaign Evaluation Measures

The types of evaluation measures used for campaigns are summarized in Tables 2 and 3.

Table 2. Reported evaluation measures in behavioral change campaigns.

Campaign	Process	Awareness	Proximal impact: engagement	Proximal impact: priming steps	Distal impact	Outcome
Tips from Former Smokers	✓	✓	✓	✓	✓	✓
Stop before the suffering starts	✓	✓	✓	✓	✓	__[a]
Stoptober	—	✓	✓	✓	✓	✓
The Real Cost	✓	✓	✓	✓	—	✓
Be a Failure	✓	✓	✓	✓	—	—
16 cancers	—	✓	✓	—	✓	—
SmokeFree Teen	✓	—	✓	—	✓	—
Fingerband campaign	✓	✓	✓	—	—	✓
Break it Off	—	—	✓	✓	✓	✓
Keep Trying	—	—	✓	—	✓	—
No judgments. Just help	—	—	✓	—	✓	✓
Personal Testimonies	—	—	✓	—	✓	—
The Smoking Kid	✓	—	—	—	✓	—

[a]No data was available on these evaluation measures.

Table 3. Reported evaluation measures in awareness raising and social norm change campaigns.

Campaign	Process	Awareness	Proximal impact: engagement	Proximal impact: priming steps	Distal impact or outcomes
Truth FinishIt	✓	✓	✓	✓	✓
The Facts Now	✓	__[a]	✓	—	✓
Take it right outside	—	✓	✓	✓	✓
Quit the Denial	✓	✓	✓	✓	—

[a]No data was available on these evaluation measures.

Process Evaluation Measures

The conceptual framework as described in Table 1 emphasizes quantitative measures for process evaluations of campaigns. Of the 10 campaigns in this review that had a television advertising component, 4 reported the number of target audience rating points (TARPs) or gross rating points (GRPs) [14-23], which are both measures of reach, describing the estimated percentage of the population that viewed the ad.

The majority of campaigns (8/10) using digital videos reported a metric about the reach of the digital video [8,15,19,24-32]. The reach of digital videos was reported using a variety of metrics, including digital TARPs (the equivalent of TARPs for content delivered on a digital platform) [33], impressions (the number of times the content was delivered) [33], exposures (opportunities for the content to be seen [34]), or video views.

The reach of web banner ads was reported as impressions or exposures by 2 campaigns [8,24], and digital impressions by 1 campaign, but it was not clear whether this was for static banner ads and/or digital video ads (*Truth FinishIt*) [35]. One campaign reported measuring banner ad reach but did not report the result (*Be a Failure*) [36].

Campaign Awareness Measures

In all, 7 campaigns evaluated whether people recalled (ie, without prompting with campaign material) or recognized (after being shown campaign material) the campaign, which was primarily measured through sampled surveys or interviews [14,15,19,23,26,35,37-54]. A total of 7 campaigns reported evaluations on the audience's response to the campaign, such as perceived effectiveness of the campaign or emotional reaction to the campaign. This was evaluated through surveys or interviews or content analysis of social media comments [14,15,25,30,36,51,53-58].

Proximal Impact Evaluation Measures I: Engagement

Proximal impact measures of engagement, such as the number of visits to a website or ad click-through rates (the percentage of times an ad is clicked) [33], represent intermediary steps

between exposure to a campaign and the desired outcomes of a campaign (see Table 1).

All but one campaign in this review reported at least one proximal impact measure of engagement. Of all the evaluation measures identified in this review, campaign website visits was the most commonly reported measure (11/17 campaigns) [8,20,24,27,36,37,59-66]. Engagement on social media—broadly encompassing numbers of likes, shares, comments, or followers on any social media platform—was reported for 8 campaigns [8,25-27,29,30,32,35,52,60]. Two of these campaigns used aggregated metrics of engagement (social media engagement rate in *The Real Cost*, and social media conversation in *Quit the Denial*) [26,29].

The number of times an ad was clicked or the click-through rate were only reported in 2 of the 11 campaigns that used web static banner ads (*SmokeFree Teen* and *Tips from Former Smokers*) [8,24].

In all, 5 campaigns reported on whether people exposed to the campaign took an intermediary action of seeking more information about the issue on the internet [14,24,26,36,67,68]. This was either measured through survey questions or through analyzing campaign keyword search trends on search engines (*Tips from Former Smokers* and *Stoptober*) [67,68].

A total of 5 campaigns used other digital media–based measures as part of the evaluation of proximal impact. These included measuring mobile phone app downloads [8,14,60,63], sign ups to the campaign [32], views of email marketing messages [69], and campaign resource downloads [63].

Proximal Impact Evaluation Measures II: Priming Steps

In all, 3 of the 17 campaigns measured knowledge-related outcomes, such as about the health-related harms of smoking or of second-hand smoke [26,40,46,50,70]. A total of 8 campaigns measured attitudes related to smoking, the tobacco industry, and the quitting process [14,23,26,36,39,40,42,43,45,46,51-54,64,67,70,71]. Overall, 8 campaigns specifically measured attitudes around intention to quit smoking [14,21,26,36,39,45,47,48,50,53,60,63]. Changes in knowledge and attitudes were measured by surveys or interviews. In addition, 3 campaigns identified whether people had spoken to a health care professional for more information on quitting [14,26,36].

Distal Impact Evaluation Measures: Trialing Behaviors

The number of people contacting smoking cessation services was reported in 9 of the 13 behavioral change campaigns [8,14,18,22,28,37,59,61,65-67,72,73]. In all, 6 campaigns evaluated the number of people making quit attempts [14,17,21,40,44,46,47,50,60,63,67,72,74,75].

Outcome Evaluation Measures

Finally, 4 campaigns evaluated the number of people with sustained quit attempts [44,47,60,63,72,76]. *The Real Cost*, which aimed to reduce smoking initiation rates in young people, evaluated smoking initiation behavior [41]. *Tips from Former Smokers*, which had nonsmokers as a secondary target audience, also measured the number of nonsmokers who had initiated conversations about smoking cessation with friends or family [44,50]. These outcomes were all measured by surveys or interviews. In addition, 2 campaigns (*Fingerband Campaign* and *The Facts Now*) used population smoking prevalence rates [25,27], and 1 campaign (*Stoptober)* measured cigarette sale volumes as part of the outcome evaluation [67].

Media Platform Attribution

In all, 7 campaigns attempted to measure media platform attribution, that is, where the audience was exposed to the campaign [8,14,19,35,37,38,40,44,59]. A total of 4 campaigns used surveys or interviews to ask participants where they recalled seeing the campaign (*Stop before the suffering starts*, *Tips from Former Smokers*, *Take it right outside*, and *Truth FinishIt*) [14,19,35,40,44], 2 campaigns used correlations between timings of campaign outcome events with waves of the campaign that used different media formats (*16 Cancers* and *Personal Testimonies*) [37,59], and 1 campaign used unique website tracking codes for ads shown on different media formats (*SmokeFree Teen*) [8].

Discussion

So Many Metrics, Which Ones to Use?

This review found that there is a wide range of metrics used in tobacco control campaign evaluations, as a consequence of the diversity of media platforms and activities employed by campaigns (see Multimedia Appendix 2 [5,8,15-32,35-63,65-85]). While this gives the impression that there is a lot of information about how a campaign performed, in reality the large number of metrics makes it difficult to meaningfully interpret the reported numbers. For process evaluations, there was a gap between evaluations of traditional media use, such as television ads which used the standardized metrics of GRPs or TARPs, compared with digital media platforms which used a variety of metrics including reach, impressions, exposures, video views, and digital GRPs. The diversity in metrics is partially because of the fragmented media landscape, with each digital media platform having its own reporting system. As all the metrics refer to slightly different measures, it makes comparisons between campaigns difficult. In addition, these raw reach metrics on social media may not reflect a broad generalized reach, as one of the criticisms of organic social media activity is that it perpetuates echo chambers, where messages are often only shared between like-minded individuals. This is less of an issue when campaigns use paid social media strategies, where they can choose the target audience of the campaign ads based on demographics, stated interests, and previous online behavior.

Another group of metrics identified in this review were engagement metrics, which result from digital media activities, and were not present in traditional broadcast media. Examples of these metrics included likes, comments, and retweets. The sheer number of these engagement metrics is overwhelming, and it is challenging to know which are meaningful [86,87]. An additional type of metric identified in this category are metrics which are amalgamations of other metrics, such as social media engagement and social conversation. These have usually been created by advertising companies, and the calculation of these

metrics is usually not transparently described. Finally, digital metrics are usually provided by the platforms themselves, which raises a number of issues. First, the platforms are constantly changing their reporting systems. For example, in 2019 Facebook and Instagram began hiding the number of likes publicly displayed [88,89]. Second, the metrics are not open to independent scrutiny as the platforms are not transparent in how the metrics are calculated. For example, Facebook has previously been reported to have inflated its video view metrics [90]. With these factors in play, campaign practitioners are faced with the great challenge of deciding which metrics to use.

There are currently moves to try to create more uniform digital metrics across the board [91-93]; however, this is a complex undertaking and it is unlikely that a standardized system will be developed in the near future. In the meantime, a published glossary explaining commonly used metrics could provide practitioners and evaluators with a greater understanding of the specific definitions of metrics. In addition, when practitioners and evaluators select metrics, they should be guided by certain principles, as opposed to overloading the reader with numbers that may or may not have relevance to the evaluation. Principles to guide the use of metrics include the following:

1. Metrics should be consistent with the objectives of the campaign [87,94]. For example, reach (the number of people who have seen a campaign) would be appropriate for awareness-raising campaigns that aim to reach as many people as possible, whereas impressions (the number of times the campaign has been shown to the target audience) could be more relevant for behavioral change campaigns that aim to communicate a message many times to a targeted audience.
2. Reported metrics should be the simplest metric available for reporting the intended concept, that is, the metric understood by most people. While complex metrics may help practitioners understand how campaigns are performing at the time, they are usually not widely understood. Furthermore, combined metrics, such as "the campaign produced XXX impressions in total," should be avoided, as they are ambiguous about how the number is calculated across different media.

Contextualizing Evaluation Metrics Through the Conceptual Framework

The conceptual framework in Table 1 provides a starting point in organizing the range of metrics identified in this review. The framework is based on an established program evaluation framework, and for the purposes of planning and evaluating campaigns, provides a structured approach to grouping the metrics. In reality, the flow of events relating to the campaign-desired outcomes may not be linear as depicted in this framework. In the public health literature, several approaches have been used to organize social media metrics [93,95-97]; however, they focus on social media metrics alone, without demonstrating how the social media metrics fit with other digital media measures or other mass media evaluation measures.

Through the use of this conceptual framework to review the range of metrics, we identified strengths and gaps in the evaluations in this review. A large proportion of campaigns reported proximal impact engagement measures, such as website visits, whereas a smaller proportion evaluated proximal impact priming step measures of health-related knowledge and attitudes. The review also identified that marketing reports generally focused more on process evaluation measures and proximal impact engagement measures, whereas peer-reviewed articles focused more on priming step measures. This distinction has practical implications, as campaigns with smaller evaluation budgets often rely on marketing reports to evaluate the effectiveness of a campaign. Conversely, researchers may only look at peer-reviewed articles to identify best practice in campaign development. As all levels of evaluation are of value, it is important that the full spectrum of evaluation measures is reported to understand the effectiveness of a campaign.

Many mass media campaigns are based on behavioral change theories that have priming steps of changes in knowledge, attitudes, or beliefs as intermediary stages before the behavioral change outcome [15,98]. This conceptual framework demonstrates that there is a gap in understanding of whether there is any relationship between proximal impact engagement measures (such as Facebook likes) and proximal impact priming steps of changes, or other impact or outcome measures. Social media is inherently performative, with the user's social network serving as an audience that observes what content users interact with and share. Motivations for engaging may or may not be linked to processing of campaign messaging. For example, it is possible that content that is highly engaging (eg, humorous or controversial content) does not drive behavioral change, that the desired behavioral change is not personally relevant to advocates who are keen to engage and promote the campaign (eg, ex-smokers), or that people do not engage (by liking, sharing, or commenting) with hard-hitting content that does drive behavioral change, as they may not want their peers to see their engagement with this type of content. Despite looking for indication of a relationship between engagement measures and priming step measures in this review, none of the included campaigns provided data that could allow for the analysis of correlations between these two types of measures. To understand whether engagement metrics are meaningful, future research studies need to specifically design campaign evaluations that look at whether people who undertake digital engagement actions are more or less likely to have changes in knowledge or attitudes, or even make the desired behavioral change [99]. It is only by gaining a greater understanding of the relationship of engagement measures with other evaluation measures that we know whether reporting engagement measures is at all meaningful [99,100].

Measuring Media Platform Attribution

One of the major challenges facing practitioners is knowing where to invest resources given the diverse media landscape. The number of platforms is overwhelming, and without evidence of which are more useful at achieving campaign objectives, decisions are sometimes made based on opinions or trends. Therefore, this review examined whether campaign evaluations measured attribution, that is, how activity on each media platform used by the campaign contributed to the campaign's outcomes. Despite this being important information, only a low

proportion of campaigns (7/17) measured attribution. The methods used to measure attribution included survey self-report, using unique website tracking codes for different media format ads, and using an ecological study approach of correlating exposure of different media use combinations with reported campaign awareness and outcomes.

The majority of mass media campaigns use more than one media platform, as reflected in the campaigns included in this review. Previous research has shown that advertising campaigns on multiple platforms produces higher return-on-investment, and campaigns in sectors that are higher-involvement, such as pharmaceuticals, benefit most from synergistic campaigns using both traditional and digital media [101]. Therefore, while the trend toward multiplatform campaigns is clear, there is a great deal of uncertainty on how to accurately measure attribution in cross-platform marketing campaigns [102-104]. This is an even greater challenge in public health campaigns in comparison to marketing campaigns, as the final outcome to determine return-on-investment is not a purchase, but rather an attitudinal or behavioral change.

In all, 4 of the campaigns in this review used surveys or interviews to determine where people had encountered the campaign. However, this method has widely been found to be inaccurate, particularly where different media interact with one another or are viewed at the same time, making it difficult for people to recall where they encountered the campaign [105]. The study by Pettigrew et al [38] identified that people would often attribute their encounter with a campaign to television, even if this was unlikely to be the case. One campaign in this review (*SmokeFree Teen*) used unique website tracking codes on different media format ads to identify attribution. While this has the benefit of being objective, ad click-throughs underestimate the true impact of campaigns. Ad click-through rates have been steadily dropping over time to an average of 0.1% and have been shown not to have any relationship with ad effectiveness [86]. This may be because people instead search for the campaign on a search engine or manually type in a website address at a later time, rather than clicking on an ad at the time of viewing [24]. In addition, using ad click-throughs to measure attribution only captures the most recent encounter that an individual has with the campaign, not taking into account that earlier encounters with the campaign could have influenced their decision to click on the ad. Other methods of measuring attribution include passive systems of tracking exposure to campaigns, such as household meters to record when the TV is on or computer meters that monitor what websites are visited [106]. These methods are used by market research companies for population samples but were not used by any of the campaign evaluations in this review and are not widely used in public health campaigns as they are expensive to implement.

Given the absence of practical methods for campaign evaluators to accurately measure attribution for individual campaigns, there needs to be guidance provided to practitioners on what are generally the most effective combinations of media use. To develop such best practice guidelines, more studies examining the synergistic effects of different combinations of media platforms for public health mass media campaigns are required. The study design used by Allom et al [37] provides a good approach to developing a stronger understanding of the effectiveness of different combinations of media. By testing individual and combinations of media platform use at different times (such as TV only, TV and digital video, and web display and digital video) and then measuring campaign awareness and campaign-related events (website visits, calls to Quitline, registrations to quit program), the study provides an understanding of which combinations are more effective. This approach captures the synergistic effect of multiple media platforms, rather than attempting to simplify measurement to the first encounter with a campaign (eg, asking in a survey, “Where did you first see the campaign?”) or the last touchpoint with a campaign (eg, tracking click-throughs to a quit website). Further research building on this study would help generate evidence for best practice in cross-platform tobacco control campaigns. This could include replicating the study design with another campaign to validate findings and developing it further by asking about priming steps (eg, attitudes toward smoking) and/or trialing behaviors (eg, quit attempts) in addition to campaign awareness. Furthermore, future studies could explore the effect of varying the order of campaign exposure on different platforms, as it has been shown in advertising campaigns that TV first, then followed by digital, has a much larger synergistic effect than vice versa [101].

Strengths and Limitations

One of the key strengths of this review is the use of peer-reviewed literature, marketing reports, grey literature, campaign websites, and social media sites to collect data for the campaigns. The triangulation of data provides a more comprehensive and practical view of how campaigns are currently evaluated.

This review included a wide range of campaigns in terms of scale, making comparison between campaigns difficult. However, the challenges in campaign evaluation identified in this review are common to all health-related campaigns, regardless of size and resourcing. The inclusion of English-only articles and the high representation of campaigns from English-speaking countries may limit the generalizability of this review’s findings and miss potential advances in non-English speaking countries. In addition, the large number of evaluation studies emanating from one campaign *(Tips from Former Smokers*) may also unevenly influence the findings of this review. The exclusion of campaigns about the use of e-cigarettes and waterpipe smoking is another limitation of this review, particularly as these forms of tobacco use are increasing in many populations, and campaigns in these areas may contain advances in the evaluation of digital media. Another limitation of this review is that a large proportion of articles were identified through hand-searching reference lists of included articles. This highlights the complexities in defining appropriate keywords for searching in this area and also supports the value of using this snowball method to ensure the majority of relevant literature is captured. Of note, specific social media–focused keywords were not included in the search strategy; however, many of the campaigns identified in this review use various social media platforms, suggesting that the overall approach has captured the main forms of social media use by mass media campaigns. In addition, future reviews could benefit from using PubMed

searches to ensure newer journals not yet indexed by MEDLINE are included as well. The fragmented amount of information publicly available for some of the included campaigns is also a limitation of this review. Contacting organizations responsible for the campaign could provide more information; however, another review study found this method did not yield much additional information [107].

Conclusions

This review examined how recent tobacco control campaigns that used traditional and digital media platforms were evaluated. It found that in today's fragmented and rapidly evolving media environment, a wide and diverse range of measures and metrics were used in campaign evaluations, particularly for campaign activities relating to digital media use. Purposeful selection of metrics, and utilization of a conceptual framework can help practitioners and researchers make sense of the multitude of metrics and conduct evaluations that further our understanding of how best to use traditional and digital media to communicate health messages to target audiences.

Acknowledgments

LC received support for this research through a PhD scholarship under the Prevention Research Support Program, funded by the New South Wales Ministry of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Eligibility criteria for literature review.
[PDF File (Adobe PDF File), 813 KB - jmir_v22i8e17432_app1.pdf]

Multimedia Appendix 2
Tobacco control campaigns including a digital media component and their evaluation methods.
[PDF File (Adobe PDF File), 252 KB - jmir_v22i8e17432_app2.pdf]

References

1. McDonald J. Internet is Set to Account for Half of Global Media Spend in 2020. WARC. 2019 May. URL: https://www.warc.com/paywall/content/article/warc-datapoints/internet-is-set-to-account-for-half-of-global-media-spend-in-2020/126624 [accessed 2020-02-02]
2. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014 Mar 14;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
3. Burke-Garcia A, Scally G. Trending now: future directions in digital media for the public health sector. J Public Health (Oxf) 2014 Dec;36(4):527-534. [doi: 10.1093/pubmed/fdt125] [Medline: 24468840]
4. Kite J, Grunseit A, Li V, Vineburg J, Berton N, Bauman A, et al. Generating engagement on the make healthy normal campaign Facebook page: analysis of Facebook analytics. JMIR Public Health Surveill 2019 Jan 14;5(1):e11132 [FREE Full text] [doi: 10.2196/11132] [Medline: 31344679]
5. Kornfield R, Smith KC, Szczypka G, Vera L, Emery S. Earned media and public engagement with CDC's 'Tips from Former Smokers' campaign: an analysis of online news and blog coverage. J Med Internet Res 2015 Jan 20;17(1):e12 [FREE Full text] [doi: 10.2196/jmir.3645] [Medline: 25604520]
6. Carlson S, Dey A, Beard F. An evaluation of the 2016 influenza vaccination in pregnancy campaign in NSW, Australia. Public Health Res Pract 2020 Mar 10;30(1):pii: 29121908 [FREE Full text] [doi: 10.17061/phrp29121908] [Medline: 32152616]
7. Serrano WC, Chren M, Resneck JS, Aji NN, Pagoto S, Linos E. Online advertising for cancer prevention: Google ads and tanning beds. JAMA Dermatol 2016 Jan;152(1):101-102. [doi: 10.1001/jamadermatol.2015.3998] [Medline: 26444220]
8. Sanders A, Robinson C, Taylor SC, Post SD, Goldfarb J, Shi R, et al. Using a media campaign to increase engagement with a mobile-based youth smoking cessation program. Am J Health Promot 2018 Jun;32(5):1273-1279. [doi: 10.1177/0890117117728608] [Medline: 28925292]
9. Brownsell A. Trend Snapshot: TikTok. WARC. 2019 May. URL: https://www.warc.com/content/paywall/article/warc-trends/trend_snapshot_tiktok/126582 [accessed 2019-11-01]
10. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. Lancet 2010 Oct 9;376(9748):1261-1271 [FREE Full text] [doi: 10.1016/S0140-6736(10)60809-4] [Medline: 20933263]

11. Best Practices for Comprehensive Tobacco Control Programs—2014. Centers for Disease Control and Prevention. Atlanta, USA: National Center for Chronic Disease Prevention and Health Promotion; 2014. URL: https://www.cdc.gov/tobacco/stateandcommunity/best_practices/index.htm [accessed 2018-11-27]
12. Cavill N, Bauman A. Changing the way people think about health-enhancing physical activity: do mass media campaigns have a role? J Sports Sci 2004 Aug;22(8):771-790. [doi: 10.1080/02640410410001712467] [Medline: 15370487]
13. Bauman A, Smith BJ, Maibach EW, Reger-Nash B. Evaluation of mass media campaigns for physical activity. Eval Program Plann 2006 Aug;29(3):312-322. [doi: 10.1016/j.evalprogplan.2005.12.004]
14. Myers P, Blackmore D. The National Partnership Agreement on Preventive Health Tobacco Social Marketing Campaign: Evaluation Report Waves 5&6. North Melbourne: The Social Research Centre; 2013.
15. Duke JC, Alexander TN, Zhao X, Delahanty JC, Allen JA, MacMonegle AJ, et al. Youth's awareness of and reactions to the real cost national tobacco public education campaign. PLoS One 2015;10(12):e0144827 [FREE Full text] [doi: 10.1371/journal.pone.0144827] [Medline: 26679504]
16. Duke JC, Farrelly MC, Alexander TN, MacMonegle AJ, Zhao X, Allen JA, et al. Effect of a national tobacco public education campaign on youth's risk perceptions and beliefs about smoking. Am J Health Promot 2018 Jun;32(5):1248-1256. [doi: 10.1177/0890117117720745] [Medline: 28759999]
17. Zhang L, Vickerman K, Malarcher A, Mowery P. Intermediate cessation outcomes among quitline callers during a national tobacco education campaign. Nicotine Tob Res 2014 Nov;16(11):1478-1486. [doi: 10.1093/ntr/ntu105] [Medline: 25006045]
18. Davis KC, Alexander RL, Shafer P, Mann N, Malarcher A, Zhang L. The dose-response relationship between tobacco education advertising and calls to Quitlines in the United States, March-June, 2012. Prev Chronic Dis 2015 Nov 5;12:E191 [FREE Full text] [doi: 10.5888/pcd12.150157] [Medline: 26542143]
19. Davis KC, Shafer PR, Rodes R, Kim A, Hansen H, Patel D, et al. Does digital video advertising increase population-level reach of multimedia campaigns? Evidence from the 2013 tips from former smokers campaign. J Med Internet Res 2016 Sep 14;18(9):e235 [FREE Full text] [doi: 10.2196/jmir.5683] [Medline: 27627853]
20. Shafer PR, Davis KC, Patel D, Rodes R, Beistle D. Association between media dose, ad tagging, and changes in web traffic for a national tobacco education campaign: a market-level longitudinal study. J Med Internet Res 2016 Feb 17;18(2):e39 [FREE Full text] [doi: 10.2196/jmir.5343] [Medline: 26887959]
21. Davis KC, Patel D, Shafer P, Duke J, Glover-Kudon R, Ridgeway W, et al. Association between media doses of the tips from former smokers campaign and cessation behaviors and intentions to quit among cigarette smokers, 2012-2015. Health Educ Behav 2018 Feb;45(1):52-60 [FREE Full text] [doi: 10.1177/1090198117709316] [Medline: 28497703]
22. Zhang L, Babb S, Johns M, Mann N, Thompson J, Shaikh A, et al. Impact of US antismoking tv ads on Spanish-language Quitline calls. Am J Prev Med 2018 Oct;55(4):480-487. [doi: 10.1016/j.amepre.2018.05.025] [Medline: 30241616]
23. Vallone D, Cantrell J, Bennett M, Smith A, Rath JM, Xiao H, et al. Evidence of the impact of the truth FinishIt campaign. Nicotine Tob Res 2018 Apr 2;20(5):543-551. [doi: 10.1093/ntr/ntx119] [Medline: 28575421]
24. Kim A, Hansen H, Duke J, Davis K, Alexander R, Rowland A, et al. Does digital ad exposure influence information-seeking behavior online? Evidence from the 2012 tips from former smokers national tobacco prevention campaign. J Med Internet Res 2016 Mar 16;18(3):e64 [FREE Full text] [doi: 10.2196/jmir.4299] [Medline: 26983849]
25. Ministry of Health & Welfare. Ministry of Health & Welfare: The Fingerband Campaign. WARC. 2016. URL: https://www.warc.com/content/paywall/article/sabre-awards/ministry_of_health_welfare_the_fingerband_campaign/120087 [accessed 2018-10-01]
26. Ontario Ministry of Health & Long-Term Care. Ontario Ministry of Health and Long-Term Care: Social smoking Campaign. WARC. 2014. URL: https://www.warc.com/content/paywall/article/ontario_ministry_of_health_and_longterm_care_social_smoking_campaign/101945 [accessed 2018-10-01]
27. Tobacco Free Florida: Auctioneer. WARC. 2016. URL: https://www.warc.com/content/paywall/article/warc-cases/tobacco_free_florida_auctioneer/110096 [accessed 2018-10-01]
28. Thai Health Promotion Foundation. THPF: The Smoking Kid - A Personal Message to the Smokers. WARC. 2013. URL: https://www.warc.com/content/paywall/article/thpf_the_smoking_kid__a_personal_message_to_the_smokers/100511 [accessed 2018-10-01]
29. Food and Drug Administration. Food and Drug Administration: Little Lungs. WARC. 2017. URL: https://www.warc.com/content/paywall/article/warc-cases/food-and-drug-administration-little-lungs/112541 [accessed 2018-10-01]
30. Chung JE. Antismoking campaign videos on YouTube and audience response: Application of social media assessment metrics. Comput Hum Behav 2015 Oct;51:114-121. [doi: 10.1016/j.chb.2015.04.061]
31. truth: 'Finish It'. WARC. 2016. URL: https://www.warc.com/content/paywall/article/arfogilvy/truth_finish_it/106876 [accessed 2018-10-01]
32. Truth Initiative: #StopProfiling. WARC. 2018. URL: https://www.warc.com/content/paywall/article/Truth_Initiative_StopProfiling_/120965 [accessed 2018-10-01]
33. CIMM Lexicon 3.0. Coalition for Innovative Media Measurement. 2016. URL: http://cimm-us.org/initiatives-2/cimm-lexicon/ [accessed 2020-01-01]
34. Marketing Dictionary. Monash University. URL: https://www.monash.edu/business/marketing/marketing-dictionary [accessed 2020-01-20]

35. Hair E, Pitzer L, Bennett M, Halenar M, Rath J, Cantrell J, et al. Harnessing Youth and Young Adult Culture: Improving the Reach and Engagement of the truth Campaign. J Health Commun 2017 Jul;22(7):568-575. [doi: 10.1080/10810730.2017.1325420] [Medline: 28548610]
36. Ministry of Health and Long-Term Care. Ministry of Health and Long-Term Care: Be a Failure. WARC. 2018. URL: https://www.warc.com/content/paywall/article/cca/ministry_of_health_and_longterm_care_be_a_failure/118518 [accessed 2018-10-01]
37. Allom V, Jongenelis M, Slevin T, Keightley S, Phillips F, Beasley S, et al. Comparing the cost-effectiveness of campaigns delivered various combinations of television and online media. Front Public Health 2018;6:83 [FREE Full text] [doi: 10.3389/fpubh.2018.00083] [Medline: 29629366]
38. Pettigrew S, Jongenelis M, Phillips F, Slevin T, Allom V, Keightley S, et al. Assessing audience members' ability to identify the media source of a health campaign disseminated via different media. Front Public Health 2018;6:196 [FREE Full text] [doi: 10.3389/fpubh.2018.00196] [Medline: 30073159]
39. von Weiler P, Bayard B, Sheard J. The Australian National Preventive Health Agency: How Suffering Made People Quit. WARC. 2014. URL: https://www.warc.com/content/paywall/article/afa/the_australian_national_preventive_health_agency_how_suffering_made_people_quit/102685 [accessed 2018-10-01]
40. Story & The Scottish Government. 'Take it right outside' Campaign Tracking Research Key findings - Post-campaign. The Scottish Government. 2014 Sep. URL: https://www2.gov.scot/Resource/0045/00459841.pdf [accessed 2018-10-01]
41. Farrelly MC, Duke JC, Nonnemaker J, MacMonegle AJ, Alexander TN, Zhao X, et al. Association between the real cost media campaign and smoking initiation among youths - United States, 2014-2016. MMWR Morb Mortal Wkly Rep 2017 Jan 20;66(2):47-50 [FREE Full text] [doi: 10.15585/mmwr.mm6602a2] [Medline: 28103214]
42. Huang L, Lazard AJ, Pepper JK, Noar SM, Ranney LM, Goldstein AO. Impact of the real cost campaign on adolescents' recall, attitudes, and risk perceptions about tobacco use: a national study. Int J Environ Res Public Health 2017 Jan 4;14(1):pii: E42 [FREE Full text] [doi: 10.3390/ijerph14010042] [Medline: 28054993]
43. Kranzler EC, Gibson LA, Hornik RC. Recall of 'The real cost' anti-smoking campaign is specifically associated with endorsement of campaign-targeted beliefs. J Health Commun 2017 Oct;22(10):818-828 [FREE Full text] [doi: 10.1080/10810730.2017.1364311] [Medline: 28937865]
44. McAfee T, Davis KC, Alexander RL, Pechacek TF, Bunnell R. Effect of the first federally funded US antismoking national media campaign. Lancet 2013 Dec 14;382(9909):2003-2011. [doi: 10.1016/S0140-6736(13)61686-4] [Medline: 24029166]
45. Duke JC, Davis KC, Alexander RL, MacMonegle AJ, Fraze JL, Rodes RM, et al. Impact of a US antismoking national media campaign on beliefs, cognitions and quit intentions. Health Educ Res 2015 Jun;30(3):466-483 [FREE Full text] [doi: 10.1093/her/cyv017] [Medline: 25976009]
46. Huang L, Thrasher JF, Abad EN, Cummings KM, Bansal-Travers M, Brown A, et al. The US national tips from former smokers antismoking campaign: promoting awareness of smoking-related risks, cessation resources, and cessation behaviors. Health Educ Behav 2015 Aug;42(4):480-486 [FREE Full text] [doi: 10.1177/1090198114564503] [Medline: 25588934]
47. Neff LJ, Patel D, Davis K, Ridgeway W, Shafer P, Cox S. Evaluation of the national tips from former smokers campaign: the 2014 longitudinal cohort. Prev Chronic Dis 2016 Mar 24;13:E42 [FREE Full text] [doi: 10.5888/pcd13.150556] [Medline: 27010845]
48. Zhao X, Cai X. The association between exposure to 'Tips' and smoking-related outcomes among adolescents in the United States. Health Educ Res 2016 Oct;31(5):614-623. [doi: 10.1093/her/cyw037] [Medline: 27422897]
49. Zhao X, Cai X. Exposure to the tips from former smokers campaign among adolescents in the United States. Nicotine Tob Res 2016 May;18(5):971-975. [doi: 10.1093/ntr/ntv265] [Medline: 26647103]
50. McAfee T, Davis KC, Shafer P, Patel D, Alexander R, Bunnell R. Increasing the dose of television advertising in a national antismoking media campaign: results from a randomised field trial. Tob Control 2017 Jan;26(1):19-28 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052517] [Medline: 26678518]
51. Evans WD, Rath J, Pitzer L, Hair EC, Snider J, Cantrell J, et al. Design and feasibility testing of the truth FinishIt tobacco countermarketing brand equity scale. J Health Commun 2016 Jul;21(7):800-808. [doi: 10.1080/10810730.2016.1157658] [Medline: 27315354]
52. Vallone D, Smith A, Kenney T, Greenberg M, Hair E, Cantrell J, et al. Agents of social change: a model for targeting and engaging generation z across platforms: how a nonprofit rebuilt an advertising campaign to curb smoking by teens and young adults. J Adv Res 2016 Dec;56(4):414-425. [doi: 10.2501/jar-2016-046]
53. Vallone D, Greenberg M, Xiao H, Bennett M, Cantrell J, Rath J, et al. The effect of branding to promote healthy behavior: reducing tobacco use among youth and young adults. Int J Environ Res Public Health 2017 Dec 7;14(12):pii: E1517 [FREE Full text] [doi: 10.3390/ijerph14121517] [Medline: 29215555]
54. Evans WD, Rath JM, Hair EC, Snider JW, Pitzer L, Greenberg M, et al. Effects of the truth FinishIt brand on tobacco outcomes. Prev Med Rep 2018 Mar;9:6-11 [FREE Full text] [doi: 10.1016/j.pmedr.2017.11.008] [Medline: 29276667]
55. Rowa-Dewar N, Amos A. Disadvantaged parents' engagement with a national secondhand smoke in the home mass media campaign: a qualitative study. Int J Environ Res Public Health 2016 Sep 9;13(9):- [FREE Full text] [doi: 10.3390/ijerph13090901] [Medline: 27618085]

56. Emery SL, Szczypka G, Abril EP, Kim Y, Vera L. Are you scared yet?: evaluating fear appeal messages in tweets about the tips campaign. J Commun 2014 Apr;64:278-295 [FREE Full text] [doi: 10.1111/jcom.12083] [Medline: 25429162]
57. Abril EP, Szczypka G, Emery SL. LMFAO! Humor as a response to fear: Decomposing fear control within the extended parallel process model. J Broadcast Electron Media 2017;61(1):126-143 [FREE Full text] [doi: 10.1080/08838151.2016.1273921] [Medline: 29527092]
58. Davis KC, Duke J, Shafer P, Patel D, Rodes R, Beistle D. Perceived effectiveness of antismoking ads and association with quit attempts among smokers: evidence from the tips from former smokers campaign. Health Commun 2017 Aug;32(8):931-938. [doi: 10.1080/10410236.2016.1196413] [Medline: 27435919]
59. Clayforth C, Pettigrew S, Mooney K, Lansdorp-Vogelaar I, Rosenberg M, Slevin T. A cost-effectiveness analysis of online, radio and print tobacco control advertisements targeting 25-39 year-old males. Aust N Z J Public Health 2014 Jun;38(3):270-274. [doi: 10.1111/1753-6405.12175] [Medline: 24750454]
60. Baskerville NB, Azagba S, Norman C, McKeown K, Brown KS. Effect of a digital social media campaign on young adult smoking cessation. Nicotine Tob Res 2016 Mar;18(3):351-360. [doi: 10.1093/ntr/ntv119] [Medline: 26045252]
61. Alberta Health Services: Tobacco Cessation - Keep Trying. WARC. 2015. URL: https://www.warc.com/content/paywall/article/cca/alberta_health_services_tobacco_cessation_keep_trying/103624 [accessed 2018-10-01]
62. Clearway Minnesota. QUITPLAN Services: No Judgements. Just Help. WARC. 2015. URL: https://www.warc.com/content/paywall/article/effies/quitplan_services_no_judgements_just_help/104923 [accessed 2018-10-01]
63. Public Health England. Stoptober 2017: Campaign Evaluation. Government of UK. 2018. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/743844/Stoptober_2017_campaign_evaluation.pdf [accessed 2018-10-01]
64. Truth Initiative: Left Swipe Dat. WARC. 2016. URL: https://www.warc.com/content/paywall/article/sabre-awards/truth_initiative_left_swipe_dat/120265 [accessed 2018-10-01]
65. Centers for Disease Control and Prevention (CDC). Increases in quitline calls and smoking cessation website visitors during a national tobacco education campaign--March 19-June 10, 2012. MMWR Morb Mortal Wkly Rep 2012 Aug 31;61(34):667-670 [FREE Full text] [Medline: 22932300]
66. Centers for Disease Control and Prevention (CDC). Impact of a national tobacco education campaign on weekly numbers of quitline calls and website visitors--United States, March 4-June 23, 2013. MMWR Morb Mortal Wkly Rep 2013 Sep 20;62(37):763-767 [FREE Full text] [Medline: 24048152]
67. Arden J, Buckley P, Hirst N, Shardlow N, Walmsley M. Stoptober: A Radical New Way to Get England to Quit Smoking. WARC. 2016. URL: https://www.warc.com/content/paywall/article/ipa/stoptober_a_radical_new_way_to_get_england_to_quit_smoking/108012 [accessed 2018-10-01]
68. Ayers JW, Althouse BM, Emery S. Changes in internet searches associated with the 'Tips from Former Smokers' campaign. Am J Prev Med 2015 Jun;48(6):e27-e29 [FREE Full text] [doi: 10.1016/j.amepre.2015.03.015] [Medline: 25998929]
69. Public Health England. Public Health England: Stoptober 2016 Facebook Messenger Bot. WARC. 2017. URL: https://www.warc.com/content/paywall/article/dmauk/public_health_england_stoptober_2016_facebook_messenger_bot/117320 [accessed 2018-10-01]
70. Scottish Government: Second-Hand Smoke - Take It Right Outside. WARC. 2014. URL: https://www.warc.com/content/paywall/article/dmauk/scottish_government_secondhand_smoke_take_it_right_outside/103885 [accessed 2018-10-01]
71. truth: 'Left swipe dat'. WARC. 2016. URL: https://www.warc.com/content/paywall/article/truth_quot;left_swipe_datquot;/106879 [accessed 2018-10-01]
72. Keller PA, Schillo BA, Kerr AN, Lien RK, Saul J, Dreher M, et al. Increasing reach by offering choices: Results from an innovative model for statewide services for smoking cessation. Prev Med 2016 Oct;91:96-102. [doi: 10.1016/j.ypmed.2016.08.010] [Medline: 27514248]
73. Thai Health Promotion Foundation: Smoking Kid. WARC. 2015. URL: https://www.warc.com/content/paywall/article/Thai_Health_Promotion_Foundation_Smoking_kid/104782 [accessed 2018-10-01]
74. Brown J, Kotz D, Michie S, Stapleton J, Walmsley M, West R. How effective and cost-effective was the national mass media smoking cessation campaign 'Stoptober'? Drug Alcohol Depend 2014 Feb 1;135:52-58 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.11.003] [Medline: 24322004]
75. Public Health England. Stoptober 2016: Campaign Evaluation. The Government of UK. 2017. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/743814/Stoptober_2016_campaign_evaluation.pdf [accessed 2018-10-01]
76. England L, Tong VT, Rockhill K, Hsia J, McAfee T, Patel D, et al. Evaluation of a federally funded mass media campaign and smoking cessation in pregnant women: a population-based study in three states. BMJ Open 2017 Dec 19;7(12):e016826 [FREE Full text] [doi: 10.1136/bmjopen-2017-016826] [Medline: 29259054]
77. Chew R, Kim A, Chen V, Ruddle P, Morgan-Lopez A. Assessing target audiences of digital public health campaigns: A computational approach. In: Thomson R, Dancy C, Hyder A, Bisgin H, editors. Social, Cultural, and Behavioral Modeling. Cham: Springer; 2018:286-291.

78. MacMonegle AJ, Nonnemaker J, Duke JC, Farrelly MC, Zhao X, Delahanty JC, et al. Cost-effectiveness analysis of the real cost campaign's effect on smoking prevention. Am J Prev Med 2018 Sep;55(3):319-325 [FREE Full text] [doi: 10.1016/j.amepre.2018.05.006] [Medline: 30122214]
79. Duke JC, Hansen H, Kim AE, Curry L, Allen J. The use of social media by state tobacco control programs to promote smoking cessation: a cross-sectional study. J Med Internet Res 2014 Jul 10;16(7):e169 [FREE Full text] [doi: 10.2196/jmir.3430] [Medline: 25014311]
80. Xu X, Alexander RL, Simpson SA, Goates S, Nonnemaker JM, Davis KC, et al. A cost-effectiveness analysis of the first federally funded antismoking campaign. Am J Prev Med 2015 Mar;48(3):318-325 [FREE Full text] [doi: 10.1016/j.amepre.2014.10.011] [Medline: 25498550]
81. Zhang L, Vickerman K, Malarcher A, Carpenter K. Changes in Quitline caller characteristics during a national tobacco education campaign. Nicotine Tob Res 2015 Sep;17(9):1161-1166. [doi: 10.1093/ntr/ntu271] [Medline: 25561612]
82. Chung JE. A smoking cessation campaign on Twitter: understanding the use of Twitter and identifying major players in a health campaign. J Health Commun 2016 May;21(5):517-526. [doi: 10.1080/10810730.2015.1103332] [Medline: 27115179]
83. Zhang L, Malarcher A, Mann N, Campbell K, Davis K, Anderson C, et al. The influence of state-specific Quitline numbers on call volume during a national tobacco education campaign promoting 1-800-QUIT-NOW. Nicotine Tob Res 2016 Aug;18(8):1780-1785. [doi: 10.1093/ntr/ntw100] [Medline: 27073208]
84. Murphy-Hoefer R, Davis KC, Beistle D, King BA, Duke J, Rodes R, et al. Impact of the tips from former smokers campaign on population-level smoking cessation, 2012-2015. Prev Chronic Dis 2018 May 31;15:E71 [FREE Full text] [doi: 10.5888/pcd15.180051] [Medline: 29862960]
85. Weir BW, Cantrell J, Holtgrave DR, Greenberg MS, Kennedy RD, Rath JM, et al. Cost and threshold analysis of the campaign to prevent youth smoking in the United States. Int J Environ Res Public Health 2018 Aug 6;15(8):- [FREE Full text] [doi: 10.3390/ijerph15081662] [Medline: 30082612]
86. Fulgoni GM. In the digital world, not everything that can be measured matters. J Adv Res 2016;56(1):9-13. [doi: 10.2501/jar-2016-008]
87. Rappaport SD. Lessons learned from 197 metrics, 150 studies, and 12 essays. J Adv Res 2014;54(1):110-118. [doi: 10.2501/jar-54-1-110-118]
88. Hutchinson A. Instagram Announces Global Expansion of Hidden Like Counts Test. Social Media Today. 2019 Nov 15. URL: https://www.socialmediatoday.com/news/instagram-announces-global-expansion-of-hidden-like-counts-test/567361/ [accessed 2019-11-18]
89. Reichert C. Facebook Has Begun Hiding Likes. CNET. 2019 Sep 30. URL: https://www.cnet.com/news/facebook-has-begun-hiding-likes/ [accessed 2019-11-18]
90. Hutchinson A. Facebook to Pay $40 Million in Settlement Over Inflated Video Metrics. Social Media Today. 2019 Oct 8. URL: https://www.socialmediatoday.com/news/facebook-to-pay-40-million-in-settlement-over-inflated-video-metrics/564535/ [accessed 2019-11-14]
91. Fulgoni GM. Is the GRP really dead in a cross-platform ecosystem? J Adv Res 2015;55(4):358-361. [doi: 10.2501/jar-2015-019]
92. Weprin A. Nielsen Adds YouTube To Total Ad Ratings. MediaPost. 2019 Jan 7. URL: https://www.mediapost.com/publications/article/330163/nielsen-adds-youtube-to-total-ad-ratings.html [accessed 2019-10-14]
93. Neiger BL, Thackeray R, van Wagenen SA, Hanson CL, West JH, Barnes MD, et al. Use of social media in health promotion: purposes, key performance indicators, and evaluation metrics. Health Promot Pract 2012 Mar;13(2):159-164. [doi: 10.1177/1524839911433467] [Medline: 22382491]
94. Chen J. The Most Important Social Media Metrics to Track. Social Sprout. 2020 Jan 10. URL: https://sproutsocial.com/insights/social-media-metrics/ [accessed 2019-10-30]
95. Lim MS, Wright CJ, Carrotte ER, Pedrana AE. Reach, engagement, and effectiveness: a systematic review of evaluation methodologies used in health promotion via social networking sites. Health Promot J Austr 2016 Feb;27(3):187-197. [doi: 10.1071/HE16057] [Medline: 27719734]
96. Platt T, Platt J, Thiel DB, Kardia SL. Facebook advertising across an engagement spectrum: a case example for public health communication. JMIR Public Health Surveill 2016 May 30;2(1):e27 [FREE Full text] [doi: 10.2196/publichealth.5623] [Medline: 27244774]
97. Thackeray R, Neiger BL, Smith AK, van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012 Mar 26;12:242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]
98. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
99. Freeman B, Potente S, Rock V, McIver J. Social media campaigns that make a difference: what can public health learn from the corporate sector and other social change marketers? Public Health Res Pract 2015 Mar 30;25(2):e2521517 [FREE Full text] [doi: 10.17061/phrp2521517] [Medline: 25848735]
100. Heldman AB, Schindelar J, Weaver JB. Social media engagement and public health communication: implications for public health organizations being truly 'social'. Public Health Rev 2013;35(1):13. [doi: 10.1007/BF03391698]

101. Snyder J, Garcia-Garcia M. Advertising across platforms: conditions for multimedia campaigns: a method for determining optimal media investment and creative strategies across platforms. J Adv Res 2016;56(4):352-367. [doi: 10.2501/jar-2016-042]
102. Ford JB. What do we know about digital attribution? J Adv Res 2018 Dec 11;58(4):383-384. [doi: 10.2501/jar-2018-044]
103. Nisar TM, Yeung M. Attribution modeling in digital advertising: an empirical investigation of the impact of digital sales channels. J Adv Res 2018;58(4):399-413. [doi: 10.2501/jar-2017-055]
104. What We Know About Attribution Marketing Mix Modelling. WARC. 2019 Aug. URL: https://www.warc.com/content/paywall/article/bestprac/what_we_know_about_attribution_and_marketing_mix_modelling/107151 [accessed 2019-10-30]
105. Niederdeppe J. Meeting the challenge of measuring communication exposure in the digital age. Commun Method Meas 2016 Apr 20;10(2-3):170-172. [doi: 10.1080/19312458.2016.1150970]
106. de Vreese CH, Neijens P. Measuring media exposure in a changing communications environment. Commun Method Meas 2016;10(2-3):69-80. [doi: 10.1080/19312458.2016.1150441]
107. Kite J, Grunseit A, Bohn-Goldbaum E, Bellew B, Carroll T, Bauman A. A systematic search and review of adult-targeted overweight and obesity prevention mass media campaigns and their evaluation: 2000-2017. J Health Commun 2018;23(2):207-232. [doi: 10.1080/10810730.2018.1423651] [Medline: 29336684]

Abbreviations

GRPs: gross rating points
MEDLINE: Medical Literature Analysis and Retrieval System Online
TARPs: target audience rating points

Edited by G Eysenbach; submitted 12.12.19; peer-reviewed by M Navarro, S Beasley; comments to author 01.01.20; revised version received 12.02.20; accepted 21.02.20; published 11.08.20.

Please cite as:
Chan L, O'Hara B, Phongsavan P, Bauman A, Freeman B
Review of Evaluation Metrics Used in Digital and Traditional Tobacco Control Campaigns
J Med Internet Res 2020;22(8):e17432
URL: https://www.jmir.org/2020/8/e17432
doi:10.2196/17432
PMID:32348272

Original Paper

Online Guide for Electronic Health Evaluation Approaches: Systematic Scoping Review and Concept Mapping Study

Tobias N Bonten[1,2*], MD, PhD; Anneloek Rauwerdink[3*], MD; Jeremy C Wyatt[4], MD, PhD; Marise J Kasteleyn[1,2], PhD; Leonard Witkamp[5,6], MD, PhD; Heleen Riper[7], PhD; Lisette JEWC van Gemert-Pijnen[8], PhD; Kathrin Cresswell[9], PhD; Aziz Sheikh[9], PhD; Marlies P Schijven[3], MD, PhD; Niels H Chavannes[1,2], MD, PhD; EHealth Evaluation Research Group[10]

[1]Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, Netherlands
[2]National eHealth Living Lab, Leiden, Netherlands
[3]Department of Surgery, Amsterdam Gastroenterology and Metabolism, Amsterdam UMC, Amsterdam, Netherlands
[4]Wessex Institute, University of Southampton, Southampton, United Kingdom
[5]Department of Medical Informatics, Amsterdam UMC, Amsterdam, Netherlands
[6]Ksyos Health Management Research, Amstelveen, Netherlands
[7]Department of Clinical, Neuro and Developmental Psychology, Vrije Universiteit, Amsterdam, Netherlands
[8]Department of Psychology, Health and Technology, Centre for eHealth and Wellbeing Research, University of Twente, Enschede, Netherlands
[9]Centre of Medical Informatics, Usher Institute, The University of Edinburgh, Medical School, Edinburgh, United Kingdom
[10] Please see acknowledgements section for list of collaborators
[*]these authors contributed equally

Corresponding Author:
Tobias N Bonten, MD, PhD
Department of Public Health and Primary Care
Leiden University Medical Centre
Department of Public Health & Primary Care, Room V6-22
PO Box 9600
Leiden, 2300 RC
Netherlands
Phone: 31 715268433
Email: t.n.bonten@lumc.nl

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2020/8/e23642/

Abstract

Background: Despite the increase in use and high expectations of digital health solutions, scientific evidence about the effectiveness of electronic health (eHealth) and other aspects such as usability and accuracy is lagging behind. eHealth solutions are complex interventions, which require a wide array of evaluation approaches that are capable of answering the many different questions that arise during the consecutive study phases of eHealth development and implementation. However, evaluators seem to struggle in choosing suitable evaluation approaches in relation to a specific study phase.

Objective: The objective of this project was to provide a structured overview of the existing eHealth evaluation approaches, with the aim of assisting eHealth evaluators in selecting a suitable approach for evaluating their eHealth solution at a specific evaluation study phase.

Methods: Three consecutive steps were followed. Step 1 was a systematic scoping review, summarizing existing eHealth evaluation approaches. Step 2 was a concept mapping study asking eHealth researchers about approaches for evaluating eHealth. In step 3, the results of step 1 and 2 were used to develop an “eHealth evaluation cycle” and subsequently compose the online “eHealth methodology guide.”

Results: The scoping review yielded 57 articles describing 50 unique evaluation approaches. The concept mapping study questioned 43 eHealth researchers, resulting in 48 unique approaches. After removing duplicates, 75 unique evaluation approaches remained. Thereafter, an "eHealth evaluation cycle" was developed, consisting of six evaluation study phases: conceptual and planning, design, development and usability, pilot (feasibility), effectiveness (impact), uptake (implementation), and all phases. Finally, the "eHealth methodology guide" was composed by assigning the 75 evaluation approaches to the specific study phases of the "eHealth evaluation cycle."

Conclusions: Seventy-five unique evaluation approaches were found in the literature and suggested by eHealth researchers, which served as content for the online "eHealth methodology guide." By assisting evaluators in selecting a suitable evaluation approach in relation to a specific study phase of the "eHealth evaluation cycle," the guide aims to enhance the quality, safety, and successful long-term implementation of novel eHealth solutions.

(J Med Internet Res 2020;22(8):e17774) doi:10.2196/17774

KEYWORDS

eHealth; mHealth; digital health; methodology; study design; health technology assessment; evaluation; scoping review; concept mapping

Introduction

Background

Electronic health (eHealth) solutions play an increasingly important role in the sustainability of future health care systems. An increase in the use and adoption of eHealth has been observed in the last decade. For instance, 59% of the member states of the European Union had a national eHealth record system in 2016 [1]. Despite the increase in use and high expectations about the impact of eHealth solutions, scientific evidence about the effectiveness, along with other aspects such as usability and accuracy, is often lagging behind [2-6]. In addition, due to rising demands such as time and cost restrictions from policymakers and commercial interests, the quality of eHealth evaluation studies is under pressure [7-9]. Although most eHealth researchers are aware of these limitations and threats, they may find it difficult to determine the most suitable evaluation approach to evaluate their novel eHealth solution since a clear overview of the wide array of evaluation approaches is lacking. However, to safely and successfully implement novel eHealth solutions into existing health care pathways, and to facilitate long-term implementation, robust scientific evaluation is paramount [10].

Limitations of Classic Methodologies in eHealth Research

The most rigorous method to study the effects of health interventions is considered to be the double blinded parallel-group randomized controlled trial (RCT). Randomization has the unique ability to distribute both known and unknown confounders between study arms equally [11]. Although many RCTs of eHealth solutions have been published, limitations of this method are frequently described in the literature [12]. For instance, information bias could occur due to blinding difficulties because of the visibility of an eHealth solution [13-16]. Moreover, conducting an RCT can be very time-consuming, whereas eHealth technology develops rapidly. Consequently, before the trial results are known, the tested eHealth solution may be outdated [17]. Further, "contamination" in which the control group also uses a digital intervention, despite being randomized to the no-intervention group, easily occurs in eHealth research. Another drawback of placing too much focus on the classical research methodologies that are generally used to evaluate effectiveness is that the need for significant evaluation during the development and implementation phases of eHealth is often neglected. Additionally, validating the quality and evaluating behavioral aspects of an eHealth solution may be lacking [18,19]. Although it is not wrong to use classical research methods such as an RCT to study eHealth solutions, given the fact that eHealth solutions are considered to be "complex" interventions, more awareness about the wide array of eHealth evaluation approaches may be required.

Evaluation of eHealth as a Complex Intervention

As described by the Medical Research Council (MRC) Framework 2000, eHealth solutions typically have multiple interacting components presenting several additional problems for evaluators, besides the already practical and methodological difficulties described above [20,21]. Because of these difficulties, eHealth solutions are considered as complex interventions. To study such interventions, multiple evaluation approaches are needed that are capable of answering the many different questions that arise during the consecutive phases of intervention development and implementation, including the "development," "feasibility and piloting," "evaluation," and "implementation" phases [21]. For instance, to assess the effectiveness of complex interventions, the MRC Framework authors suggest the following experimental designs: individually randomized trials, cluster randomized trials, stepped wedge designs, preference trials, randomized consent designs, and N-of-1 designs. Unfortunately, the authors did not offer suggestions of evaluation approaches to use in the other phases of the MRC Framework. Murray et al [20] proposed a staged approach to the evaluation of eHealth that is modeled on the MRC Framework for Complex Interventions with 10 core questions to help developers quantify the costs, scalability, sustainability, and risks of harm of the eHealth solution. Greenhalgh et al [22] developed the Nonadoption, Abandonment, and challenges to Scale-up, Spread, and Sustainability (NASSS) framework to identify, understand, and address the interacting challenges around achieving sustained adoption, local scale-up, distant spread, and long-term

sustainability of eHealth programs. Both of these studies illustrated and justified the necessity of a variety of evaluation approaches for eHealth beyond the RCT; however, this research does not assist evaluators in choosing which approach to use in a selected evaluation study phase. Another suggestion to improve the quality of eHealth research was proposed by Nykanen et al [23,24], who developed the guideline for Good Evaluation Practice in Health Informatics (GEP-HI), which precisely describes how to design and carry out a health informatics evaluation study in relation to the evaluation study phases. However, this guideline also did not include information on which specific evaluation approaches could be used in the related study phases. Besides the individual studies described above, there have been several books published concerning eHealth evaluation research. Among one of the first books on the topic is the “Handbook of Evaluation Methods for Health Informatics,” which was published in 2006 [25]. The aim of this book was to suggest options for finding appropriate tools to support the user in accomplishing an evaluation study. The book contains more than 30 evaluation methods, which are related to the phases of the system lifecycle, and the reliability, degree of difficulty, and resource requirements for each method are described. Moreover, the book “Evidence-Based Health Informatics,” published in 2016 [26], provides the reader with a better understanding of the need for robust evidence to improve the quality of health informatics. The book also provides a practical overview of methodological considerations for health information technology, such as using the best study design, stakeholder analysis, mixed methods, clinical simulation, and evaluation of implementation.

Although useful work has been performed by these previous authors, no single source is able to provide clear guidance in selecting appropriate evaluation approaches in relation to the specific evaluation phases of eHealth. Therefore, to enhance quality and safety, and to facilitate long-term implementation of eHealth solutions into daily practice, raising the awareness of eHealth evaluators about the wide array of eHealth evaluation approaches and thereby enhancing the completeness of evidence is sorely needed [27].

Aim and Objectives

The overall aim of the present study was to raise awareness among eHealth evaluators about the wide array of eHealth evaluation approaches and the existence of multiple evaluation study phases. Therewith, quality, safety, and successful long-term implementation of novel eHealth solutions may be enhanced.

To achieve this aim, we pursued the following objectives: (1) systematically map the current literature and expert knowledge on methods, study designs, frameworks, and philosophical approaches available to evaluate eHealth solutions; and (2) provide eHealth evaluators with an online “eHealth methodology guide” to assist them with selecting a suitable evaluation approach to evaluate their eHealth solution in a specific study phase.

Methods

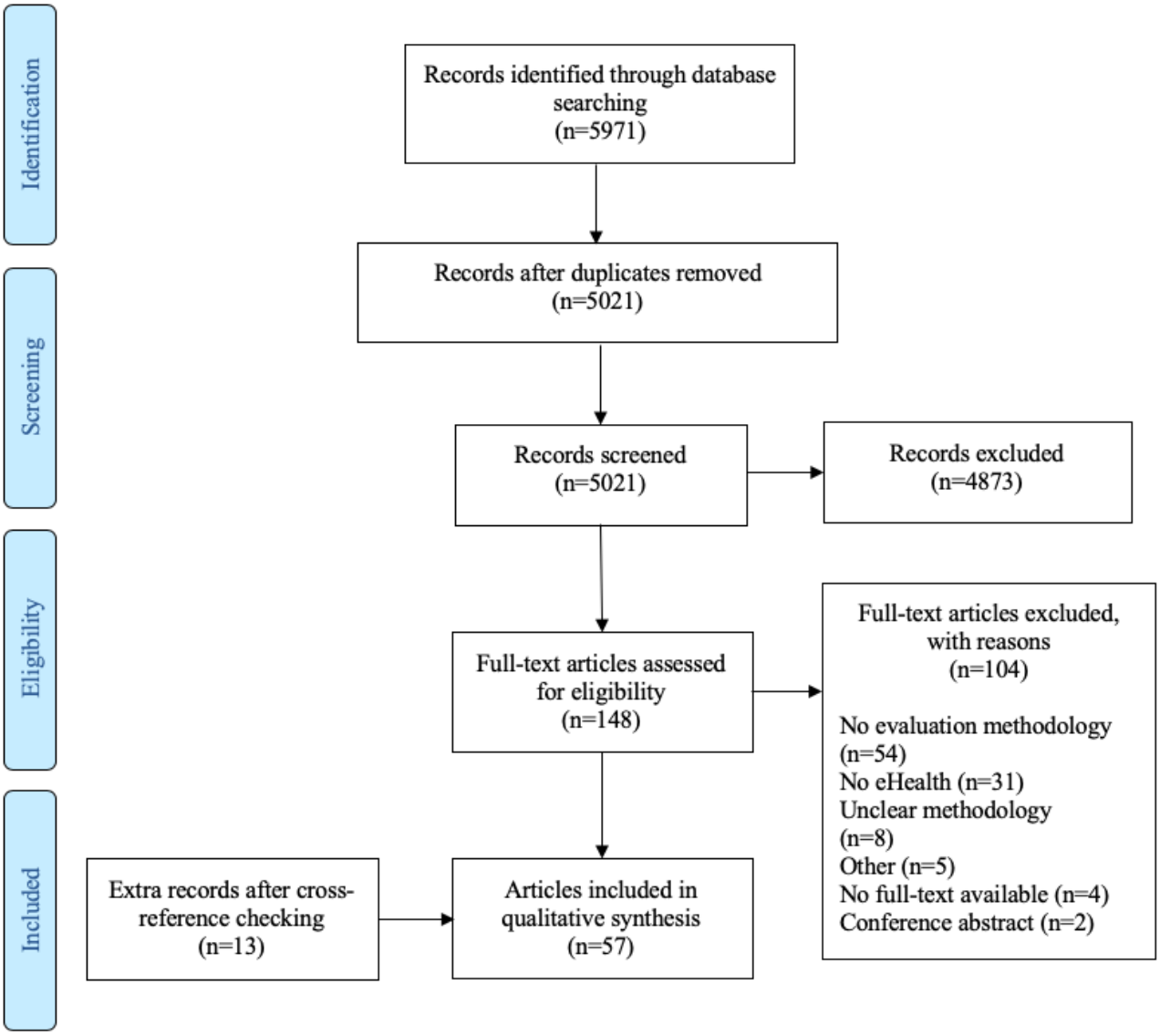
Overall Design

The project consisted of three consecutive steps: (1) a systematic scoping review, (2) concept mapping study, and (3) development of the "eHealth methodology guide" with content based on the results from steps 1 and 2.

Step 1: Systematic Scoping Review

To describe the methods, study designs, frameworks, and other philosophical approaches (collectively referred to as “evaluation approach[es]”) currently used to evaluate eHealth solutions, a systematic scoping review was conducted. The online databases Pubmed, Embase, and PsycINFO were systematically searched using the term ”eHealth” in combination with ”evaluation” OR “methodology.” The search included Medical Subject Headings or Emtree terms and free-text terms. A complete list of the search strings is shown in Multimedia Appendix 1. Broad inclusion criteria were applied. All types of peer-reviewed English language articles published from January 1, 2006 until November 11, 2016 and a subsequent update from November 12, 2016 until October 21, 2018 describing any eHealth evaluation approach were included. We reasoned that articles published before January 1, 2006 would not necessarily need to be screened because the annual number of publications related to eHealth evaluation approaches was still low at that time, suggesting that the field was just starting to take its first scientific steps. In addition, if an article did describe a useful evaluation approach, it would have also been described by articles that were published later. Two reviewers (TB and AR) independently screened the titles and abstracts of the articles according to the inclusion criteria described above. Cohen kappa coefficient was calculated to measure the initial interrater reliability. Disagreements between the reviewers were resolved by the decision of a third independent reviewer (MK). Full-text assessment of the selected articles after screening of titles and abstracts was performed by both reviewers (TB and AR). Exclusion criteria after full-text assessment were: no eHealth evaluation approach described, article did not concern eHealth, the described methodology was unclear, full-text version was not available, or the article was a conference abstract. The reference list of eligible articles was checked for relevant additional studies. These studies were also checked for eligibility and included as crossreferenced articles in the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) diagram (Figure 1). In the qualitative synthesis, the eHealth evaluation approach was extracted from eligible articles, and duplicates and synonyms were merged to develop a single list of all the methods.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the article selection process.



Step 2: Concept Mapping Study

Overview of Phases

Although the systematic scoping review was performed rigorously, it was possible that not all of the current or possible approaches to evaluate eHealth solutions would have been described in the eligible studies. Therefore, to achieve a reasonably complete overview of eHealth evaluation approaches, it was considered essential to incorporate eHealth researchers' knowledge on these approaches. A concept mapping study was selected as the most suitable method for structuring the suggested evaluation approaches from the researchers and for exploring their views on the different phases of the "eHealth evaluation cycle." Concept mapping is a qualitative research methodology that was introduced by Trochim and Linton in 1986 [28]. It can be used by a group of individuals to first determine the scope of ideas on a certain topic and then to structure these ideas [29]. There is no interaction between the participants. A typical concept mapping study consists of 5 phases: (1) selection of the participants; (2) brainstorm, generation of the evaluation approaches by participants; (3) sorting and rating of the evaluation approaches; (4) concept mapping analysis; (5) and interpretation and utilization of the concept map. In the next subsections, these 5 phases are described in more detail. Concept System 2017 Global MAX online software was used for these tasks [30]. A Venn diagram was drawn to visualize the overlap between the results of the scoping review (step 1) and the evaluation approaches suggested by participants (step 2).

Selection of the Participants

To include a wide cross-section of eHealth researchers and reduce the influence of "group think," any researchers in contact with the authors and with any level of expertise in eHealth or evaluation research (to help assure that all major perspectives on the eHealth evaluation topic were represented) were approached as being suitable participants for this concept mapping study. Snowball sampling (ie, asking participants to recruit other researchers) was also included in the recruitment strategy. The target participants received an email describing the objective of the study and instructions on how they could participate. A register was kept of the number of participants that were approached and that refused. In general, in a concept

mapping study, there are no "rules" established as to how many participants should be included [31]. However, we estimated that 25 or more participants would be appropriate to generate a sufficient number of evaluation approaches and to have representative sorting and rating results.

Brainstorm Phase: Generation of the List of Evaluation Approaches

In this phase, participants were asked to enter all of the evaluation approaches they were aware of into an online form using Global MAX software. We intentionally did not include a strict definition of "evaluation approaches" so as to maintain the concept mapping phase as broad as possible and to avoid missing any methods due to an overly restrictive definition. The participants were not familiar with the results of the systematic scoping review. Participants were also asked 8 general background questions about their age, gender, background, years of experience in research, type of health care institute they work at, whether their daily work comprised eHealth, self-rated expertise in eHealth in general (grade 1-10), and self-rated expertise (grade 1-10) in eHealth evaluation approaches.

Sorting and Rating Phases

The coordinating researcher (AR) reviewed the evaluation approaches suggested by the participants, checking if each suggested approach truly represented a specific evaluation approach rather than, for instance, a broad methodological category such as "qualitative research." If the coordinating researcher was unfamiliar with the suggested approach, Pubmed or Google Scholar was searched for supporting information. The cleaned results were combined with the results from the systematic scoping review, omitting duplicate approaches. The resulting set of approaches was then presented back to the participants who were instructed to sort these approaches into categories that had to be created by the participants. The participant was instructed to keep the following question in mind while sorting each approach into a self-created category: "To which phase of the research cycle (eg, planning, testing, implementation) does this evaluation approach belong?" To gain insights about opinions of the researchers with respect to the use in daily practice and suitability for effectiveness testing of the evaluation approaches, the participants were asked the following three rating questions about each approach: (1) Does your research group use this approach, or did it do so in the past? (yes or no); (2) In your opinion, how important is it that researchers with an interest in eHealth are familiar with this approach? (1, unimportant; 2, less important; 3, very important; 4, absolutely essential); (3) In your opinion, how important is the approach for proving the effectiveness of eHealth? (1, unimportant; 2, less important; 3, very important; 4, absolutely essential).

Results of the first rating question are reported as percentages of how many participants use or used the approach. For the second and third questions related to familiarity with the approach and importance for proving effectiveness, respectively, average rating scores ranging from 1 to 4 for each evaluation approach and the proportion of participants who selected categories 3 or 4 are reported.

Concept Mapping Analysis

Global MAX software uses a 3-step analysis to compute the concept map [32]. First, the sorting data from each participant were compiled into a similarity matrix. The matrix illustrates how many times each approach was sorted into similar categories. Second, the software applied a multidimensional scaling algorithm to plot points that were frequently sorted close together on a point map. A stress value (0-1), indicating the goodness of fit of the configuration of the point map, was calculated; the lower the stress value, the better the fit. In the last step, a hierarchical cluster analysis using the Ward algorithm was applied to group approaches into clusters (see also pages 87-100 of Kane and Trochim [33] for a detailed description of the data analyses to compute concept maps).

Two authors (TN and AR) reviewed the concept maps ranging from a 7-cluster to a 3-cluster option. The guidance of Kane and Trochim [33] was followed to select the best fitting number of clusters. Once the best fitting number of clusters was identified, each evaluation approach on the concept map was reviewed by the two authors to check if the approach truly belonged to the assigned cluster. If the approach seemed to belong in an adjacent cluster, it was reassigned to that particular cluster. If an approach could be assigned to multiple clusters, the best fitting cluster was selected.

The average rating scores for the rating questions on familiarity with the approach and importance for proving effectiveness were used to create a 4-quadrant Go-Zone graph. The Go-Zone graph easily visualizes the evaluation approaches with above-average rating scores on both questions, which are represented in the upper right quadrant. Approaches in the upper right quadrant that were also mentioned in the effectiveness testing cluster of the concept map are asterisked in the "eHealth methodology guide," meaning that participants in general used these approaches and that these approaches were recommended by participants for evaluating effectiveness.

Interpretation and Utilization of the Concept Map

The initial concept map clusters represented names that participants suggested when sorting the evaluation approaches into self-created categories. Because these cluster names were used to constitute the phases of the "eHealth evaluation cycle" later in the project, three authors (TN, AR, and JW) determined (after multiple discussion sessions) the most appropriate names for the final concept map clusters. A name was found to be appropriate when it was suggested by multiple participants and was considered to be representative for the "eHealth evaluation cycle," meaning that all of the evaluation approaches could be logically subdivided. After updating the names, the concept map clusters still contained the evaluation approaches allocated by the participants. This subdivision of eHealth evaluation approaches was used as the content for the "eHealth evaluation guide."

Step 3: eHealth Methodology Guide

The unique evaluation approaches identified in the systematic scoping review and unique evaluation approaches described by the participants in the concept mapping study were brought together by authors TB and AR, and used as the content to

develop the "eHealth methodology guide." To logically subdivide the eHealth evaluation approaches and to increase researchers' awareness of the existence of multiple evaluation study phases, an "eHealth evaluation cycle" was developed. The cycle was based on the cluster names of the concept map and on the common denominators of the "all phases" evaluation approaches from the systematic scoping review. Each unique evaluation approach was assigned to a specific evaluation study phase. If an approach could belong to multiple study phases, it was assigned to all applicable phases.

Results

Step 1: Systematic Scoping Review

The systematic search retrieved 5971 articles from the databases. After removing duplicates, 5021 articles were screened using title and abstract review. A total of 148 articles were selected for full-text assessment. Among these, 104 articles were excluded because of the following reasons: not containing any named eHealth evaluation approach, not being an eHealth article, unclear description of approach, no full-text version available, conference abstract, and other reasons. Through crossreferencing, 13 additional articles were added to the final selection. In total, 57 articles were included in the qualitative synthesis. Calculation of Cohen kappa showed an interrater reliability of 0.49, which corresponds to "moderate agreement" between both reviewers. Figure 1 presents the PRISMA flow diagram describing the selection process. The 57 articles described 50 unique eHealth evaluation approaches (Table 1). Of the 50 methods, 19 were described by more than 1 article.

Table 1. Articles included in the systematic scoping review according to the evaluation approach adopted.

Reference	Year	Country	Evaluation approach
Chiasson et al [34]	2007	United Kingdom	Action research
Campbell and Yue [35]	2014	United States	Adaptive design; propensity score
Law and Wason [36]	2014	United Kingdom	Adaptive design
Mohr et al [16]	2015	United States	Behavioral intervention technology model (bit) in Trials of Intervention Principles; SMART[a]
Van Gemert-Pijnen et al [37]	2011	Netherlands	CeHRes[b] Roadmap
Alpay et al [38]	2018	Netherlands	CeHRes Roadmap; Fog model; Oinas-Kukkonen model
Shaw [39]	2002	United Kingdom	CHEATS[c]: a generic ICT[d] evaluation framework
Kushniruk and Patel [40]	2004	Canada	Cognitive task analysis; user-centered design
Jaspers [41]	2009	Netherlands	Cognitive walkthrough; heuristic evaluation; think-aloud method
Khajouei et al [42]	2017	Iran	Cognitive walkthrough; heuristic evaluation
Van Engen-Verheul et al [43]	2015	Netherlands	Concept mapping
Mohr et al [44]	2013	United States	CEEBIT[e] framework
Nicholas et al [45]	2016	Australia	CEEBIT framework; single-case experiment (N=1)
Bongiovanni-Delaroziere and Le Goff Pronost [46]	2017	France	Economic evaluation; HAS[f] methodological framework
Fatehi et al [47]	2017	Australia	Five-stage model for comprehensive research on telehealth
Baker et al [48]	2014	United States	Fractional-factorial (ANOVA[g]) design; SMART
Collins et al [49]	2007	United States	Fractional-factorial (ANOVA) design; MOST[h]; SMART
Chumbler et al [14]	2008	United States	Interrupted time-series analysis; matched cohort study design
Grigsby et al [50]	2006	United States	Interrupted time-series analysis; pretest-posttest design
Liu and Wyatt [51]	2001	United Kingdom	Interrupted time-series analysis
Kontopantelis et al [52]	2015	United Kingdom	Interrupted time-series analysis
Catwell and Shiekh [53]	2009	United Kingdom	Life cycle–based approach
Han [54]	2011	United States	Life cycle–based approach
Sieverink [55]	2017	Netherlands	Logfile analysis
Kramer-Jackman Popkess-Vawter [56]	2008	United States	Method for technology-delivered health care measures
Wilson et al [57]	2018	Canada	mHealth[i] agile and user-centered research and development lifecycle
Jacobs and Graham[58]	2016	United States	mHealth development and evaluation framework; MOST
Dempsey et al [59]	2015	United States	Microrandomized trial; single-case experiment (N=1)

Reference	Year	Country	Evaluation approach
Klasnja et al [60]	2015	United States	Microrandomized trial; single-case experiment (N=1)
Law et al [61]	2016	United Kingdom	Microrandomized trial
Walton et al [62]	2018	United States	Microrandomized trial
Caffery et al [63]	2017	Australia	Mixed methods
Lee and Smith [64]	2012	United States	Mixed methods
Kidholm et al [65]	2017	Denmark	MAST[j]
Kidholm et al [66]	2018	Denmark	MAST
Kummervold et al [67]	2012	Norway	Noninferiority trial
May [68]	2006	United Kingdom	Normalization process theory and checklist
Borycki et al [69]	2016	Canada	Participatory design; user-centered design
Clemensen et al [70]	2017	Denmark	Participatory design
Glasgow [71]	2007	United States	Practical clinical trial; RE-AIM[k] framework
Danaher and Seeley [72]	2007	United States	Pragmatic randomized controlled trial; SMART; Stage model of behavioral therapies research
Sadegh et al [73]	2018	Iran	Proposed framework for evaluated mHealth services
Harker and Kleinen [74]	2012	United Kingdom	Rapid review
Glasgow et al [75]	2014	United States	RE-AIM framework
Almirall et al [76]	2014	United States	SMART
Ammenwerth et al [77]	2012	Austria	Simulation study
Jensen et al [78]	2015	Denmark	Simulation study
Dallery et al [79]	2013	United States	Single case experiment (N=1)
Cresswell and Shiekh [80]	2014	United Kingdom	Sociotechnical evaluation
Kaufman et al [81]	2006	United States	Stead et al [82] evaluation framework
Brown and Lilford [83]	2006	United Kingdom	Stepped wedge (cluster) randomized trial
Hussey and Hughes [84]	2007	United States	Stepped wedge (cluster) randomized trial
Spiegelman [85]	2016	United States	Stepped wedge (cluster) randomized trial
Langbecker et al [86]	2017	Australia	Survey methods
Rönnby et al [87]	2018	Sweden	Technology acceptance model
Bastien [88]	2010	France	User-based evaluation

Reference	Year	Country	Evaluation approach
Nguyen et al [89]	2007	Canada	Waitlist control group design

[a]SMART: Sequential Multiple Assignment Randomized Trial.

[b]CeHRes: Centre for eHealth Research and Disease management.

[c]CHEATS: Clinical, human and organizational, educational, administrative, ethnical and social explanatory factors in a randomized controlled trial intervention.

[d]ICT: information and communication technology.

[e]CEEBIT: continuous evaluation of evolving behavioral intervention technology.

[f]HAS: Haute Autorité de Santé (French National Authority for Health).

[g]ANOVA: analysis of variance.

[h]MOST: multiphase optimization strategy.

[i]mHealth: mobile health.

[j]MAST: Model of Assessment of Telemedicine Applications.

[k]RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance.

Step 2: Concept Mapping Study

Characteristics of the Participants

In total, 52 researchers were approached to participate in the concept mapping study, 43 (83%) of whom participated in the "brainstorm" phase. Reasons for refusal to participate were a lack of time or not feeling skilled enough to contribute. From the 43 initial participants, 27 (63%) completed the "sorting" phase and 32 (74%) answered the three rating questions of the "rating" phase. The characteristics of participants for each phase are shown in Table 2. Participant characteristics did not change substantially throughout the study phases, with a mean participant age ranging from 39.9 to 40.5 years, a mean of 13 years of eHealth research experience, and more than 70% of participants working in a university medical center. The majority of participants gave themselves high grades for their knowledge about eHealth but lower scores for their expertise in eHealth evaluation approaches.

Table 2. Characteristics of study participants for each phase of the concept mapping study.

Characteristic	Brainstorm phase	Sorting phase	Rating phase
Participants (n)	43[a]	27	32[b]
Age (years), mean (SD)	39.9 (12.1)	39.0 (12.6)	40.5 (13)
Female gender, n (%)	21 (49)	16 (53)	16 (50)
Research experience (years), mean (SD)	13.5 (10.8)	12.6 (10.5)	13.9 (11)
Working in university medical center, n (%)	37 (73)	26 (72)	27 (71)
Use of eHealth[c] in daily practice, n (%)			
During clinic work, not EHR[d]	4 (7)	3 (9)	3 (8)
During research	32 (59)	21 (60)	23 (59)
During clinic work and research	10 (19)	7 (20)	8 (21)
No	1 (2)	0 (0)	1 (3)
Other	7 (13)	4 (11)	4 (10)
Knowledge about eHealth, n (%)			
Grade 1-2	0 (0)	0 (0)	0 (0)
Grade 3-4	1 (2)	1 (4)	1 (3)
Grade 5-6	2 (5)	1 (4)	1 (3)
Grade 7-8	29 (71)	17 (63)	21 (68)
Grade 9-10	9 (22)	8 (30)	8 (26)
Expertise about eHealth research methods, n (%)			
Grade 1-2	0 (0)	0 (0)	0 (0)
Grade 3-4	1 (2)	1 (4)	1 (3)
Grade 5-6	15 (37)	8 (30)	11 (36)
Grade 7-8	19 (46)	15 (56)	15 (48)
Grade 9-10	6 (15)	3 (11)	4 (13)
Background, n (%)			
Biology	2 (3)	1 (2)	1 (2)
Data science	2 (3)	1 (2)	1 (2)
Economics	1 (1)	1 (2)	1 (2)
Medicine	24 (35)	14 (30)	18 (34)
(Health) Science	9 (13)	6 (13)	7 (13)
Industrial design	1 (1)	1 (2)	1 (2)
Informatics	4 (6)	3 (7)	3 (6)
Communication and culture	4 (6)	3 (7)	3 (6)
Psychology	14 (21)	11 (24)	12 (23)
Other	7 (10)	5 (11)	6 (11)

[a]43 participants participated in the sorting phase, but 41 participants answered the characteristics questions.

[b]One of the 32 participants did not finish the third rating question: "importance for proving effectiveness."

[c]eHealth: electronic health.

[d]EHR: electronic health record.

Brainstorm Phase

Forty-three participants participated in an online brainstorm phase and generated a total of 192 evaluation approaches. After removing duplicate or undefined approaches, 48 unique approaches remained (Multimedia Appendix 2). Only 23 of these 48 approaches (48%) overlapped with those identified in the systematic scoping review (Figure 2).

Based on the update of the scoping literature review at the end of the project, 13 additional evaluation approaches were found that were not incorporated into the sorting and rating phases. Therefore, in total, only 62 of the 75 unique methods were presented to the participants in the sorting and rating phases. Participants were asked to sort the 62 evaluation approaches into as many self-created categories as they wished. Twenty-seven individuals participated in this sorting exercise, and they suggested between 4 and 16 categories each, with a mean of 8 (SD 4) categories.

The rating questions on use of the approach, familiarity with the approach, and importance for proving effectiveness were answered by 32, 32, and 31 participants, respectively. An analysis of responses to these three questions is presented in Table 3 and the mean ratings for familiarity with the approach and importance for proving effectiveness are plotted on the Go-Zone graph shown in Figure 3. The evaluation approach used most frequently by the participants was the questionnaire, with 100% responding "yes." The approach that the participants used the least often was the Evaluative Questionnaire for E-health Tools at 3%. The average rating score for familiarity with the approach ranged from 1.9 for stage model of behavioral therapies to 3.6 for feasibility study. In addition, 88% of the participants thought that it is essential that researchers are familiar with the feasibility study method. The average rating score for importance for proving effectiveness ranged from 1.6 for vignette study to 3.3 for pragmatic RCT. In addition, 90% of the participants considered the stepped wedge trial design to be essential for proving the effectiveness of eHealth solutions.

Figure 2. Venn diagram showing the origin of the 75 unique evaluation approaches.

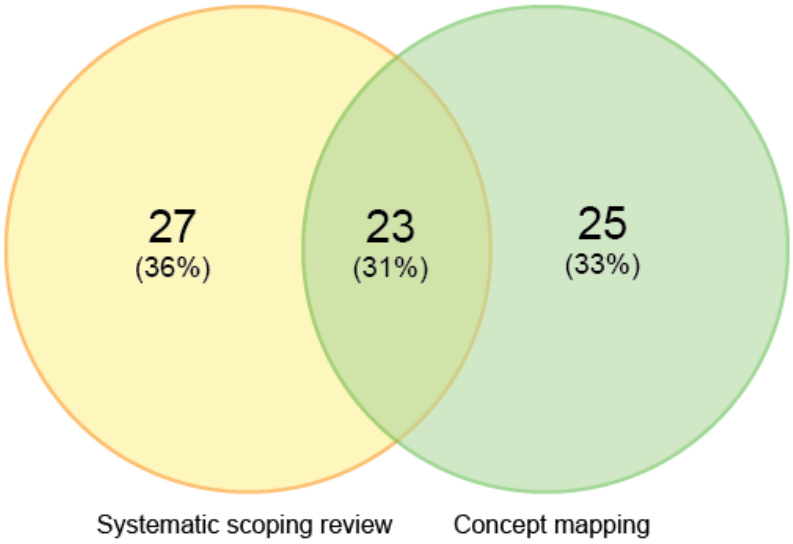


Table 3. Results of step 2: concept mapping study.

Evaluation approach[a]	Use of approach[b], % "yes" response	Familiarity with approach[c]		Proving effectiveness[d]	
		Mean	% of 3 + 4 (n/N)	Mean	% of 3 + 4 (n/N)
Pilot/feasibility	58 (SD 32.7)	2.9 (SD 0.5)		2.3 (SD 0.3)	
3. Feasibility study[e]	94	3.6	88 (28/42)	2.6	52 (16/31)
4. Questionnaire[e]	100	3.4	84 (27/63)	2.5	52 (16/31)
8. Single-case experiments or n-of-1 study (N=1)	28	2.5	43 (13/60)	2.0	27 (8/30)
12. Action research study	41	2.6	50 (15/58)	2.3	38 (11/29)
44. A/B testing	25	2.5	45 (13/58)	2.2	36 (10/28)
Development and usability	37 (SD 29.1)	2.5 (SD 0.4)		2.1 (SD 0.3)	
5. Focus group (interview)	91	3.2	81 (26/62)	2.3	32 (10/31)
6. Interview	94	3.1	75 (24/62)	2.3	35 (11/31)
23. Think-aloud method	66	2.6	52 (15/59)	1.7	14 (4/29)
25. Cognitive walkthrough	31	2.4	37 (11/59)	1.8	17 (5/30)
27. eHealth[f] Analysis and Steering Instrument	12	2.4	55 (16/58)	2.4	48 (14/29)
28. Model for Assessment of Telemedicine applications (MAST)	22	2.5	48 (14/59)	2.4	37 (11/30)
29. Rapid review	31	2.0	23 (7/58)	1.8	7 (2/29)
30. eHealth Needs Assessment Questionnaire (ENAQ)	6	2.4	45 (13/58)	2.0	24 (7/29)
31. Evaluative Questionnaire for eHealth Tools (EQET)	3	2.4	52 (15/58)	2.3	41 (12/29)
32. Heuristic evaluation	19	2.2	31 (9/57)	2.1	24 (7/29)
33. Critical incident technique	9	2.0	24 (7/59)	1.8	4 (1/28)
36. Systematic review[e]	94	3.1	67 (20/62)	2.9	69 (20/29)
39. User-centered design methods[e]	53	3.2	73 (22/62)	2.5	50 (14/28)
43. Vignette study	41	2.2	31 (9/58)	1.6	7 (2/28)
45. Living lab	34	2.5	41 (12/58)	2.3	54 (15/28)
50. Method for technology-delivered health care measures	9	2.3	39 (11/58)	2.1	25 (7/28)
54. Cognitive task analysis (CTA)	16	2.1	23 (7/59)	1.9	18 (5/28)
60. Simulation study	41	2.5	50 (15/60)	2.2	34 (10/29)
62. Sociotechnical evaluation	22	2.3	37 (11/60)	2.1	29 (8/28)
All phases	11 (SD 4)	2.3 (SD 0.2)		2.2 (SD 0.2)	
21. Multiphase Optimization Strategy (MOST)	6	2.3	45 (13/58)	2.3	39 (11/28)
26. Continuous evaluation of evolving behavioral intervention technologies (CEEBIT) framework	6	2.4	48 (14/60)	2.3	38 (11/29)
40. RE-AIM[g] framework[e]	19	2.6	61 (17/59)	2.4	52 (14/27)
46. Normalization process model	9	2.0	25 (7/57)	1.9	18 (5/28)
48. CeHRes[h] Roadmap	16	2.4	43 (12/58)	2.3	41 (11/27)

Evaluation approach[a]	Use of approach[b], % "yes" response	Familiarity with approach[c]		Proving effectiveness[d]	
		Mean	% of 3 + 4 (n/N)	Mean	% of 3 + 4 (n/N)
49. Stead et al [82] evaluation framework	12	2.2	38 (11/58)	2.1	22 (6/27)
51. CHEATS[i]: a generic information communication technology evaluation framework	6	2.3	41 (12/58)	2.1	26 (7/27)
52. Stage Model of Behavioral Therapies Research	9	1.9	21 (6/58)	2.0	22 (6/27)
53. Life cycle–based approach to evaluation	12	2.3	45 (13/58)	2.0	21 (6/28)
Effectiveness testing	45 (SD 23)	2.6 (SD 0.3)		2.6 (0.4)	
1. Mixed methods[e]	87	3.2	81 (26/63)	2.9	65 (20/31)
2. Pragmatic randomized controlled trial[e]	62	3.1	77 (24/63)	3.3	83 (25/30)
7. Cohort study[e] (retrospective and prospective)	81	2.7	58 (18/61)	2.5	58 (18/31)
9. Randomized controlled trial[e]	91	3.3	71 (22/63)	3.3	74 (23/31)
10. Crossover study[e]	44	2.7	57 (17/61)	2.7	59 (17/29)
11. Case series	50	2.1	20 (6/60)	1.8	10 (3/29)
13. Pretest-posttest study design[e]	62	2.6	45 (14/60)	2.5	50 (15/30)
14. Interrupted time-series study	44	2.5	43 (13/59)	2.7	59 (17/29)
15. Nested randomized controlled trial	31	2.3	37 (11/59)	2.8	55 (16/29)
16. Stepped wedge trial design[e]	56	2.8	70 (21/60)	3.2	90 (26/29)
17. Cluster randomized controlled trial[e]	50	2.8	60 (18/60)	3.1	69 (20/29)
19. Trials of intervention principles (TIPs)[e]	23	2.5	42 (13/61)	2.5	43 (13/30)
20. Sequential Multiple Assignment Randomized Trial (SMART)	9	2.4	45 (13/58)	2.7	62 (18/29)
35. (Fractional-)factorial design	22	2.3	45 (13/58)	2.2	36 (10/28)
37. Controlled before-after study (CBA)[e]	37	2.6	50 (15/60)	2.4	52 (15/29)
38. Controlled clinical trial /nonrandomized controlled trial (CCT/NRCT)[e]	47	2.9	70 (21/60)	2.9	71 (20/28)
41. Preference clinical trial (PCT)	19	2.1	24 (7/58)	2.1	25 (7/28)
42. Microrandomized trial	9	2.2	24 (7/59)	2.4	50 (14/28)
55. Cross-sectional study	72	2.5	40 (12/60)	2.1	29 (8/28)
56. Matched cohort study	37	2.2	30 (9/59)	2.3	46 (13/28)
57. Noninferiority trial design[e]	53	2.6	47 (14/60)	2.6	48 (14/29)
58. Adaptive design[e]	19	2.6	52 (15/58)	2.5	50 (14/28)
59. Waitlist control group design	34	2.1	28 (8/59)	2.0	32 (9/28)
61. Propensity score methodology	31	2.1	30 (9/59)	2.0	21 (6/29)

Evaluation approach[a]	Use of approach[b], % "yes" response	Familiarity with approach[c]		Proving effectiveness[d]	
		Mean	% of 3 + 4 (n/N)	Mean	% of 3 + 4 (n/N)
Implementation	54 (SD 28)	2.8 (SD 0.5)		2.6 (SD 0.5)	
18. Cost-effectiveness analysis	81	3.4	87 (27/63)	3.2	70 (21/30)
22. Methods comparison study	16	2.0	17 (5/59)	2.0	21 (6/28)
24. Patient reported outcome measures (PROMs)[e]	84	3.1	80 (24/60)	2.9	73 (22/30)
34. Transaction logfile analysis	25	2.4	45 (13/57)	2.1	21 (6/28)
47. Big data analysis[e]	62	3.0	73 (22/61)	2.8	59 (17/29)

[a]Approach identification numbers correspond with the numbers used in Figure 3 and Figure 4.

[b]Based on the rating question: "does your research group use this approach, or did it do so in the past?"; the percentage of "yes" responses is shown.

[c]Based on the rating question: "according to your opinion, how important is it that researchers with an interest in eHealth will become familiar with this approach?"; average rating scores ranging from unimportant (1) to absolutely essential (4) and percentages of categories 3 plus 4 are represented.

[d]The "proving effectiveness" column corresponds with the rating question: "according to your opinion, how important is the approach for proving the effectiveness of eHealth?" Average rating scores ranging from unimportant (1) to absolutely essential (4) and percentages of categories 3 plus 4 are presented.

[e]This approach scored above average on the rating questions "familiarity with the approach" and "proving effectiveness, " which is plotted in the upper right quadrant of the Go-Zone graph (Figure 3).

[f]eHealth: electronic health.

[g]RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance.

[h]CeHRes: Centre for eHealth Research and Disease management.

[i]CHEATS: Clinical, human and organizational, educational, administrative, ethnical and social explanatory factors in a randomized controlled trial intervention.

Figure 3. Go-Zone graph. The numbers refer to the evaluation approaches listed in Table 3.

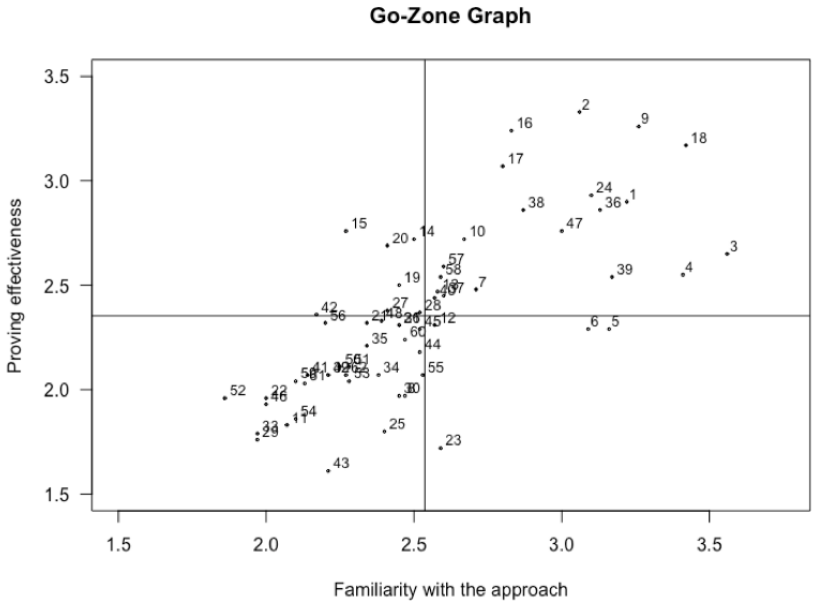
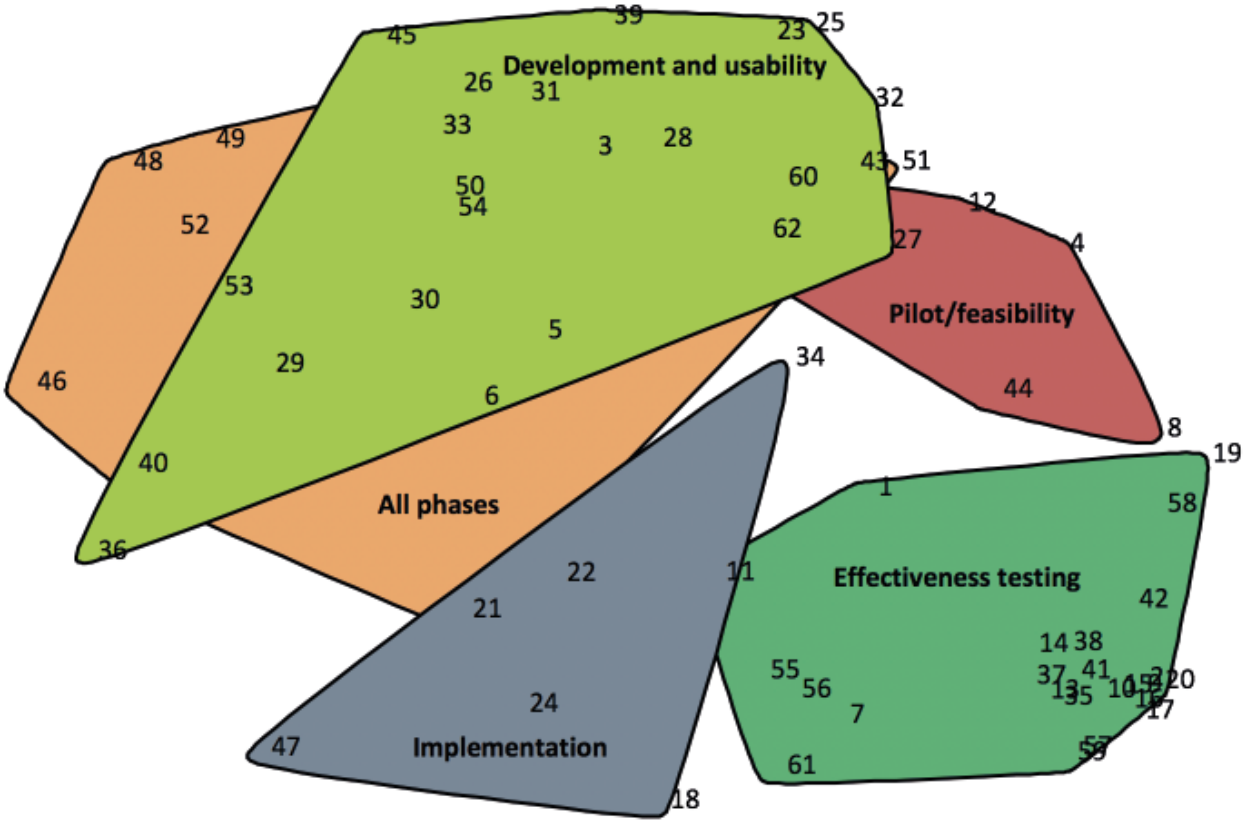


Figure 4. Concept map showing evaluation approaches grouped into five labeled clusters. The numbers refer to the approaches listed in Table 3.



Concept Mapping Analysis

Based on sorting data from 27 participants, a point map with a stress value of 0.27 was created. Compared with previous concept mapping study stress values, this represents a good fit [90,91]. In the next step, the software automatically clustered the points into the clusters shown on the concept map in Figure 4. A 5-cluster concept map was judged to represent the best fit for aggregating similar evaluation approaches into one cluster. Table 3 lists these clusters with average rating scores for the three rating questions and the approaches belonging in each cluster. With an average score of 2.9, the pilot/feasibility cluster showed the highest score on the familiarity with approach scale, whereas the "all phases" cluster showed the lowest average score at 2.3. With respect to responses to the importance for proving effectiveness question, the implementation cluster presented the highest average score at 2.6 and the development and usability cluster presented the lowest average score at 2.1.

Twenty of the 62 methods (32%) received above-average scores for both the questions related to familiarity with the approach and importance for proving effectiveness, and therefore appear in the upper right quadrant of the Go-Zone graph (Figure 3) and are indicated in Table 3. The majority of these approaches (12/20, 60%) fall into the effectiveness testing cluster.

Interpretation and Utilization of the Concept Mapping Study

The results of the concept map study were discussed within the team and the following names for the clusters were selected: "Development and usability," "Pilot/feasibility," "Effectiveness testing," "Implementation," and "All phases."

Step 3: eHealth Methodology Guide

Fifty evaluation approaches were identified in the systematic scoping review and 48 approaches were described by participants in the brainstorm phase of the concept mapping study. As visualized in the Venn diagram (Figure 2), 23 approaches were identified in both studies. Therefore, in total, 75 (50 + 48 – 23) unique evaluation approaches were identified. Examining the 23 approaches identified in both the literature and concept maps, 14 (67%) were described by more than one article.

Based on the cluster names from the concept map (Figure 4), development and usability, pilot/feasibility, effectiveness testing, implementation, and the all phases evaluation approaches found in the systematic scoping review, an empirically based "eHealth evaluation cycle" was developed (Figure 5). The concept map did not reveal a conceptual and planning phase; however, based on the results of the systematic scoping review, and since there are evaluation approaches that belong to this phase, it was added to the "eHealth evaluation cycle."

This evaluation cycle is iterative with consecutive evaluation study phases and an "all phases" cluster in the middle, which includes "all phases" evaluation frameworks such as Model for Assessment of Telemedecine that are capable of evaluating multiple study phases [65]. The "eHealth evaluation cycle" was used to construct the "eHealth methodology guide" by subdividing the guide into the evaluation study phase themes. Within the guide, each of the 75 unique evaluation approaches are briefly described and allocated to their respective evaluation study phase(s). Note that a single evaluation approach may belong to multiple evaluation phases.

The "eHealth methodology guide" can be found in Multimedia Appendix 3 and is available online [92]. Because the "eHealth methodology guide" is web-based, it is easy to maintain and, more importantly, it is easy to add content as new evaluation approaches may be proposed.

Figure 5. The "eHealth evaluation cycle" derived from empirical results of the scoping literature review and concept map study.

Discussion

Principal Findings

By carrying out a systematic scoping review and concept mapping study with eHealth researchers, we identified and aggregated 75 unique evaluation approaches into an online "eHealth methodology guide." This online guide supports researchers in the field of eHealth to identify the appropriate study phase of the research cycle and choose an evaluation approach that is suitable for each particular study phase.

As stipulated by the participants in the concept mapping study, the most frequently used eHealth evaluation approaches were questionnaire (100%) and feasibility study (88%). The participants were most familiar with cost-effectiveness analysis (87%) and feasibility study (84%). In addition, they found pragmatic RCT (83%) and the stepped wedge trial design (90%) to be the most suitable approaches for proving effectiveness in eHealth research. Although a wide array of alternative evaluation approaches are already available, well-known traditional evaluation approaches, including all of the evaluation approaches described above, seemed to be most relevant for the participants. This suggests that eHealth research is still an immature field with too much focus on traditional evaluation approaches. However, to facilitate long-term implementation and safe use of novel eHealth solutions, evaluations performed by less-known evaluation approaches such as those described in the online "eHealth evaluation guide" are required.

The Go-Zone graph (Figure 3) confirms the practicing researchers' familiarity with—and judged importance for proving the effectiveness of—the traditional evaluation approaches. The majority of the 20 approaches in the upper right quadrant of this graph are well-known study designs such as cohort study, (pragmatic) RCT, and controlled before-after study. Alternative and novel study designs (eg, instrumental variable analysis, interrupted time-series analysis) were not mentioned in the upper right quadrant, possible due to unfamiliarity.

Comparison with Previous Work

Ekeland et al [93] performed a systematic review of reviews to summarize methodologies used in telemedicine research, analyze knowledge gaps, and suggest methodological recommendations for further research. They assessed and extracted data from 50 reviews and performed a qualitative summary and analysis of methodologies. They recommended that larger and more rigorous controlled studies are needed, including standardization of methodological aspects, to produce better evidence for the effectiveness of telemedicine. This is in line with our study, which provides easy access to, and an overview of, current approaches for eHealth evaluation throughout the research cycle. However, our work extends beyond effectiveness to cover the

many other questions arising when developing and implementing eHealth tools. Aldossary et al [94] also performed a review to identify evaluations of deployed telemedicine services in hospitals, and to report methods used to evaluate service implementation. The authors included 164 papers describing 137 studies in the qualitative synthesis. They showed that 83 of the 137 studies used a descriptive evaluation methodology to report information about their activities, and 27 of the 137 studies evaluated clinical outcomes by the use of "traditional" study designs such as nonrandomized open intervention studies. Although the authors also reported methods to evaluate implementation, an overview of all evaluation study phases was lacking. In addition, no suggestions for alternative evaluation approaches were provided. Enam et al [27] developed an evaluation model consisting of multiple evaluation phases. The authors conducted a literature review to elucidate how the evidence of effectiveness and efficiency of eHealth can be generated through evaluation. They emphasized that generation of robust evidence of effectiveness and efficiency would be plausible when the evaluation is conducted through all distinct phases of eHealth intervention development (design, pretesting, pilot study, pragmatic trial, evaluation, and postintervention). This is partially in line with our study aim, and matches the "eHealth evaluation cycle" and online "eHealth methodology guide" developed as a result of our study. However, we added specific evaluation approaches to be used for each study phase and also incorporated other existing "all phases" research models.

Strengths and Limitations

One of the greater strengths of this study was the combination of the scoping review and concept mapping study. The scoping review focused on finding eHealth-specific evaluation approaches. In contrast, in the concept mapping study, the participants were asked to write down any approach they were aware of that could contribute to the evaluation of eHealth. This slight discrepancy was intentional because we particularly wanted to find evaluation approaches that are actually being used in daily research practice to evaluate eHealth solutions. Therefore, the results from the systematic scoping review and the concept mapping study complement and reinforce each other, and therewith contribute to delivering a complete as possible "eHealth methodology guide."

Another strength of this project was the level of knowledge and experience of the eHealth researchers who participated in the concept mapping study. They had approximately 13 years of eHealth research experience and the majority of participants graded themselves high for knowledge about eHealth. Interestingly, they gave themselves lower grades for their expertise in eHealth evaluation approaches. This means that we likely included an average group of eHealth researchers and did not only include the top researchers in the field of eHealth methodology. In our view, we had a representative sample of average eHealth researchers, who are also the target end users for our online "eHealth methodology guide." This supports the generalizability and implementability of our project. However, the fact that more than 70% of participants worked in university medical centers may slightly limit the generalizability of our work to nonacademic researchers. It would be wise to keep an eye out for positive deviants outside university medical centers and users that are not senior academic "expert" eHealth researchers [95]. Slight wandering off the beaten track might be very necessary to find the needed innovative evaluation approaches and dissemination opportunities for sustainable implementation.

A limitation of our study was the date restriction of the systematic scoping review. We performed a broad systematic search but limited the search to only English language articles published from January 1, 2006 so as to keep the number of articles manageable. This could explain why some approaches, especially those published before 2006, were not found.

Another weakness of our study was that the systematic search was updated after the concept mapping exercise was complete. Therefore, 13 of the 75 evaluation approaches were not reviewed by the participants in the sorting and rating phases of the concept mapping study. However, this will also occur in the future with every new approach added to the online "eHealth methodology guide," as the aim is to frequently update the guide.

Future Perspectives

This first version of the "eHealth evaluation guide" contains short descriptions of the 75 evaluation approaches and references describing the approaches in more detail. Our aim is to include information on the level of complexity in the following version and other relevant resource requirements. Moreover, case example references will be added to the evaluation approaches to support the user in selecting an appropriate approach. Further, in the coming years, we aim to subject the "eHealth methodology guide" to an expert evaluation to assess the quality and ranking of the evaluation approaches, since this was not part of this present study. Finally, we are discussing collaboration and integration with the European Federation for Medical Informatics EVAL-Assessment of Health Information Systems working group.

Conclusion

In this project, 75 unique eHealth evaluation approaches were identified in a scoping review and concept mapping study and served as content for the online "eHealth methodology guide." The online "eHealth methodology guide" could be a step forward in supporting developers and evaluators in selecting a suitable evaluation approach in relation to the specific study phase of the "eHealth evaluation cycle." Overall, the guide aims to enhance quality and safety, and to facilitate long-term implementation of novel eHealth solutions.

Acknowledgments

We thank the following individual study participants of the eHealth Evaluation Research Group for their contributions to the concept mapping study: AM Hooghiemstra, ASHM van Dalen, DT Ubbink, E Tensen, HAW Meijer, H Ossebaard, IM Verdonck,

JK Sont, J Breedvelt, JFM van den Heuvel, L Siemons, L Wesselman, MJM Breteler, MJ Schuuring, M Jansen, MMH Lahr, MM van der Vlist, NF Keizer, P Kubben, PM Bossuyt, PJM van den Boog, RB Kool, VT Visch, and WA Spoelman. We would like to acknowledge the Netherlands Organization for Health Research and Development (ZonMw) and the Netherlands Federation of University Medical Centres for their financial support through the means of the "Citrienfund - program eHealth" (grant number 839201005). We also acknowledge Terralemon for development and support of the online "eHealth methodology guide."

Authors' Contributions

TB, AR, MS, MK, and NC designed the study. TB, AR, and MK performed the systematic scoping review. AR set up the online concept mapping software, invited participants, and coordinated data collection. TB, AR, JW, MK, LW, HR, LGP, MS, and NC engaged, alongside the eHealth Evaluation Collaborators Group, in the exercises of the concept mapping study. TB, AR, and JW analyzed data and interpreted the study results. TB and AR wrote the first draft. AR created the tables and figures. TB, AR, JW, MK, HR, LGP, KC, AS, MS, and NC contributed to the redrafting of the manuscript. All authors approved the final version of the manuscript for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 13 KB - jmir_v22i8e17774_app1.docx]

Multimedia Appendix 2
List of 48 unique electronic health (eHealth) evaluation approaches suggested by participants of the concept mapping study.
[DOCX File , 13 KB - jmir_v22i8e17774_app2.docx]

Multimedia Appendix 3
eHealth methodology guide.
[DOCX File , 288 KB - jmir_v22i8e17774_app3.docx]

References

1. From innovation to implementation: eHealth in the WHO European Region. World Health Organization. 2016. URL: http://www.euro.who.int/__data/assets/pdf_file/0012/302331/From-Innovation-to-Implementation-eHealth-Report-EU.pdf?ua=1 [accessed 2020-01-01]
2. de la Torre-Díez I, López-Coronado M, Vaca C, Aguado JS, de Castro C. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. Telemed J E Health 2015 Feb;21(2):81-85 [FREE Full text] [doi: 10.1089/tmj.2014.0053] [Medline: 25474190]
3. Sanyal C, Stolee P, Juzwishin D, Husereau D. Economic evaluations of eHealth technologies: A systematic review. PLoS One 2018;13(6):e0198112 [FREE Full text] [doi: 10.1371/journal.pone.0198112] [Medline: 29897921]
4. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015 Sep 07(9):CD002098 [FREE Full text] [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
5. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
6. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
7. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6:27882 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
8. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
9. Sheikh A, Cornford T, Barber N, Avery A, Takian A, Lichtner V, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in "early adopter" hospitals. BMJ 2011 Oct 17;343:d6054 [FREE Full text] [doi: 10.1136/bmj.d6054] [Medline: 22006942]
10. Scott RE, Mars M. Principles and framework for eHealth strategy development. J Med Internet Res 2013 Jul 30;15(7):e155 [FREE Full text] [doi: 10.2196/jmir.2250] [Medline: 23900066]

11. Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. PLoS Med 2008 Mar 11;5(3):e67 [FREE Full text] [doi: 10.1371/journal.pmed.0050067] [Medline: 18336067]
12. Brender J. Evaluation of health information applications--challenges ahead of us. Methods Inf Med 2006;45(1):62-66. [Medline: 16482372]
13. Kaplan B. Evaluating informatics applications--some alternative approaches: theory, social interactionism, and call for methodological pluralism. Int J Med Inform 2001 Nov;64(1):39-56. [doi: 10.1016/s1386-5056(01)00184-8] [Medline: 11673101]
14. Chumbler NR, Kobb R, Brennan DM, Rabinowitz T. Recommendations for research design of telehealth studies. Telemed J E Health 2008 Nov;14(9):986-989. [doi: 10.1089/tmj.2008.0108] [Medline: 19035813]
15. de Lusignan S, Crawford L, Munro N. Creating and using real-world evidence to answer questions about clinical effectiveness. J Innov Health Inform 2015 Nov 04;22(3):368-373. [doi: 10.14236/jhi.v22i3.177] [Medline: 26577427]
16. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of Intervention Principles: Evaluation Methods for Evolving Behavioral Intervention Technologies. J Med Internet Res 2015 Jul 08;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
17. Riley WT, Glasgow RE, Etheredge L, Abernethy AP. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. Clin Transl Med 2013 May 10;2(1):10. [doi: 10.1186/2001-1326-2-10] [Medline: 23663660]
18. Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? BMC Med 2018 Dec 03;16(1):225 [FREE Full text] [doi: 10.1186/s12916-018-1211-7] [Medline: 30501638]
19. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
20. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating Digital Health Interventions: Key Questions and Approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
21. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
22. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
23. Nykänen P, Brender J, Talmon J, de Keizer N, Rigby M, Beuscart-Zephir M, et al. Guideline for good evaluation practice in health informatics (GEP-HI). Int J Med Inform 2011 Dec;80(12):815-827. [doi: 10.1016/j.ijmedinf.2011.08.004] [Medline: 21920809]
24. Nykänen P, Kaipio J. Quality of Health IT Evaluations. Stud Health Technol Inform 2016;222:291-303. [Medline: 27198111]
25. Brender J. Handbook of Evaluation Methods for Health Informatics. Cambridge, MA: Academic Press/Elsevier; 2006.
26. Ammenwerth E, Rigby M. Evidence-Based Health Informatics. In: Studies in Health Technology and Informatics. Amsterdam: IOS press; 2016.
27. Enam A, Torres-Bonilla J, Eriksson H. Evidence-Based Evaluation of eHealth Interventions: Systematic Literature Review. J Med Internet Res 2018 Nov 23;20(11):e10971 [FREE Full text] [doi: 10.2196/10971] [Medline: 30470678]
28. Trochim WM, Linton R. Conceptualization for planning and evaluation. Eval Program Plann 1986;9(4):289-308. [doi: 10.1016/0149-7189(86)90044-3] [Medline: 10301179]
29. Trochim WM. An introduction to concept mapping for planning and evaluation. Eval Program Plan 1989 Jan;12(1):1-16. [doi: 10.1016/0149-7189(89)90016-5]
30. Concept Systems Incorporated. Global MAXTM. URL: http://www.conceptsystems.com [accessed 2019-03-06]
31. Trochim WM, McLinden D. Introduction to a special issue on concept mapping. Eval Program Plann 2017 Feb;60:166-175. [doi: 10.1016/j.evalprogplan.2016.10.006] [Medline: 27780609]
32. Group Concept Mapping Resource Guide. groupwisdom. URL: https://conceptsystems.com/GCMRG [accessed 2019-01-16]
33. Kane M, Trochim W. Concept Mapping for Planning and Evaluation. Thousand Oaks, CA: Sage Publications Inc; 2007.
34. Chiasson M, Reddy M, Kaplan B, Davidson E. Expanding multi-disciplinary approaches to healthcare information technologies: what does information systems offer medical informatics? Int J Med Inform 2007 Jun;76(Suppl 1):S89-S97. [doi: 10.1016/j.ijmedinf.2006.05.010] [Medline: 16769245]
35. Campbell G, Yue LQ. Statistical innovations in the medical device world sparked by the FDA. J Biopharm Stat 2016 Sep 15;26(1):3-16. [doi: 10.1080/10543406.2015.1092037] [Medline: 26372890]
36. Law LM, Wason JMS. Design of telehealth trials--introducing adaptive approaches. Int J Med Inform 2014 Dec;83(12):870-880 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.09.002] [Medline: 25293533]
37. van Gemert-Pijnen JEWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]

38. Alpay L, Doms R, Bijwaard H. Embedding persuasive design for self-health management systems in Dutch healthcare informatics education: Application of a theory-based method. Health Informatics J 2019 Dec;25(4):1631-1646. [doi: 10.1177/1460458218796642] [Medline: 30192696]
39. Shaw NT. 'CHEATS': a generic information communication technology (ICT) evaluation framework. Comput Biol Med 2002 May;32(3):209-220. [doi: 10.1016/s0010-4825(02)00016-1]
40. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76 [FREE Full text] [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
41. Jaspers MWM. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
42. Khajouei R, Zahiri Esfahani M, Jahani Y. Comparison of heuristic and cognitive walkthrough usability evaluation methods for evaluating health information systems. J Am Med Inform Assoc 2017 Apr 01;24(e1):e55-e60. [doi: 10.1093/jamia/ocw100] [Medline: 27497799]
43. van Engen-Verheul M, Peek N, Vromen T, Jaspers M, de Keizer N. How to use concept mapping to identify barriers and facilitators of an electronic quality improvement intervention. Stud Health Technol Inform 2015;210:110-114. [Medline: 25991112]
44. Mohr DC, Cheung K, Schueller SM, Hendricks Brown BC, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523 [FREE Full text] [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]
45. Nicholas J, Boydell K, Christensen H. mHealth in psychiatry: time for methodological change. Evid Based Ment Health 2016 May;19(2):33-34. [doi: 10.1136/eb-2015-102278] [Medline: 27044849]
46. Bongiovanni-Delarozière I, Le Goff-Pronost M. Economic evaluation methods applied to telemedicine: From a literature review to a standardized framework. Eur Res Telemed 2017 Nov;6(3-4):117-135. [doi: 10.1016/j.eurtel.2017.08.002]
47. Fatehi F, Smith AC, Maeder A, Wade V, Gray LC. How to formulate research questions and design studies for telehealth assessment and evaluation. J Telemed Telecare 2017 Oct;23(9):759-763. [doi: 10.1177/1357633X16673274] [Medline: 29070001]
48. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014 Feb 19;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
49. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(Suppl 5):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
50. Grigsby J, Bennett RE. Alternatives to randomized controlled trials in telemedicine. J Telemed Telecare 2006;12(Suppl 2):S77-S84. [doi: 10.1258/135763306778393162] [Medline: 16989679]
51. Liu JLY, Wyatt JC. The case for randomized controlled trials to assess the impact of clinical information systems. J Am Med Inform Assoc 2011;18(2):173-180 [FREE Full text] [doi: 10.1136/jamia.2010.010306] [Medline: 21270132]
52. Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. BMJ 2015 Jun 09;350:h2750 [FREE Full text] [doi: 10.1136/bmj.h2750] [Medline: 26058820]
53. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009 Aug;6(8):e1000126 [FREE Full text] [doi: 10.1371/journal.pmed.1000126] [Medline: 19688038]
54. Han JY. Transaction logfile analysis in health communication research: challenges and opportunities. Patient Educ Couns 2011 Mar;82(3):307-312. [doi: 10.1016/j.pec.2010.12.018] [Medline: 21277146]
55. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the Black Box of Electronic Health: Collecting, Analyzing, and Interpreting Log Data. JMIR Res Protoc 2017 Aug 07;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]
56. Kramer-Jackman KL, Popkess-Vawter S. Method for technology-delivered healthcare measures. Comput Inform Nurs 2011 Dec;29(12):730-740. [doi: 10.1097/NCN.0b013e318224b581] [Medline: 21694585]
57. Wilson K, Bell C, Wilson L, Witteman H. Agile research to complement agile development: a proposal for an mHealth research lifecycle. NPJ Digit Med 2018 Sep 13;1(1):46 [FREE Full text] [doi: 10.1038/s41746-018-0053-1] [Medline: 31304326]
58. Jacobs MA, Graham AL. Iterative development and evaluation methods of mHealth behavior change interventions. Curr Opin Psychol 2016 Jun;9:33-37. [doi: 10.1016/j.copsyc.2015.09.001]
59. Dempsey W, Liao P, Klasnja P, Nahum-Shani I, Murphy SA. Randomised trials for the Fitbit generation. Signif (Oxf) 2015 Dec 01;12(6):20-23 [FREE Full text] [doi: 10.1111/j.1740-9713.2015.00863.x] [Medline: 26807137]
60. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
61. Law L, Edirisinghe N, Wason JM. Use of an embedded, micro-randomised trial to investigate non-compliance in telehealth interventions. Clin Trials 2016 Aug;13(4):417-424. [doi: 10.1177/1740774516637075] [Medline: 26968939]

62. Walton A, Nahum-Shani I, Crosby L, Klasnja P, Murphy S. Optimizing Digital Integrated Care via Micro-Randomized Trials. Clin Pharmacol Ther 2018 Jul 19;104(1):53-58 [FREE Full text] [doi: 10.1002/cpt.1079] [Medline: 29604043]
63. Caffery LJ, Martin-Khan M, Wade V. Mixed methods for telehealth research. J Telemed Telecare 2017 Oct;23(9):764-769. [doi: 10.1177/1357633X16665684] [Medline: 27591744]
64. Lee S, Smith CAM. Criteria for quantitative and qualitative data integration: mixed-methods research methodology. Comput Inform Nurs 2012 May;30(5):251-256. [doi: 10.1097/NXN.0b013e31824b1f96] [Medline: 22411415]
65. Kidholm K, Clemensen J, Caffery LJ, Smith AC. The Model for Assessment of Telemedicine (MAST): A scoping review of empirical studies. J Telemed Telecare 2017 Oct;23(9):803-813. [doi: 10.1177/1357633X17721815] [Medline: 28758525]
66. Kidholm K, Jensen LK, Kjølhede T, Nielsen E, Horup MB. Validity of the Model for Assessment of Telemedicine: A Delphi study. J Telemed Telecare 2018 Feb;24(2):118-125. [doi: 10.1177/1357633X16686553] [Medline: 28024442]
67. Kummervold PE, Johnsen JK, Skrøvseth SO, Wynn R. Using noninferiority tests to evaluate telemedicine and e-health services: systematic review. J Med Internet Res 2012 Sep 28;14(5):e132 [FREE Full text] [doi: 10.2196/jmir.2169] [Medline: 23022989]
68. May C. A rational model for assessing and evaluating complex interventions in health care. BMC Health Serv Res 2006 Jul 07;6:86 [FREE Full text] [doi: 10.1186/1472-6963-6-86] [Medline: 16827928]
69. Borycki E, Dexheimer JW, Hullin Lucay Cossio C, Gong Y, Jensen S, Kaipio J, et al. Methods for Addressing Technology-induced Errors: The Current State. Yearb Med Inform 2016 Nov 10(1):30-40 [FREE Full text] [doi: 10.15265/IY-2016-029] [Medline: 27830228]
70. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2017 Oct;23(9):780-785. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
71. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(Suppl 5):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
72. Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug;38(1):28-39 [FREE Full text] [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]
73. Sadegh SS, Khakshour Saadat P, Sepehri MM, Assadi V. A framework for m-health service development and success evaluation. Int J Med Inform 2018 Apr;112:123-130. [doi: 10.1016/j.ijmedinf.2018.01.003] [Medline: 29500009]
74. Harker J, Kleijnen J. What is a rapid review? A methodological exploration of rapid reviews in Health Technology Assessments. Int J Evid Based Healthc 2012 Dec;10(4):397-410. [doi: 10.1111/j.1744-1609.2012.00290.x] [Medline: 23173665]
75. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. Int J Med Inform 2014 Jul;83(7):e1-e11. [doi: 10.1016/j.ijmedinf.2013.07.002] [Medline: 23910896]
76. Almirall D, Nahum-Shani I, Sherwood NE, Murphy SA. Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. Transl Behav Med 2014 Sep;4(3):260-274 [FREE Full text] [doi: 10.1007/s13142-014-0265-0] [Medline: 25264466]
77. Ammenwerth E, Hackl WO, Binzer K, Christoffersen TEH, Jensen S, Lawton K, et al. Simulation studies for the evaluation of health information technologies: experiences and results. Health Inf Manag 2012 Jun;41(2):14-21. [doi: 10.1177/183335831204100202] [Medline: 22700558]
78. Jensen S, Kushniruk AW, Nøhr C. Clinical simulation: A method for development and evaluation of clinical information systems. J Biomed Inform 2015 Apr;54:65-76 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.002] [Medline: 25684129]
79. Dallery J, Cassidy RN, Raiff BR. Single-case experimental designs to evaluate novel technology-based health interventions. J Med Internet Res 2013 Feb 08;15(2):e22 [FREE Full text] [doi: 10.2196/jmir.2227] [Medline: 23399668]
80. Cresswell KM, Sheikh A. Undertaking sociotechnical evaluations of health information technologies. Inform Prim Care 2014 Mar 18;21(2):78-83. [doi: 10.14236/jhi.v21i2.54] [Medline: 24841408]
81. Kaufman D, Roberts WD, Merrill J, Lai T, Bakken S. Applying an evaluation framework for health information system design, development, and implementation. Nurs Res 2006;55(Suppl 2):S37-S42. [doi: 10.1097/00006199-200603001-00007] [Medline: 16601633]
82. Stead M, Hastings G, Eadie D. The challenge of evaluating complex interventions: a framework for evaluating media advocacy. Health Educ Res 2002 Jun;17(3):351-364. [doi: 10.1093/her/17.3.351] [Medline: 12120850]
83. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC Med Res Methodol 2006 Nov 08;6:54 [FREE Full text] [doi: 10.1186/1471-2288-6-54] [Medline: 17092344]
84. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007 Feb;28(2):182-191. [doi: 10.1016/j.cct.2006.05.007] [Medline: 16829207]
85. Spiegelman D. Evaluating Public Health Interventions: 2. Stepping Up to Routine Public Health Evaluation With the Stepped Wedge Design. Am J Public Health 2016 Mar;106(3):453-457. [doi: 10.2105/AJPH.2016.303068] [Medline: 26885961]
86. Langbecker D, Caffery LJ, Gillespie N, Smith AC. Using survey methods in telehealth research: A practical guide. J Telemed Telecare 2017 Oct;23(9):770-779. [doi: 10.1177/1357633X17721814] [Medline: 28728502]

87. Rönnby S, Lundberg O, Fagher K, Jacobsson J, Tillander B, Gauffin H, et al. mHealth Self-Report Monitoring in Competitive Middle- and Long-Distance Runners: Qualitative Study of Long-Term Use Intentions Using the Technology Acceptance Model. JMIR Mhealth Uhealth 2018 Aug 13;6(8):e10270 [FREE Full text] [doi: 10.2196/10270] [Medline: 30104183]
88. Bastien JMC. Usability testing: a review of some methodological and technical aspects of the method. Int J Med Inform 2010 Apr;79(4):e18-e23. [doi: 10.1016/j.ijmedinf.2008.12.004] [Medline: 19345139]
89. Nguyen HQ, Cuenco D, Wolpin S, Benditt J, Carrieri-Kohlman V. Methodological considerations in evaluating eHealth interventions. Can J Nurs Res 2007 Mar;39(1):116-134. [Medline: 17450708]
90. Brennan LK, Brownson RC, Kelly C, Ivey MK, Leviton LC. Concept mapping: priority community strategies to create changes to support active living. Am J Prev Med 2012 Nov;43(5 Suppl 4):S337-S350 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.015] [Medline: 23079266]
91. van Engen-Verheul MM, Peek N, Haafkens JA, Joukes E, Vromen T, Jaspers MWM, et al. What is needed to implement a web-based audit and feedback intervention with outreach visits to improve care quality: A concept mapping study among cardiac rehabilitation teams. Int J Med Inform 2017 Jan;97:76-85. [doi: 10.1016/j.ijmedinf.2016.10.003] [Medline: 27919398]
92. Online eHealth methodology guide. NFU eHealth. URL: https://www.citrienfonds-ehealth.nl/e-health-toolkit/onderzoek/e-health-evaluation-methodology/overview-of-methods [accessed 2020-01-01]
93. Ekeland AG, Bowes A, Flottorp S. Methodologies for assessing telemedicine: a systematic review of reviews. Int J Med Inform 2012 Jan;81(1):1-11. [doi: 10.1016/j.ijmedinf.2011.10.009] [Medline: 22104370]
94. AlDossary S, Martin-Khan MG, Bradford NK, Smith AC. A systematic review of the methodologies used to evaluate telemedicine service initiatives in hospital facilities. Int J Med Inform 2017 Jan;97:171-194. [doi: 10.1016/j.ijmedinf.2016.10.012] [Medline: 27919377]
95. Baxter R, Taylor N, Kellar I, Lawton R. What methods are used to apply positive deviance within healthcare organisations? A systematic review. BMJ Qual Saf 2016 Mar 20;25(3):190-201 [FREE Full text] [doi: 10.1136/bmjqs-2015-004386] [Medline: 26590198]

Abbreviations

eHealth: electronic health
GEP-HI: Good Evaluation Practice in Health Informatics
MRC: Medical Research Council
NASS: Nonadoption, Abandonment, and challenges to Scale-up, Spread, and Sustainability
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 12.01.20; peer-reviewed by E Ammenwerth, P Nykänen, F Mayoral; comments to author 17.03.20; revised version received 09.05.20; accepted 03.06.20; published 12.08.20.

Please cite as:
Bonten TN, Rauwerdink A, Wyatt JC, Kasteleyn MJ, Witkamp L, Riper H, van Gemert-Pijnen LJEWC, Cresswell K, Sheikh A, Schijven MP, Chavannes NH, EHealth Evaluation Research Group
Online Guide for Electronic Health Evaluation Approaches: Systematic Scoping Review and Concept Mapping Study
J Med Internet Res 2020;22(8):e17774
URL: https://www.jmir.org/2020/8/e17774
doi:10.2196/17774
PMID:32784173

Review

Understanding Self-Guided Web-Based Educational Interventions for Patients With Chronic Health Conditions: Systematic Review of Intervention Features and Adherence

Li Feng Xie[1], MSc; Alexandra Itzkovitz[1], BSc; Amelie Roy-Fleming[1], MSc; Deborah Da Costa[2], PhD; Anne-Sophie Brazeau[1,3], PhD

[1]School of Human Nutrition, McGill University, Sainte-Anne-de-Bellevue, QC, Canada

[2]Department of Medicine, McGill University, Montreal, QC, Canada

[3]Montreal Diabetes Research Center, Montreal, QC, Canada

Corresponding Author:
Anne-Sophie Brazeau, PhD
School of Human Nutrition
McGill University
21111 Lakeshore, Ste-Anne-de-Bellevue
Sainte-Anne-de-Bellevue, QC, H9X 3V9
Canada
Phone: 1 514 398 7848
Email: anne-sophie.brazeau@mcgill.ca

Abstract

Background: Chronic diseases contribute to 71% of deaths worldwide every year, and an estimated 15 million people between the ages of 30 and 69 years die mainly because of cardiovascular disease, cancer, chronic respiratory diseases, or diabetes. Web-based educational interventions may facilitate disease management. These are also considered to be a flexible and low-cost method to deliver tailored information to patients. Previous studies concluded that the implementation of different features and the degree of adherence to the intervention are key factors in determining the success of the intervention. However, limited research has been conducted to understand the acceptability of specific features and user adherence to self-guided web interventions.

Objective: This systematic review aims to understand how web-based intervention features are evaluated, to investigate their acceptability, and to describe how adherence to web-based self-guided interventions is defined and measured.

Methods: Studies published on self-guided web-based educational interventions for people (≥14 years old) with chronic health conditions published between January 2005 and June 2020 were reviewed following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Statement protocol. The search was performed using the PubMed, Cochrane Library, and EMBASE (Excerpta Medica dataBASE) databases; the reference lists of the selected articles were also reviewed. The comparison of the interventions and analysis of the features were based on the published content from the selected articles.

Results: A total of 20 studies were included. Seven principal features were identified, with goal setting, self-monitoring, and feedback being the most frequently used. The acceptability of the features was measured based on the comments collected from users, their association with clinical outcomes, or device adherence. The use of quizzes was positively reported by participants. Self-monitoring, goal setting, feedback, and discussion forums yielded mixed results. The negative acceptability was related to the choice of the discussion topic, lack of face-to-face contact, and technical issues. This review shows that the evaluation of adherence to educational interventions was inconsistent among the studies, limiting comparisons. A clear definition of adherence to an intervention is lacking.

Conclusions: Although limited information was available, it appears that features related to interaction and personalization are important for improving clinical outcomes and users' experience. When designing web-based interventions, the selection of features should be based on the targeted population's needs, the balance between positive and negative impacts of having human involvement in the intervention, and the reduction of technical barriers. There is a lack of consensus on the method of evaluating adherence to an intervention. Both investigations of the acceptability features and adherence should be considered when designing and evaluating web-based interventions. A proof-of-concept or pilot study would be useful for establishing the required level of engagement needed to define adherence.

(J Med Internet Res 2020;22(8):e18355) doi:10.2196/18355

KEYWORDS

chronic disease; online learning; self-management; mobile phone

Introduction

Background

Chronic diseases contribute to 71% of deaths worldwide every year, which corresponds to 41 million deaths per year. It has been estimated that among these deaths, 15 million people between the ages of 30 and 69 years die mainly because of cardiovascular disease, cancer, chronic respiratory diseases, or diabetes [1]. Apart from mortality, the consequences of these chronic diseases include a decrease in the quality of life [2,3] and an economic burden for both households and countries [4-6]. The use of information and communication technology for health-related purposes has the potential to mitigate these consequences by offering numerous benefits for disease management, such as facilitating access to health information and helping to increase the understanding of the disease [7]. It is also considered a flexible, low-cost method for patients to obtain information in comparison with face-to-face education sessions [8]. Web-based interventions are an example of information and communication technology that has the potential to educate people living with a specific chronic disease condition and can help to improve their self-care over the long term through education and peer support [8,9]. These web-based interventions can be in a guided format by including features such as electronic counseling (e-counseling) and long-distance monitoring by health care professionals (HCPs) [10] or can be self-guided, defined in this paper as an absence of individual or face-to-face contact between HCPs and the users. Previous studies have investigated the integration of various features (eg, reminders and opportunities for social support) and the design of these web-based interventions. They concluded that the implementation of specific features and degree of adherence to the intervention are key factors in determining their success [11,12]. However, these studies do not distinguish between interventions with one-on-one or in-person contact among users with (guided) and without (self-guided) an HCP. As contact with HCPs or e-consultations can lead to a higher cost per usage and decrease the accessibility of the intervention [13], it is important to understand the inclusion of specific features and evaluation of adherence to these self-guided interventions.

The definition and measurement of adherence to self-guided interventions are still subject to debate [14,15]. Adherence is defined by the World Health Organization as “the extent to which a person’s behaviour – taking medication, following a diet, and/or executing lifestyle changes, corresponds with agreed recommendations from a health care provider” [16]. However, this definition is not adapted in the context of information and communication technology; there is no prescribed dosage that users of specific web-based interventions should be using to have the expected behavioral change [12]. The difficulty in defining adherence to web-based self-guided interventions is further accentuated by the differences in which they have been measured across studies with the use of parameters, such as the number of log-ins, the content viewed, and/or the time spent on the intervention [14].

Objective

A deeper understanding of previously published evaluations of self-guided educational interventions is required. The goals of this systematic review are to investigate how web-based intervention features are evaluated to determine their acceptability and to explore how adherence to web-based self-guided interventions are defined and measured. An understanding of the specific features and standardization of the definition of adherence to web-based self-guided interventions can help increase their efficacy and help to develop future web-based interventions for disease management.

Methods

Design and Search Strategies

To achieve these objectives, a systematic review of studies investigating the acceptability of the included features in web-based educational interventions on chronic health conditions was conducted based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses framework [17]. For this review, chronic health conditions also include chronic diseases. Nine chronic health conditions were selected from a list of common chronic diseases in Canada [18]. The selection of these studies was related to the implication of a web-based educational intervention on patients’ self-management and their commonality across different age groups. Cancer and mental illness were excluded from this systematic review because of the broad variety of disease and treatment methods [19-21]. The selected categories were defined as follows: (1) arthritis, (2) celiac disease, (3) epilepsy, (4) inflammatory bowel disease (including Crohn disease and ulcerative colitis), (5) metabolic disorders (including hypertension, dyslipidemia, atherosclerosis, heart failure, gestational diabetes mellitus [GDM], and type 1 and type 2 diabetes mellitus), (6) multiple sclerosis, (7) overweight and obesity, (8) respiratory disease (including chronic respiratory disease, asthma, and chronic obstructive pulmonary disease [COPD]), and (9) kidney diseases (including end-stage renal disease and nephritis).

The search method was elaborated with the help of a librarian. The PubMed, Cochrane Library, and EMBASE (Excerpta Medica dataBASE) databases were used to ensure that all articles related to the topic were covered. Keywords (Textbox 1), derived from Medical Subject Headings (MeSH), were searched in the titles or abstracts. The search combined each medical condition with the web-based, education, and intervention terms. A full list of the search methods is included in Multimedia Appendix 1. If the clinical trial protocol was available, the corresponding author’s name and the study title were further searched on these databases to find the relevant publications. The reference lists of the selected articles were

also screened to capture potential articles. The screening and selection of the articles were performed independently by 2 reviewers (LFX and AI), and consensus was reached through a discussion to ensure agreeability. A third researcher (ASB) was consulted for a nonunanimous discussion for the selection of the articles.

Textbox 1. Keywords used for the article searches for different categories.

Web-based

- "social media" OR Internet OR "web based" OR web OR online

Education

- "distance education" OR education OR "patient education" OR teaching

Intervention

- learning OR intervention OR treatment OR program OR "Program development" OR platform

Arthritis

- arthritis

Celiac disease

- celiac

Epilepsy

- epilepsy

Inflammatory bowel disease

- IBD OR "inflammatory bowel disease" or "crohn disease" or "ulcerative colitis"

Metabolic disorders

- CVD OR hypertension OR diabetes OR "diabetes mellitus" OR "diabetes insipidus" OR "gestational diabetes" OR "type 2 diabetes mellitus" OR "type 1 diabetes mellitus" OR "Juvenile diabetes" OR "heart failure" OR atherosclerosis OR dyslipidemia OR "Cardiovascular disease"

Multiple sclerosis

- "multiple sclerosis"

Obesity

- "pediatric obesity" OR "abdominal obesity" OR "morbid obesity" OR "obesity management" OR "Abdominal obesity" OR "metabolic syndrome" OR "overweight" OR "metabolic syndrome" OR "weight reduction program"

Respiratory disease

- "respiratory disease" or "respiratory tract disease" or "respiratory disorder" or "asthma" or "chronic respiratory disease" or "copd" or "chronic obstructive pulmonary disease"

Kidney disease

- "chronic kidney disease" or "chronic renal insufficiency" or "kidney disease" or "chronic kidney failure" or "diabetic nephropathies" or "esrd" or "end stage renal disease" or "nephritis"

Study Selection

Inclusion criteria were as follows: (1) the study included a web-based educational intervention designed for people living with this health condition (eg, transfer of knowledge to this population), (2) the intervention aimed to improve clinical outcomes defined as the result of a health care intervention, which includes a change in clinical laboratory values (eg, level of blood glucose, blood lipid profile), lifestyle behavior (eg, improvement in eating habits and level of physical activity), use of health care system (eg, use of emergency department and length of hospitalization), and quality of life [22] related to an existing chronic health condition, (3) no in-person or one-to-one contact with an HCP within the intervention, (4) only contacted the research team for technical support or an introductory meeting during the intervention (to ensure the pragmatism of the study results [23] and limit the impact of these contacts on the adherence to the intervention), (5) the included population is ≥14 years old (age cutoff where people can make their own health care decisions in Quebec, Canada [24]), (6) the articles (published or in-press, to have a full portrait of the intervention and have peer-reviewed evidence) were published between

January 1, 2005, and June 15, 2020, in English or in French, (7) the articles are fully available to the researchers, and (8) no restriction on the design of the study but only original research was included.

Studies corresponding to any of the following criteria were excluded from this systematic review: (1) the intervention is for family members or HCPs only, (2) the intervention has only a purpose of prevention/assessment/screening aftercare, (3) the web-based intervention included a live session or personalized e-counseling, (4) the intervention consisted of only emails, discussion forums, and/or recording functions, (5) the study explicitly stated an inclusion of participants with severe depression, and (6) the primary target outcome was related to mental health.

Data Extraction and Analysis

For each study, the following information was collected and compared: the year of publication, country where the study took place, study design, targeted chronic health conditions, primary clinical outcomes, age group of the population, sample size, intervention given to the experimental and control groups, and length of the intervention.

In this study, a feature is defined as any functionality within a web-based educational intervention other than text-based educational modules, supporting users to have a better learning or navigation experience or to improve clinical outcomes. The term feature and functionality are used interchangeably for this review. Both analyses of acceptability of the features and adherence to the intervention were based on reported information contained in the articles or the complete protocol cited from the selected articles. The method for evaluating features and their acceptability on the outcomes of the intervention are discussed. The measurement and criteria used to evaluate adherence to the intervention were collected and compared between studies.

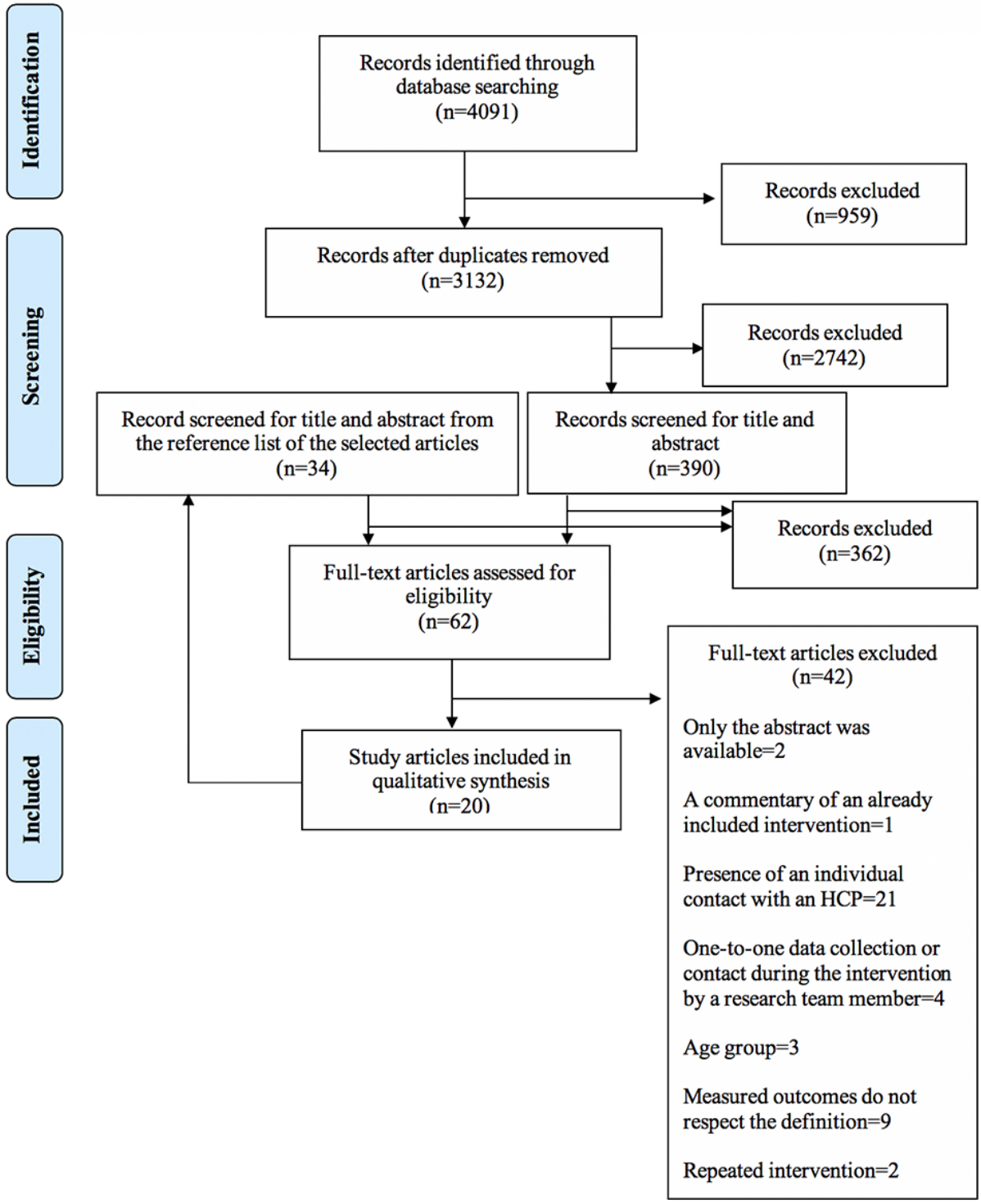
All the data were collected from information within the articles, the related published supplementary documents, or the cited references. If >1 article reported the same intervention and outcomes but had different sample sizes, then articles stating results of the acceptability of the features or adherence to the intervention were reported. If none or all the articles reported these details, the latest publication was analyzed. However, information related to the acceptability of the features and adherence was collected from all related articles. If 2 interventions within the same study correspond to the inclusion criteria of this review, the intervention with the highest number of features was analyzed. The data from each study were then grouped into themes. EndNote X9.2 for Macintosh was used to regroup the articles.

Results

Study Selection

The searches on the PubMed, Cochrane Library, and EMBASE databases resulted in 4091 potentially eligible articles (Figure 1). The titles and abstracts were reviewed, resulting in 390 articles. The titles and abstracts of potential articles from the reference list of the selected articles were also reviewed (n=34). After reading the full articles, 20 studies were selected.

Figure 1. Study selection.



Study Characteristics

The study characteristics are presented in Table 1. The identified articles included 6 areas of chronic health conditions: arthritis (n=1) [25], celiac disease (n=1) [26], metabolic disorders other than weight issues (n=8; metabolic syndrome [27], cardiac condition [28], hypertension [29], type 2 diabetes [30-33], and GDM [34]), multiple sclerosis (n=1) [35], overweight and obesity (n=7 studies) [13,36-41], and respiratory diseases (n=2; asthma [42] and COPD [43]). No study related to epilepsy, inflammatory bowel disease, or chronic kidney disease was found. The primary clinical outcomes were mainly related to changes in weight [13,36-41]. The studies were predominantly conducted in the United States [13,27,28,31,33,39-41,43] and Australia [25,26,34,36-38,42]. All the selected studies were randomized controlled trials, except for the study by Hutchesson et al [36], which was a pre-post design, and the study by Umapathy et al [25], which had a quasi-experimental design. All selected articles were published in English.

Table 1. Study characteristics and description of the interventions.

Study; primary clinical outcomes	Health conditions	Study design; length of the intervention	Population, sample size	Descriptions of the interventions given to the experimental group	Descriptions of the interventions given to the control group
Bosak, 2010, United States [27]; minutes of PA[a], energy expenditure per week	Metabolic syndrome	RCT[b]; 6 weeks	Adults ≥19 years; n=22	In-person introductory session, weekly new content, goal setting, self-monitoring, quiz, feedback (by email and after a quiz), use of persona, discussion board monitored by the PI[c]; general discussion question posted by the PI on the forum. Requested at least weekly participation in the discussion[d] (n=12, with 57% men)	Usual care (assessment by physicians and a consultation with a dietitian); (n=10, with 80% men)
Burns, 2013, Australia [42]; asthma control, self-efficacy, QoL[e]	Asthma	RCT; 3 months	Adults with asthma ≥55 years; n=51	Six 15-min modules, reminder email to the nonresponders[d] (with 33% men)	None
Carolan-Olah, 2019, Australia [34]; BMI, blood pressure, glycemic level	GDM[f]	RCT; ND[g]	Singleton pregnant women aged 18-45 years with recently diagnosed GDM; n=110	Standard GDM program and an additional 41-module web-based program including a one-on-one 30-min introductory session and quizzes[d] (n=52)	Standard GDM program (1.5 hours of in-person class education given by HCPs[h]; n=58)
Hansel, 2017, France [30]; change of the dietary score	T2DM[i]	RCT; 16 weeks	Adults with T2DM and abdominal obesity, 18-75 years; n=120	4 modules, videos, hotline technical support, and feedback on the self-monitoring data and pedometer outcomes; requested at least 11 weekly log-in[d] (n=60, with 33% men)	Usual follow-up with HCPs (n=60, with 33% men)
Hutchesson, 2016, Australia [36]; weight, BMI, WC[j]	Overweight	Pre-post design; 3 months	Women aged 18-30 years; n=26	Web-based quizzes to assess current health status (diet, exercise, weight) with email feedback report, goal setting, discussion forum monitored by a dietitian, smartphone app, email newsletters, text messages, graphic design reflecting the image of the population[d]	None
Jane, 2017, Australia [37]; weight	Obesity	RCT; 24 weeks	Adults aged 21-65 years; n=67	1. Leaflet group with pedometer: weight loss information contained in a booklet (n=23, with 9% men); 2. Facebook group with pedometer: same weight loss information within a booklet but with pages only accessible via the Facebook group. The group was monitored by the study coordinator and this person made a weekly post[d] (n=23, with 17% men); all the groups: 30-min introductory session	Standard care following Australian dietary and physical activity guidelines (n=21, with 19% men)
Kessel, 2016, New Zealand [35]; fatigue severity and impact	MS[k]	RCT; 8-10 weeks	Adults[l] experiencing MS fatigue; n=39	MSInvigor 8 plus: MSInvigor8 intervention with email-based support provided by a clinical psychologist for guidance and personal feedback (n=19, with 42% men)	MSInvigor8: cognitive behavior therapy–based 8 sessions with printable document, audio, and video; 25-50 min to complete; automated email reminders[d] (n=20, with 10% men)

Study; primary clinical outcomes	Health conditions	Study design; length of the intervention	Population, sample size	Descriptions of the interventions given to the experimental group	Descriptions of the interventions given to the control group
Kerfoot, 2017, United States [31]; HbA_{1c}[m]	T2DM	RCT; 6 months	Veterans with T2DM; n=456	Team-based web game with questions related to DSME[n] and a civic booklet about American history; other features: multiple-choice questions via email or smartphone app, same questions resent in a cycled pattern, points given for the quiz answer, feedback after the quiz, team and individual financial reward (US $100 gift certificate[d]; n=227, with 95% men)	Same game format as the intervention group but with game questions on civics and a DSME booklet (n=229, with 93% men)
Leahey, 2014, United States [13]; weight	Obesity	RCT; 3 months	Adults aged 18-70 years; n=230	Group 1: the ShapeUp Rhode Island 2011 (SURI) program plus an internet behavioral weight loss program. Included a 60-min introductory session, self-monitoring, and feedback on the progress[d] (n=90, with 18% men); group 2: the previous program plus optional weekly face-to-face group sessions (n=94, with 14% men)	SURI program alone: team participation, self-monitoring, pedometer, newsletters, community workshops, and recognition for meeting goals (n=46, with 18% men)
Liu, 2018, Canada [29]; systolic blood pressure	HTN[o]	RCT; 4 months	Adults aged 35-74 years with HTN; n=128	1. Web expert-driven group with a prescribed weekly exercise and diet plan (n=43, with 51% men); 2. web user-driven group with weekly email where participants can choose their exercise and diet goals[d] (n=42, with 48% men); in both groups, the same contents are under video and text format	Weekly email newsletter on HTN management only (n=43, with 57% men)
Morgan, 2011, Australia [38]; weight	Obesity	RCT; 3 months	Men aged 18-60 years; n=65	75-min face-to-face introductory session, self-monitoring, goal setting, feedback, and online forum weekly monitored by the research team[d] (n=31)	60-min face-to-face introductory session and a weight loss program booklet (n=34)
Moy, 2016, United States [43]; HRQoL[p]	COPD[q]	RCT; 4 months	Veterans with COPD, n=239	Goal setting, self-monitoring, feedback for the self-monitoring data, reminder, discussion forum, technical support, and pedometer[d] (n=155, with 95% men)	Pedometer with 12-month delayed access to the web intervention (n=84, with 92% men)
Noh, 2010, Korea [32]; postprandial glucose, HbA_{1c}	T2DM	RCT; 6 months	Adults with T2DM aged 18-80 years; n=40	6-module program, adaptation to smartphones[d] (n=20, with 80% men)	Same educational content in a printed booklet (n=20, with 75% men)
Richardson, 2007, United States [33]; steps	T2DM	RCT; 6 weeks	Nonpregnant adults with T2DM; n=35	Basic intervention with automated step goals based on the previous weekly total accumulated steps[d] (n=17, with 29% men)	Basic intervention (60-min introductory session, pedometer, access to web-based educational information, tailored motivational messages, feedback for the performance) with step goals based on walking bouts >10 min with at least 60 steps per minute (n=13, with 62% men)

Study; primary clinical outcomes	Health conditions	Study design; length of the intervention	Population, sample size	Descriptions of the interventions given to the experimental group	Descriptions of the interventions given to the control group
Rothert, 2006, United States, [39]; weight	Overweight and obesity	RCT; 6 weeks	Adult with BMI 27-40 kg/m^2; n=286	Tailored expert system: automated personal weight management plan delivered at 1, 3, and 6 weeks of the study; reminders and choice of encouragement message via email[d] (n=1475, with 17% men)	Information-only: standard Kaiser Permanente weight loss website (n=1378, with 13% men)
Sainsbury, 2013, Australia [26]; gluten-free diet adherence	Celiac disease	RCT; 8 weeks	Patients ≥16 years with biopsy-confirmed celiac disease (n=189, with 13% men)	Six 30-min modules[d] (n=101)	Access to the intervention after 8 weeks of randomization (n=88)
Tate, 2006, United States, [40]; weight	Overweight and obesity	RCT; 6 months	Adults (20-55 years) with a BMI 27-40 kg/m^2; n=122	1. Basic intervention with an additional website that includes electronic diary, message board, additional weekly reminder emails, weekly automated email feedback[d] (n=61, with 13% men); 2. same intervention as in 1 but email feedback was given by a human counselor (n=64, with 16% men)	Basic intervention: introductory face-to-face group session, diet and energy expenditure goal, access to SlimFast website, meal-replacement coupon, optional web matching with another participant, weekly report, email communications (n=67, with 18% men)
Thomas, 2015, United States [41]; weight	Obesity	RCT; 3 months	Adults aged 18-70 years; n=154	60-min introductory session, video, animation, quiz, self-monitoring, weekly feedback about participant's progress, reminders, and recognition for meeting the goals[d] (n=15, with 20% men)	Introductory session, printable newsletters with educational information on diet and physical activity; requested at least weekly login (n=16, with 21% men)
Umpathy, 2015, Australia [25]; heiQ[r]	OA[s]	Quasi-experimental study; 12 months	Adults with self-assessed hip and/or knee OA; n=195	My Joint Pain: educational modules (text or video) with self-assessment tools[d] (n=104, with 24% men)	No intervention was provided from the study (n=91, with 20% men)

Study; primary clinical outcomes	Health conditions	Study design; length of the intervention	Population, sample size	Descriptions of the interventions given to the experimental group	Descriptions of the interventions given to the control group
Widmer, 2017 2015, United States [28,44][t]; CV-related ED visits[u] and rehospitalizations	Cardiac condition	RCT; 3 months	Eligible patients to a regular cardiac rehabilitation; n=80	Regular cardiac rehabilitation with digital health: 30-min introductory session, accessibility via a smartphone app, technical support, and reminders[d] (n=40, with 78% men)	Regular cardiac rehabilitation for 36 weeks (weekly in-person meeting) (n=40, with 85% men)

[a]PA: physical activity.

[b]RCT: randomized controlled trial.

[c]PI: Principal Investigator.

[d]Interventions with a d superscript are the ones analyzed in this review.

[e]QoL: quality of life.

[f]GDM: gestational diabetes mellitus.

[g]ND: nondisposible.

[h]HCPs: health care professionals.

[i]T2DM: type 2 diabetes mellitus.

[j]WC: waist circumference.

[k]MS: multiple sclerosis.

[l]Adults refer to 18 years and older unless specified.

[m]HbA_{1c}: hemoblogin A_{1c}.

[n]DSME: diabetes self-management education.

[o]HTN: hypertension.

[p]HRQoL: health-related quality of life.

[q]COPD: chronic obstructive pulmonary disease.

[r]heiQ: health education impact questionnaire.

[s]OA: osteoarthritis.

[t]The selected article was Widmer et al, 2017 [28] and additional information about the interventions were collected from Widmer et al, 2015 [44].

[u]CV-related ED visit: cardiovascular-related emergency department visit.

Study Population

In the selected studies, 19 included an adult population (age 18 years) [13,25,27-34,37-43] and 1 included an adolescent/adult population aged 16 years [26]. The sample size varied from 22 to 456 participants. Seventeen studies included both genders [13,25-33,35,37,39-43]. The intervention length ranged from 8 weeks to 12 months, and in 1 article, the length was not specified [34].

Web Educational Components

The web-based interventions are summarized in Table 1.

Features and Acceptability

The main features included in the web-based educational intervention and their acceptability are summarized in Table 2. None of these main features were identified in the studies by Noh et al [32] and Sainsury et al [26].

Only 8 studies (8/20, 40%) discussed the acceptability of the features. Acceptability was evaluated based on feedback from the users [33,36,38], their association with clinical outcomes [13,31,33,38,40,41,43], or device (eg, pedometer) adherence [43]. The features that reported positive, negative, or mixed acceptability in the studies are presented with a "+," "−," or "±" symbol in Table 2.

Table 2. Main features included in the web-based educational intervention and their acceptability.

Articles and features	Introductory session	Goal settings	Self-monitoring	Quiz	Feedback	Reminder	Online community
Bosak, 2010, United States [27]	✓[a]	✓	✓	✓	✓	x[b]	✓
Burns, 2013, Australia [42]	x	x	x	x	x	✓	x
Carolan-Olah, 2019, Australia [34]	✓	x	x	✓	x	x	x
Hansel, 2017, France [30]	x	✓	✓	x	✓	x	x
Hutchesson, 2016, Australia [36]	x	−[c]	✓	+[d]	+	x	−
Jane, 2017, Australia [37]	✓	✓	✓	x	x	x	✓
Leahey, 2014, United States [13]	✓	✓	+	x	✓	x	x
Liu, 2018, Canada [29]	x	✓	x	x	x	x	x
Morgan, 2011, Australia [38]	✓	✓	±[e]	x	±	x	−
Moy, 2016, United States [43]	x	+	✓	x	+	✓	+
Richardson, 2007, United States [33]	✓	±	+	x	+	x	x
Rothert, 2006, United States [39]	x	x	x	x	x	✓	x
Kessel, 2016 and 2012, New Zealand [35,45]	x	x	✓	✓	x	✓	x
Kerfoot, 2017, United States [31]	x	x	x	✓	✓	x	+
Tate, 2006, United States [40]	✓	✓	✓	x	+	✓	✓
Thomas, 2015, United States [41]	✓	✓	+	✓	✓	✓	x
Umpathy, 2015, Australia [25]	x	x	✓	x	✓	x	x
Widmer, 2015 and 2017, United States [28,44]	✓	x	✓	x	x	✓	x

[a]✓: Features presented in the study but without evaluation of its acceptability.

[b]x: data not available.

[c]−: features reported having negative acceptability.

[d]+: features reported having positive acceptability.

[e]±: features with mixed acceptability.

Introductory Session

Face-to-face introductory sessions varying from 15 to 75 min were offered in 9 of the studies [13,27,28,33,34,37,38,40,41]. Among these, the study conducted by Carolan-Olah et al [34] specified that it was offered individually, and the study conducted by Tate et al [40] mentioned that it was offered in groups of 25 participants. The format was not specified in the other studies. The purposes of these sessions were mainly to introduce the study and provide instructions about navigating the website [28,33,34,37,38,40,41]. This session also allowed the development of personal goals, teach skills (eg, food intake self-monitoring), and provide the required material (eg, printed documents or meal supplement coupons) for the intervention [13,38,40,41]. In the selected articles, no information was provided on the usefulness or acceptability of this feature.

Goal Setting and Self-Monitoring

Among the selected studies, goal setting (n=11) and self-monitoring (n=13) were frequently reported. The participants were able to select their goal from a predetermined area (eg, physical activity or dietary habits) [27,29,36-38] or the goal was provided by the research team at the beginning of the intervention [13,29,30,40,41,43]. The predetermined topics were chosen according to clinical guidelines [13,29,37,41] based on participants' self-reported physical activity baseline information (eg, number of steps) [30,43] or self-reported performance from the previous week [33,46].

Three studies reported inconsistent acceptability of goal setting [33,36,43]. Participants in the study by Hutchesson et al [36] considered this feature as one of the least used. This could be related to the technical difficulty of not knowing where to find

this feature. Richardson et al [33] highlighted that more structured goals were associated with a lower level of satisfaction and adherence to the intervention among participants. However, Moy et al [43] reported that the goal-setting feature might lead to higher device (eg, pedometer) use.

Self-Monitoring

The term self-monitoring and self-assessment are used interchangeably in 2 studies [25,35]. Studies led by Umpathy et al [25] and Kessel et al [35] mainly used the term self-assessment to describe health-related risk assessment and information tracking (eg, pain, weight, use of medication). Ten other studies [13,27,28,30,33,36-38,41,43] used the term self-monitoring and referred only to the tracking function. As most of the studies used the term self-monitoring, *self-monitoring* was employed for this review.

Among all the studies with the tracking function, 6 studies requested daily self-monitoring throughout the [13,27,28,38,41,43] intervention. Other studies requested self-monitoring for a specific period (eg, participants need to complete the self-monitoring module in 1 week before going to the other modules [30,37] or by completing the module [35]), weekly, or longer self-monitoring for specific parameters (eg, weight change) [25,33,36,38,46]. The majority of the self-monitored data were entered directly into the intervention website [13,25,27,28,30,35,38,41,43,46], and one study used a smartphone app that was not designed by the research team [36]. In the study by Hutchesson et al [36], self-monitoring was captured in a quiz format where questions allowed participants to track their weight, eating habits, and physical activity level.

The acceptability of self-monitoring was evaluated in 4 studies [13,33,38,41]. Studies found that a greater frequency of self-reporting correlated with better clinical outcomes [13,38,41], increased mindfulness in food choices [38], or higher satisfaction with the intervention [33]. However, the participants in the study conducted by Morgan et al [38] expressed that it was difficult to use this feature and to remember the food eaten. These barriers might also explain the low compliance (<50%) in this study. However, the embedded *save favorite meals* feature was reported to simplify the recording process.

Quiz and Feedback

Quizzes were used in 6 studies [27,31,34-36,41]. They were mainly embedded within the web-based intervention, except in the studies by Hutchesson et al [36] and Kerfoot et al [31], where the questions were sent to participants by email or via a smartphone app. In addition to being used as a tracking method [36], the quizzes had the objective of introducing the learning material [31], learning reinforcement [27,34,35], and increasing participants' engagement [27,41]. Quizzes were included within the educational module [34,45] or sent periodically to the participants [27,31,36].

Feedback was used to reflect the progress of self-monitoring [13,25,30,33,38,40,41,43,46], the responses of the quizzes [27,31,36], and/or used as email communication with physicians [41]. In 8 of the studies, a report format was used either weekly [13,27,30,33,40,41,43,46] or periodically [38] as a follow-up to the self-monitoring data. Tate et al [40] also provided an automated weekly feedback report on the general performance of the participants for those who submitted their self-monitoring entries. In addition to summarizing the progress toward the goal [13,27,36,38,41,43], the report could also include recommendations [25,36,38,40,41], praise for achieving the goal [33,40,41], anecdotes [38], or the amount of virtual points/diamonds accumulated [36] or provide a personalized menu [30]. Among these, the use of an algorithm for generic messages or a standardized email based on the performance of each participant was used to build this report [13,25,27,33,38,40,41] and was specified in 7 of the studies. Rothert et al [39] noted the optional *buddy* feature where participants can receive email encouragement. However, no information was given on its specificity or the email content.

For feedback related to the quizzes, the correct answer and an explanation were often given immediately following the participants' responses [27,31]. The intervention led by Kessel et al [35] used the term *interactive tasks* and *homework* for the quiz feature. In this study, the completed quizzes were discussed in the following module, but the presence or absence of feedback to the participants' answers was not specified. Communication letters to physicians were used in 1 study and sent to the referring physician at 3 time points during the intervention [41].

The quiz feature was considered by the participants in 1 study as useful for providing information and feedback [36]. A similar observation was found in the study led by Richardson et al [33], where participants expressed their support for feedback on their step performance using a graph format. Morgan et al [38] explored the effect of the feedback feature, and the opinion was shared among participants. Some users positively highlighted its usefulness in helping people to realize their possible dietary issues, but others found that the feedback lacked personalization. In the study by Tate et al [40], the authors discussed that the feedback provided by both the automated computer program and the human counselor can lead to greater weight loss. This potential positive impact of the feedback feature on clinical outcomes was also reported by Moy et al [43].

Reminder

Seven studies included a reminder (eg, by email) to increase the intervention usage [35,39-42] or to recall the upload of self-monitoring data [28,38]. The frequency of sending the reminder varied between studies: weekly reminder emails to participants not using the web intervention only [41], occasional reminders to participants who did not recently log-in [28], weekly automatic reminders to all participants to upload their self-monitoring data [40,43] or the use of the intervention [35], reminder emails sent before the release of each management plan [39], or 1 reminder email midway of the intervention [42]. In addition to the email reminders, Widmer et al [28] also included reminders within the intervention to recall the completion of daily tasks and educational material. Other than reminding people participating in the intervention, Sainsbury et al [26] noted that email and text messages were used to manage participants' progress toward the goal, but the study did not explicitly use the term reminder to qualify this function.

No information was provided on the usefulness or acceptability of reminders in the selected articles.

Online Community

An online community was used in 7 studies [27,31,36-38,40,43]. Online communities included discussion forums [27,36,38,42,43], social media groups [37], game competitions [31], and buddy matching (optional pairing with another participant) [40]. The objectives of an online community were to increase social support between the participants [36,37,40,43], overcome barriers in behavioral change [27], answer questions [27,38,43], and/or increase a sense of competition [31]. The discussion forums were mainly operated by a research team member and divided into topics [27,36-38,43]. Jane et al [37] used a Facebook group to both deliver learning materials and encourage peer exchange. Tate et al [40] provided the option to the participants to be matched with another person and communicate through the web page. Kerfoot et al [28] used a game format to create an online community in which participants were grouped based on their geographic region and competed against each other by answering questions. A leaderboard displaying individual and team scores was used to increase the sense of competition.

Kerfoot et al [31] found that the positive change in mean hemoglobin A_{1c} among the participants was potentially related to participants' engagement in the online community and through competition with others. Its positive effect was further supported by a correlation between patient empowerment and game engagement, reflected by the number of earned points. The benefit of using an online community was also reported in the study by Moy et al [43]. The researchers compared the number of step counts in a population with COPD between the intervention group (access to the web intervention) with a control group having only the pedometer and a self-monitoring log. The results showed that the intervention group had significantly better device adherence, which suggested the potential benefits of the included features (discussion forum, educational content, goal setting, and feedback). In addition, more than half of the participants (67/121, 55%) expressed that the online community forum helped them learn information on their chronic condition. However, the use of the discussion forum was negatively rated in a study on weight loss among men [38]. In this study, the acceptability of the feature was based on qualitative feedback collected from the participants. Users of this discussion forum considered that weight loss was a personal issue and participants were unlikely to participate in the forum. Users also expressed a preference for having more face-to-face contact with the instructor. This negative comment was also reflected in an acceptability questionnaire in a study targeting weight loss in women [36].

Other Features

In addition to the previously mentioned features, others were presented in the studies, such as the use of a pedometer, reward, adaptation of the website intervention for smartphones, and technical support.

A pedometer was provided by 7 studies as a component of self-monitoring to increase step counts [13,27,29,30,33,37,43].

The use of rewards was mentioned in 3 studies. A social reward included praise in a weekly report to participants who reached their goal [41] and the use of online rewards (eg, virtual diamonds) [36] indicated participants' progress toward the goal. Only 1 study reported the use of material rewards [31], such as a US $100 certificate was given for the top 30% of participants based on their game points. It was also mentioned that the reward feature was included in the intervention led by Widmer et al [28], but no description was provided.

The adaptation of the website to mobile devices was specified in 3 studies [28,31,32].

The presence of technical support was mentioned in 3 studies. Participants could ask their questions by posting on a designated section of a discussion forum [43] via a link through the web-based program [28] or through hotline support [30]. In all instances, direct communication with a research team was restricted to technical support purposes.

Adherence to the Intervention

Adherence to the intervention was mentioned in 15 studies (75% of the eligible studies, 15/20) using different terms (eg, engagement, use of intervention, retention rate). The rate was reported in 4 studies. The parameters used to measure adherence to the intervention are summarized in Textbox 2.

A decrease in the use of the intervention throughout the study was observed in 6 studies [26,30,36,40,41,43]. For the length of a 16-week intervention, the percentage of log-ins in the study by Hansel et al [30] decreased by one-third in the final month. Moy et al [43] reported a similar decrease in the number of log-ins with time (from 6.8 per month in the first month to 3.0 per month at 12 months). A decrease in the use of the features was also observed, such as the number of opened newsletters [36], answered quizzes [36], and the use of the discussion forum [43]. A similar decrease in the frequency of monthly log-ins was observed in the study by Tate et al [40]. Although this decrease seemed to be progressive with time, Thomas et al [41] reported that it mainly occurred midintervention, 3 months from the beginning. Hutchesson et al [36] also observed that some features (eg, discussion forum and goal settings) had poor usage throughout the intervention and Morgan et al [38] reported that <50% of their participants complied with self-monitoring instructions. However, based on the general use of the intervention (eg, 7 weeks of submission of self-reporting data and weekly log-ins during the 3 months of the intervention), Morgan et al [38] qualified a retention rate of 41% as high. The term retention rate was also used by Sainsbury et al [26] and was measured with the use of the intervention. It was shown that 49.5% of the participants completed 4 of the 5 learning modules, but the authors considered this as a poor retention rate. Kessel et al [35] related the high dropout level (9/20, only 45% of the participants completed the intervention) to the absence of individual support, lack of feedback, and technical challenges. Bosak et al [27] explained that participants with better adherence had increased self-efficacy, but no additional information was provided.

Textbox 2. Parameters used to evaluate adherence to the intervention and the methods of measurement.

Log-in to the intervention

- Track of the total frequency of the log-in [30,32,36,40,42]
- Average log-in per participant [42]
- Average log-in per week per person [13]
- Average log-in per month per person [43]
- Number of weeks with at least one log-in [41]
- Total number of visits [42]

Exploration of the learning content

- Number of participants completed at least 4 out of the 5 modules [26]
- Number of lessons viewed [13]
- Number of participants who completed none, half, or all the 8 sessions [35]
- Mean number of sessions completed [35]

Upload of the self-monitoring data

- Total frequency of self-monitoring [13,33,38]
- Number of weeks having self-monitoring values at least 5 of the 7 days [41]
- Frequency of weekly web-based diary submission [40]

Use of other features

- Use of the discussion forum [36,43]
- Use of the discussion forum [36,43]
- Number of answered questions [31]
- Number of points earned during the game [31]
- Completion of quizzes, number of email newsletters opened, and smartphone app downloads [36]

Visit duration

- Total duration of viewing [42]
- Average viewing time by participant [42]

Discussion

Principal Findings

This systematic review highlights the use of specific features in the design of web-based self-guided interventions for people with chronic health conditions and reports on the evaluation of their acceptability. Previous researchers have investigated the importance of features included in guided web-based interventions for people with chronic diseases on their success rate (eg, adherence to the intervention and transfer of health-related information) [11,12]. However, limited data were found on the functionalities of self-guided web-based educational interventions. In-person and one-on-one interactions with an HCP might increase the adherence and use of a web-based intervention [47] but that can also increase the cost of the intervention [13]. Therefore, it is important to investigate the characteristics of web-based interventions. This review demonstrated that goal setting, self-monitoring, and feedback were the most common features. The acceptability of the different features was measured based on the comments collected from users, their influence on clinical outcomes, or device (eg, pedometer) adherence. The use of personalized features with feedback (eg, quizzes) was positively reported. The negative acceptability of the features was mainly related to technical issues and the choice of discussion topics for the intervention. This review also showed that the evaluation of adherence to the intervention was inconsistent among the studies, which limited comparison. A clear definition and measurement of adherence to web-based interventions is lacking.

Categorization of Features

Our review identified 7 features that were most commonly included in the selected studies (Table 2). Other features such as the use of a pedometer, rewards, adaptation of the website intervention for smartphones, and technical support were also observed but less frequently used. On the basis of the results of this paper, we categorized the included features under the following 3 categories: personalization, interaction, and support. Personalization refers to a function tailored to the individual needs of each participant and can be changed throughout the intervention based on the user's experience and progress [12].

Goal setting and self-monitoring have this characteristic by adjusting to the needs and progress of the user. The interactive features facilitated the engagement of the participants, increased learning retention [36], and provided a sense of community [31]. These characteristics were found in features such as quizzes, feedback, reminders, and online communities. They allowed an interaction between the intervention and participants and encouraged the users to return to the intervention [27,34,36]. Feedback and reward features correspond to both categories by personalizing the feedback report and varying the amount of rewards or type of written encouragement given to the participants based on the individual's progress [31,36]. Other features not included in these 2 categories were providing support and reducing the technical barriers of the intervention.

Importance of Evaluating the Features

Web-based educational interventions have been shown to be cost-effective compared with traditional face-to-face formats [48-51] and can reduce the production of physical materials (eg, printed documents) [52]. However, the cost related to the development of web-based educational interventions is still significant [52]. Creation of web-based educational modules can be classified into 3 levels: (1) basic content with text, graphics, simple audio, video, and test questions, (2) level 1 content with 25% interactive content (exercise, audio, video, and animations), and (3) level 2 content with highly interactive features (eg, adding game, avatars, custom interactions, and competitions) [53]. According to a study published in 2010, the average number of working hours to produce 1 hour of finished training associated with each of these levels is at least 79, 184, and 490 hours, respectively [53], and the average cost in US dollars is $10,054, $18,583, and $50,371, respectively [53]. Other factors such as the addition of new content and interactive features will further increase the cost [53]. Therefore, it is important to consider the choice of the features and their evaluation to minimize the cost and distribute the financial resources effectively. Our systematic review highlights that features are not frequently evaluated, with only 8 studies (8/20, 40%) reporting on the evaluation of some of the features used. In addition, the negative acceptability of a feature on the user's experience, clinical outcomes, or device adherence was shown to be related to a lack of responding to the population's needs, low human contact, and technical difficulties.

Factors Impacting the Acceptability of a Feature

Lack of Responding to the Population's Needs

A previous systematic review investigating features to be included in a commercial smartphone app for people with type 1 diabetes highlights the importance of integrating features related to personalization and patient empowerment for optimal disease self-management [54]. Similar to this study, our review showed the benefits of these groups of features [36,38]. For instance, the self-monitoring feature showed positive acceptability for the user's experience, clinical outcomes, or device adherence. Participants in a weight loss intervention conducted by Morgan et al [38] expressed that the self-monitoring features helped to increase mindfulness of their dietary choices. The participants also liked the *save favorite meals* option, which was associated with their eating habits and facilitated their diet entries [38]. Another feature that can increase patient empowerment is feedback, but it was found to lack personalization. Being able to effectively provide information [36] and improve behaviors [38] are some of the benefits of providing feedback through self-monitoring and quizzes. However, the use of a generic message was criticized by some participants and they expressed a preference for having more personalized communication [38]. This evidence shows the potential benefits of these features and highlights the necessity of adapting them to patients' needs.

Indeed, the effectiveness of a feature can only be maximized when there is a deep understanding of the targeted population's needs [15,38]. For example, peer support is often identified as an essential component in web-based interventions across different areas of health care [55-58], but its use should be based on the specific population's preferences. Kerfoot et al [31] and Moy et al [43] found a positive correlation between participants' engagement, learning, and use of an online community. However, men in a weight loss study also expressed their resistance in using the discussion forum mainly because of the personal nature of the topic and they preferred to have face-to-face contact with their instructor [38]. Similar feedback was also reported in a weight loss study in women [36]. As the interest and needs of patients vary with different types of chronic diseases, the topics involved in these discussion forums should also be based on the interests of the population group being targeted. For instance, Lanoye et al [59] found the importance of discussing the stigma and peer pressure related to obesity within a young adult population, whereas Cook et al [60] found that emotional support and use of medication are priorities in an older population with obesity. Therefore, the demographic background [11,61,62] and type of chronic diseases [7] are all factors potentially influencing the acceptability of a feature and should be considered when designing and evaluating web-based interventions.

Low Human Contact

In addition to the lack of responding to the population's needs, the frequency of human contact was another element mentioned in the selected studies that could interfere with the acceptability of a feature [36]. Hutchesson et al [36] suggested that the low level of human contact in their weight loss intervention could have been a reason for the low usage of the discussion forum. Leahey et al [13] verified this hypothesis in their study on weight loss by adding a face-to-face component to their web-based intervention; however, it was shown that improved clinical outcomes also resulted in a higher monetary cost. Kessel et al [35] also mentioned that having human contact (eg, telephone support) might lead to a higher engagement with the intervention. Therefore, a greater in-person or one-on-one consultation with an HCP in the intervention has the potential to increase its efficacy, but the cost should also be considered. As the goal of this systematic review is to investigate the features presented in self-guided web-based interventions, with the primary inclusion criteria of the studies being the absence of face-to-face contact, it would be contradictory to suggest the addition of a face-to-face component for an intervention. However, having patient moderators implicated in the intervention can be a potential solution for this barrier [63].

Moderators have the role of being the *housekeeper* of the discussion forum. They adopt an objective point of view by balancing the opinions of different sources in a respective environment. It also acts as a conversation stimulator, conflict resolver, feedback provider, and discussion supporter [63,64]. Previous studies highlighted the importance of their role by showing that participants can develop an attachment with community moderators and that their departure can lead to cessation in the use of the forum among some participants [65]. Having HCPs and peer moderators will combine the expertise for the delivery of web-based interventions [12]. As the use of the intervention is also associated with its impact (eg, on clinical outcomes or behavioral change) [12], it is important to be able to define and measure the level of adherence [12]. Adherence can be associated with factors such as chronic health conditions [26,42], study design, and inclusion of a variety of features [12,66]. In our review, the eligible studies reported different ways of measuring adherence to the interventions (eg, log-ins to the intervention [42], exploration of the learning content [13], and uploading of the self-monitoring data [41]) using different terms (eg, engagement [36], retention rate [38]), and none of them defined the effective engagement or intended usage of the intervention.

Technical Difficulties

Technical barriers were a third reason for the lower acceptability of a feature. Users in the weight loss trial conducted by Morgan et al [38] expressed that despite an improvement in behavioral changes related to the use of self-monitoring, the difficulty in tracking their food decreased their use of the intervention. Hutchesson et al [36] also suggested that the lack of usage of the goal-setting feature might be related to the difficulty in finding this feature in the intervention. This low usage was attributed to technical issues, and was previously reported in the literature [14]. The action planning feature usage was reported as relatively low in a study of people with type 2 diabetes conducted by Glasgow et al [14], and this could be related to navigational difficulties. These observations highlight the importance of simplifying the intervention navigation and including technical support features (eg, introductory session), providing contact information of the research team, and technology usage learning to help decrease these barriers [67].

Adherence and Future Direction

Intended usage is estimated by the developers and refers to the usage level needed to have the maximum benefit from the intervention (eg, clinical outcomes), and defining the intended usage would allow for standardization in the calculation of adherence [12]. Although Kelders et al [12] used the term intended usage, others adopted the term effective engagement [68,69], defined as "sufficient engagement with the intervention to achieve intended outcomes" [69]. As both terminologies focused on the identification of the parameters and the related minimum threshold that can have an impact on the intended behavior [12,68,69], these terms were used interchangeably.

Effective engagement should reflect the multidimension of the intervention in relation to the primary outcome, and both objective and subjective measurements should be evaluated [70]. The back-ended intervention usage data are considered an objective measurement [70] and can be assessed by using the Analyzing and Measuring Usage and Engagement Data framework [68]. This framework is designed for web-based interventions and can be used during the intervention development phase or after data collection. It contains 3 stages, and each stage is guided by a checklist of generic questions. In stage 1, the usage of data is classified into 3 categories: intervention characteristics (eg, architecture and content), accrued data (eg, data collected during the use of the intervention), and contextual data (eg, factors influencing the use of the intervention). Stage 2 consists of the selection of meaningful measures of usage and generation of research questions related to the primary outcome, usage data collected, and characteristics of the target population (eg, a web-based intervention focusing on the reduction of hospital visits can have “Will the number of content views be associated with hospital visits?” as a research question [68]). The final stage focuses on the selection of analytical tools and data preparation. A plan of analyses can then be conceived if the intervention is in the developmental phase or the analyses can be performed if data have already been collected [68]. In addition to the usage data, qualitative analysis (eg, with a semistructured interview or focus group) should be performed and combined with the quantitative methods [70] to reflect participants’ experiences. The threshold of effective engagement found with the combination of these 2 methods can then be compared with the actual intervention usage of each participant. Those who failed to reach this threshold will then be categorized as nonadherent to the intervention. Therefore, adherence to the intervention and its cutoff should only be defined after data collection is completed and a proof-of-concept or pilot study is recommended for testing [71].

Limitations

Our systematic review had some limitations. The search terms were selected based on MeSH terms; however, other important keywords could have been included. Exclusion of these important keywords might decrease the level of comprehensiveness of the search results. All the qualitative analyses were based on the content of the articles; the omission of information within the published articles might have led to a different interpretation of the results. For example, authors might only have listed the major features in their intervention instead of providing a complete list of all the available features. Only 8 studies (8/20, 40%) reported the acceptability of the features on the clinical outcome, users’ experience, or device adherence, which is a limitation for extrapolating the conclusions of the interventions. The articles included in this review were only selected from 3 databases, limited to published or in-press articles in English and French. In addition, to ensure a higher level of effectiveness in the results, this review also excluded self-guided interventions having individual contact between participants and research professionals during the study for reasons other than technical support or introductory sessions. Therefore, the results of this review might have limited external validity and cannot be applied to all web-based self-guided interventions or specific to any of the selected disease categories.

Conclusions

In conclusion, this systematic review investigated features included in 20 self-guided web-based educational interventions focusing on the self-management of chronic health conditions. It demonstrated the positive implication of specific features related to personalization and interactivity in the interventions on clinical outcomes, users' experience, or device adherence. However, only a few studies reported the acceptability of the included features; therefore, future research is needed to gain a greater understanding of the roles that each feature plays on the use of web-based interventions. The results of this systematic review provide evidence on the choice and implementation of specific features for future web-based health education interventions, highlighting the importance of understanding the needs of the target population and the need to incorporate more human contact and reducing technical barriers for the effectiveness of self-guided web-based interventions. Moreover, this study also found poor consensus related to the definitions and measurements of adherence in self-guided interventions used to target chronic health conditions. A method for evaluating the level of adherence is proposed in this review but requires future studies for its validation.

Acknowledgments

The authors would like to acknowledge Dr Rémi Rabasa-Lhoret for providing suggestions for protocol development and Meryem Talbo, Kayla Wong, and Melinda Prevost for their feedback to improve this manuscript.

LFX holds a PhD scholarship from the Fonds de Recherche Santé Québec.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Keywords used for the article searches.
[DOCX File , 16 KB - jmir_v22i8e18355_app1.docx]

References

1. Noncommunicable Diseases: Key Facts. World Health Organization. 2018. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2019-06-01]
2. Megari K. Quality of life in chronic disease patients. Health Psychol Res 2013 Sep 24;1(3):e27 [FREE Full text] [doi: 10.4081/hpr.2013.e27] [Medline: 26973912]
3. Preto O, Amaral O, Duarte J, Chaves C, Coutinho E, Nelas P. Quality of Life and Chronic Disease in Patients Receiving Primary Health Care. In: 2nd International Conference on Health and Health Psychology. 2016 Presented at: HHP'16; July 6-9, 2016; Porto, Portugal p. 217-226. [doi: 10.15405/epsbs.2016.07.02.21]
4. Engelgau MM, Karan A, Mahal A. The economic impact of non-communicable diseases on households in India. Global Health 2012 Apr 25;8:9 [FREE Full text] [doi: 10.1186/1744-8603-8-9] [Medline: 22533895]
5. Bloom D, Chen S, Kuhn M, McGovern M, Oxley L, Prettner K. The economic burden of chronic diseases: estimates and projections for China, Japan, and South Korea. J Econ Ageing 2018 Sep 26:100163 epub ahead of print. [doi: 10.1016/j.jeoa.2018.09.002]
6. Muka T, Imo D, Jaspers L, Colpani V, Chaker L, van der Lee SJ, et al. The global impact of non-communicable diseases on healthcare spending and national income: a systematic review. Eur J Epidemiol 2015 Apr;30(4):251-277. [doi: 10.1007/s10654-014-9984-2] [Medline: 25595318]
7. Win KT, Hassan NM, Bonney A, Iverson D. Benefits of online health education: perception from consumers and health professionals. J Med Syst 2015 Mar;39(3):27. [doi: 10.1007/s10916-015-0224-4] [Medline: 25666928]
8. Duke DC, Barry S, Wagner DV, Speight J, Choudhary P, Harris MA. Distal technologies and type 1 diabetes management. Lancet Diabetes Endocrinol 2018 Feb;6(2):143-156. [doi: 10.1016/S2213-8587(17)30260-7] [Medline: 28867311]
9. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013 Feb 14;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
10. Smailhodzic E, Hooijsma W, Boonstra A, Langley D. Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16:442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
11. van de Belt TH, Engelen LJ, Berben SA, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. J Med Internet Res 2013 Oct 2;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
12. Kelders S, Kok R, Ossebaard H, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]

13. Leahey TM, Thomas G, Fava JL, Subak LL, Schembri M, Krupel K, et al. Adding evidence-based behavioral weight loss strategies to a statewide wellness campaign: a randomized clinical trial. Am J Public Health 2014 Jul;104(7):1300-1306. [doi: 10.2105/AJPH.2014.301870] [Medline: 24832424]
14. Glasgow R, Christiansen S, Kurz D, King D, Woolley T, Faber A, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011 Jan 25;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]
15. Collins CE, Morgan PJ, Hutchesson MJ, Callister R. Efficacy of standard versus enhanced features in a web-based commercial weight-loss program for obese adults, part 2: randomized controlled trial. J Med Internet Res 2013 Jul 22;15(7):e140 [FREE Full text] [doi: 10.2196/jmir.2626] [Medline: 23876832]
16. Burkhart PV, Sabaté E. Adherence to long-term therapies: evidence for action. J Nurs Scholarsh 2003;35(3):207. [Medline: 14562485]
17. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
18. Chronic Diseases. Government of Canada. 2019. URL: https://www.canada.ca/en/public-health/services/chronic-diseases.html [accessed 2020-07-29]
19. Pucci C, Martinelli C, Ciofani G. Innovative approaches for cancer treatment: current perspectives and new challenges. Ecancermedicalscience 2019;13:961 [FREE Full text] [doi: 10.3332/ecancer.2019.961] [Medline: 31537986]
20. Malla A, Joober R, Garcia A. 'Mental illness is like any other medical illness': a critical examination of the statement and its impact on patient care and society. J Psychiatry Neurosci 2015 May;40(3):147-150 [FREE Full text] [doi: 10.1503/jpn.150099] [Medline: 25903034]
21. NIH Curriculum Supplement Series. New York, USA: National Institutes of Health; 2007.
22. Velentgas P, Dreyer N, Nourjah P, Smith S, Torchia M. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Washington, DC: Government Printing Office; 2013.
23. Lipman P, Loudon K, Dluzak L, Moloney R, Messner D, Stoney C. Framing the conversation: use of PRECIS-2 ratings to advance understanding of pragmatic trial design domains. Trials 2017 Nov 10;18(1):532 [FREE Full text] [doi: 10.1186/s13063-017-2267-y] [Medline: 29126437]
24. Coughlin K. Medical decision-making in paediatrics: infancy to adolescence. Paediatr Child Health 2018 May;23(2):138-146 [FREE Full text] [doi: 10.1093/pch/pxx127] [Medline: 30653623]
25. Umapathy H, Bennell K, Dickson C, Dobson F, Fransen M, Jones G, et al. The web-based osteoarthritis management resource my joint pain improves quality of care: a quasi-experimental study. J Med Internet Res 2015 Jul 7;17(7):e167 [FREE Full text] [doi: 10.2196/jmir.4376] [Medline: 26154022]
26. Sainsbury K, Mullan B, Sharpe L. A randomized controlled trial of an online intervention to improve gluten-free diet adherence in Celiac disease. Am J Gastroenterol 2013 May;108(5):811-817. [doi: 10.1038/ajg.2013.47] [Medline: 23458849]
27. Bosak KA, Yates B, Pozehl B. Effects of an internet physical activity intervention in adults with metabolic syndrome. West J Nurs Res 2010 Feb;32(1):5-22. [doi: 10.1177/0193945909333889] [Medline: 19357421]
28. Widmer RJ, Allison TG, Lennon R, Lopez-Jimenez F, Lerman LO, Lerman A. Digital health intervention during cardiac rehabilitation: a randomized controlled trial. Am Heart J 2017 Jun;188:65-72. [doi: 10.1016/j.ahj.2017.02.016] [Medline: 28577682]
29. Liu S, Brooks D, Thomas SG, Eysenbach G, Nolan RP. Effectiveness of user- and expert-driven web-based hypertension programs: an RCT. Am J Prev Med 2018 Apr;54(4):576-583. [doi: 10.1016/j.amepre.2018.01.009] [Medline: 29456025]
30. Hansel B, Giral P, Gambotti L, Lafourcade A, Peres G, Filipecki C, et al. A fully automated web-based program improves lifestyle habits and HbA1c in patients with type 2 diabetes and abdominal obesity: randomized trial of patient e-coaching nutritional support (the ANODE study). J Med Internet Res 2017 Nov 8;19(11):e360 [FREE Full text] [doi: 10.2196/jmir.7947] [Medline: 29117929]
31. Kerfoot B, Gagnon D, McMahon G, Orlander J, Kurgansky K, Conlin P. A team-based online game improves blood glucose control in veterans with type 2 diabetes: a randomized controlled trial. Diabetes Care 2017 Sep;40(9):1218-1225. [doi: 10.2337/dc17-0310] [Medline: 28790131]
32. Noh J, Cho Y, Nam H, Kim J, Kim D, Yoo H, et al. Web-based comprehensive information system for self-management of diabetes mellitus. Diabetes Technol Ther 2010 May;12(5):333-337. [doi: 10.1089/dia.2009.0122] [Medline: 20388042]
33. Richardson CR, Mehari KS, McIntyre LG, Janney AW, Fortlage LA, Sen A, et al. A randomized trial comparing structured and lifestyle goals in an internet-mediated walking program for people with type 2 diabetes. Int J Behav Nutr Phys Act 2007 Nov 16;4:59 [FREE Full text] [doi: 10.1186/1479-5868-4-59] [Medline: 18021411]
34. Carolan-Olah M, Sayakhot P. A randomized controlled trial of a web-based education intervention for women with gestational diabetes mellitus. Midwifery 2019 Jan;68:39-47. [doi: 10.1016/j.midw.2018.08.019] [Medline: 30343264]
35. van Kessel K, Wouldes T, Moss-Morris R. A New Zealand pilot randomized controlled trial of a web-based interactive self-management programme (MSInvigor8) with and without email support for the treatment of multiple sclerosis fatigue. Clin Rehabil 2016 May;30(5):454-462. [doi: 10.1177/0269215515584800] [Medline: 25952587]

36. Hutchesson M, Morgan P, Callister R, Pranata I, Skinner G, Collins C. Be positive be healthe: development and implementation of a targeted e-health weight loss program for young women. Telemed J E Health 2016 Jun;22(6):519-528. [doi: 10.1089/tmj.2015.0085] [Medline: 26701611]
37. Jane M, Hagger M, Foster J, Ho S, Kane R, Pal S. Effects of a weight management program delivered by social media on weight and metabolic syndrome risk factors in overweight and obese adults: a randomised controlled trial. PLoS One 2017;12(6):e0178326 [FREE Full text] [doi: 10.1371/journal.pone.0178326] [Medline: 28575048]
38. Morgan PJ, Lubans DR, Collins CE, Warren JM, Callister R. 12-month outcomes and process evaluation of the SHED-IT RCT: an internet-based weight loss program targeting men. Obesity (Silver Spring) 2011 Jan;19(1):142-151. [doi: 10.1038/oby.2010.119] [Medline: 20523304]
39. Rothert K, Strecher VJ, Doyle LA, Caplan WM, Joyce JS, Jimison HB, et al. Web-based weight management programs in an integrated health care setting: a randomized, controlled trial. Obesity (Silver Spring) 2006 Feb;14(2):266-272 [FREE Full text] [doi: 10.1038/oby.2006.34] [Medline: 16571852]
40. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
41. Thomas JG, Leahey TM, Wing RR. An automated internet behavioral weight-loss program by physician referral: a randomized controlled trial. Diabetes Care 2015 Jan;38(1):9-15 [FREE Full text] [doi: 10.2337/dc14-1474] [Medline: 25404659]
42. Burns P, Jones SC, Iverson D, Caputi P. AsthmaWise - a field of dreams? The results of an online education program targeting older adults with asthma. J Asthma 2013 Sep;50(7):737-744. [doi: 10.3109/02770903.2013.799688] [Medline: 23745621]
43. Moy ML, Martinez CH, Kadri R, Roman P, Holleman RG, Kim HM, et al. Long-term effects of an internet-mediated pedometer-based walking program for chronic obstructive pulmonary disease: randomized controlled trial. J Med Internet Res 2016 Aug 8;18(8):e215 [FREE Full text] [doi: 10.2196/jmir.5622] [Medline: 27502583]
44. Widmer RJ, Allison TG, Lerman LO, Lerman A. Digital health intervention as an adjunct to cardiac rehabilitation reduces cardiovascular risk factors and rehospitalizations. J Cardiovasc Transl Res 2015 Jul;8(5):283-292 [FREE Full text] [doi: 10.1007/s12265-015-9629-1] [Medline: 25946990]
45. Moss-Morris R, McCrone P, Yardley L, van Kessel K, Wills G, Dennison L. A pilot randomised controlled trial of an internet-based cognitive behavioural therapy self-management programme (MS Invigor8) for multiple sclerosis fatigue. Behav Res Ther 2012 Jun;50(6):415-421. [doi: 10.1016/j.brat.2012.03.001] [Medline: 22516321]
46. Moy ML, Janney AW, Nguyen HQ, Matthess KR, Cohen M, Garshick E, et al. Use of pedometer and Internet-mediated walking program in patients with chronic obstructive pulmonary disease. J Rehabil Res Dev 2010;47(5):485-496 [FREE Full text] [doi: 10.1682/jrrd.2009.07.0091] [Medline: 20803392]
47. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
48. Maloney S, Haas R, Keating JL, Molloy E, Jolly B, Sims J, et al. Breakeven, cost benefit, cost effectiveness, and willingness to pay for web-based versus face-to-face education delivery for health professionals. J Med Internet Res 2012 Apr 2;14(2):e47 [FREE Full text] [doi: 10.2196/jmir.2040] [Medline: 22469659]
49. Schulz D, Smit E, Stanczyk N, Kremers S, de Vries H, Evers S. Economic evaluation of a web-based tailored lifestyle intervention for adults: findings regarding cost-effectiveness and cost-utility from a randomized controlled trial. J Med Internet Res 2014 Mar 20;16(3):e91 [FREE Full text] [doi: 10.2196/jmir.3159] [Medline: 24650860]
50. Krishnan A, Finkelstein E, Levine E, Foley P, Askew S, Steinberg D, et al. A digital behavioral weight gain prevention intervention in primary care practice: cost and cost-effectiveness analysis. J Med Internet Res 2019 May 17;21(5):e12201 [FREE Full text] [doi: 10.2196/12201] [Medline: 31102373]
51. Smit ES, Evers SM, de Vries H, Hoving C. Cost-effectiveness and cost-utility of internet-based computer tailoring for smoking cessation. J Med Internet Res 2013 Mar 14;15(3):e57 [FREE Full text] [doi: 10.2196/jmir.2059] [Medline: 23491820]
52. Weller M. Learning objects and the e-learning cost dilemma. Open Learn 2010 Aug 19;19(3):293-302. [doi: 10.1080/0268051042000280147]
53. Chapman B. How Long Does It Take to Create Learning. Chapman Alliance. 2010. URL: http://www.chapmanalliance.com/howlong/ [accessed 2020-08-03]
54. Martinez-Millana A, Jarones E, Fernandez-Llatas C, Hartvigsen G, Traver V. App features for type 1 diabetes support and patient empowerment: systematic literature review and benchmark comparison. JMIR Mhealth Uhealth 2018 Nov 21;6(11):e12237 [FREE Full text] [doi: 10.2196/12237] [Medline: 30463839]
55. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]
56. Balfe M, Doyle F, Smith D, Sreenan S, Brugha R, Hevey D, et al. What's distressing about having type 1 diabetes? A qualitative study of young adults' perspectives. BMC Endocr Disord 2013 Jul 25;13:25 [FREE Full text] [doi: 10.1186/1472-6823-13-25] [Medline: 23885644]

57. Yu CH, Parsons JA, Hall S, Newton D, Jovicic A, Lottridge D, et al. User-centered design of a web-based self-management site for individuals with type 2 diabetes - providing a sense of control and community. BMC Med Inform Decis Mak 2014 Jul 23;14:60 [FREE Full text] [doi: 10.1186/1472-6947-14-60] [Medline: 25056379]
58. Joglekar S, Sastry N, Coulson N, Taylor S, Patel A, Duschinsky R, et al. How online communities of people with long-term conditions function and evolve: network analysis of the structure and dynamics of the asthma UK and British lung foundation online communities. J Med Internet Res 2018 Jul 11;20(7):e238 [FREE Full text] [doi: 10.2196/jmir.9952] [Medline: 29997105]
59. Lanoye A, Gorin A, LaRose J. Young adults' attitudes and perceptions of obesity and weight management: implications for treatment development. Curr Obes Rep 2016 Mar;5(1):14-22 [FREE Full text] [doi: 10.1007/s13679-016-0188-9] [Medline: 26923688]
60. Cook N, Tripathi P, Weiss O, Walda S, George A, Bushell A. Patient needs, perceptions, and attitudinal drivers associated with obesity: a qualitative online bulletin board study. Adv Ther 2019 Apr;36(4):842-857 [FREE Full text] [doi: 10.1007/s12325-019-00900-1] [Medline: 30859499]
61. Fergie G, Hunt K, Hilton S. What young people want from health-related online resources: a focus group study. J Youth Stud 2013 Aug;16(5):579-596 [FREE Full text] [doi: 10.1080/13676261.2012.744811] [Medline: 24748849]
62. Senft N, Everson J. Ehealth engagement as a response to negative healthcare experiences: cross-sectional survey analysis. J Med Internet Res 2018 Dec 5;20(12):e11034 [FREE Full text] [doi: 10.2196/11034] [Medline: 30518513]
63. Huh J, McDonald D, Hartzler A, Pratt W. Patient moderator interaction in online health communities. AMIA Annu Symp Proc 2013;2013:627-636 [FREE Full text] [Medline: 24551364]
64. Wright S. The Role of the Moderator: Problems and Possibilities for Government-Run Online Discussion Forums. Berkeley, USA: Online Deliberation; 2009.
65. Huh J, Marmor R, Jiang X. Lessons learned for online health community moderator roles: a mixed-methods study of moderators resigning from WebMD communities. J Med Internet Res 2016 Sep 8;18(9):e247 [FREE Full text] [doi: 10.2196/jmir.6331] [Medline: 27608721]
66. Schubart J, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92 [FREE Full text] [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21164337]
67. Kontos EZ, Bennett GG, Viswanath K. Barriers and facilitators to home computer and internet use among urban novice computer users of low socioeconomic position. J Med Internet Res 2007 Oct 22;9(4):e31 [FREE Full text] [doi: 10.2196/jmir.9.4.e31] [Medline: 17951215]
68. Miller S, Ainsworth B, Yardley L, Milton A, Weal M, Smith P, et al. A framework for analyzing and measuring usage and engagement data (AMUSED) in digital interventions: viewpoint. J Med Internet Res 2019 Feb 15;21(2):e10966 [FREE Full text] [doi: 10.2196/10966] [Medline: 30767905]
69. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
70. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. Mhealth 2018;4:37 [FREE Full text] [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]
71. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]

Abbreviations

COPD: chronic obstructive pulmonary disease
e-counseling: electronic counseling
EMBASE: Excerpta Medica dataBASE
GDM: gestational diabetes mellitus
HCP: health care professional
MeSH: Medical Subject Headings

Edited by G Eysenbach; submitted 21.02.20; peer-reviewed by C Loiselle, F Fatehi, K Fitzner; comments to author 12.06.20; revised version received 16.07.20; accepted 19.07.20; published 13.08.20.

Please cite as:
Xie LF, Itzkovitz A, Roy-Fleming A, Da Costa D, Brazeau AS
Understanding Self-Guided Web-Based Educational Interventions for Patients With Chronic Health Conditions: Systematic Review of Intervention Features and Adherence
J Med Internet Res 2020;22(8):e18355
URL: http://www.jmir.org/2020/8/e18355/
doi:10.2196/18355
PMID:32788152

Review

Hazards of the Cytokine Storm and Cytokine-Targeted Therapy in Patients With COVID-19: Review

Miguel Quirch[1], MD; Jeannie Lee[1], MD; Shabnam Rehman[1], MD

Texas Tech University Health Sciences Center, Lubbock, TX, United States

Corresponding Author:
Miguel Quirch, MD
Texas Tech University Health Sciences Center
3601 4th St
Lubbock, TX, 79430
United States
Phone: 1 806 743 3543
Email: miguel.quirch@ttuhsc.edu

Abstract

Background: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has challenged medicine and health care on a global scale. Its impact and frightening mortality rate are in large part attributable to the fact that there is a lack of available treatments. It has been shown that in patients who are severely ill, SARS-CoV-2 can lead to an inflammatory response known as cytokine storm, which involves activation and release of inflammatory cytokines in a positive feedback loop of pathogen-triggered inflammation. Currently, cytokine storm is one of the leading causes of morbidity and mortality in SARS-CoV-2, but there is no proven treatment to combat this systemic response.

Objective: The aim of this paper is to study the cytokine storm response in SARS-CoV-2 and to explore the early treatment options for patients who are critically ill with the coronavirus disease (COVID-19) in the early stages of the pandemic by reviewing the literature.

Methods: A literature review was performed from December 1, 2000, to April 4, 2020, to explore and compare therapies that target cytokine storm among SARS-CoV-2 and prior coronavirus cases.

Results: A total of 38 eligible studies including 24 systematic reviews, 5 meta-analyses, 5 experimental model studies, 7 cohort studies, and 4 case reports matched the criteria.

Conclusions: The severity of the cytokine storm, measured by elevated levels of interleukin-1B, interferon-γ, interferon-inducible protein 10, and monocyte chemoattractant protein 1, was associated with COVID-19 disease severity. Many treatment options with different targets have been proposed during the early stages of the COVID-19 pandemic, ranging from targeting the virus itself to managing the systemic inflammation caused by the virus and the excessive cytokine response. Among the different agents to manage cytokine storm in patients with COVID-19, there is developing support for convalescent plasma therapy particularly for patients who are critically ill or mechanically ventilated and resistant to antivirals and supportive care. Treatment options that were proposed in the beginning phases of the pandemic were multidimensional, and further research is needed to develop a more established treatment guideline.

(J Med Internet Res 2020;22(8):e20193) doi:10.2196/20193

KEYWORDS

coronavirus; COVID-19; convalescent plasma therapy; cytokine storm; SARS-CoV-2; cytokine; immunology; review; mortality; inflammation; therapy

Introduction

In December 2019, a novel coronavirus emerged from Wuhan, China. Since then, it has rapidly spread around the world. The coronavirus disease (COVID-19) ranges from a self-limiting upper respiratory tract illness to severe pneumonia, multiorgan failure, and death. It was declared a pandemic by the World Health Organization (WHO) on March 11, 2020 [1-3]. Despite its impact, there is still much uncertainty regarding the frontline treatment of choice for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

Current research indicates that cytokine storm is one of the leading causes of morbidity and mortality in patients with COVID-19. This evidence is based on elevated levels of interleukins (ILs) and other cytokine signaling pathways that are implicated in inflammatory cascades. Zhang et al [4] explored risk factors among patients with COVID-19 who are at higher risk of adverse events such as admission to an intensive care unit (ICU), mechanical ventilation dependence, or death among three hospitals in Wuhan, China from January 13, 2020, to February 26, 2020. As part of their analysis, 28 patients with cancer and COVID-19 were included. Lung cancer was the most frequent type of cancer involved with 7 of 28 (25%) patients; 8 of 28 (29%) patients were suspected to have received COVID-19 from hospital-associated transmission; 15 of 28 (54%) patients had a severe event, with an overall mortality rate of 29% [4]. Receiving antineoplastic therapy within 2 weeks of analysis significantly increased the risk of developing severe events, with a hazard ratio of 4.079 (P=.03) [4].

Convalescent plasma therapy is one of the proposed treatments for patients who are critically ill with COVID-19 and have exaggerated inflammatory response to the virus. Plasma therapy has been used in prior viral pandemics such as pandemic influenza A (H1N1), severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), and Ebola with moderate success. Here, we present a literature review regarding current and potential pharmacologic treatments that specifically aim to target the cytokine storm in the battle against COVID-19.

Methods

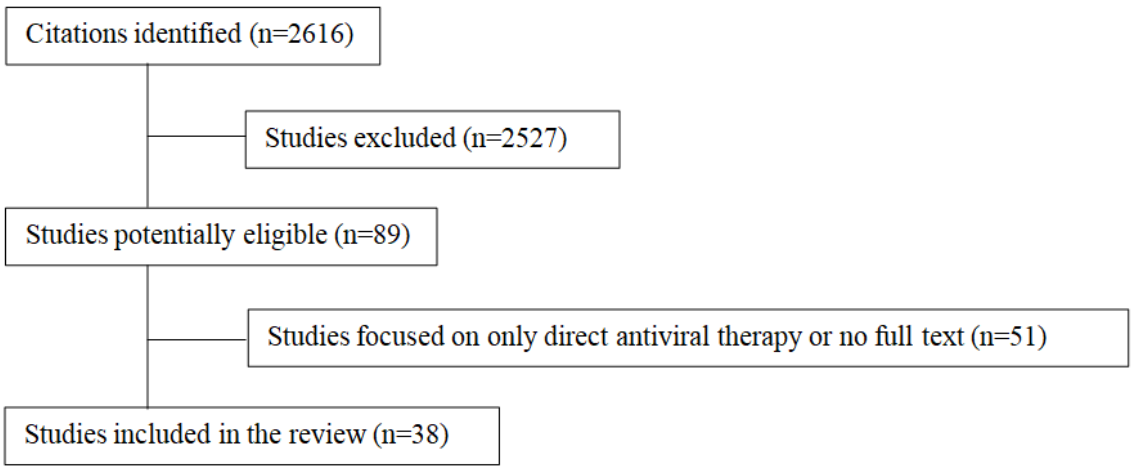
We searched the PubMed and EMBASE databases. We specifically screened studies that were published between December 1, 2000, and April 4, 2020. The following research terms were used: coronavirus, COVID-19, SARS, SARS-CoV-2, cytokine storm, inflammatory response, cytokines, ILs, immunomodulatory, anti-inflammation treatment, and immunocompromised. We excluded studies without detailed methodological reporting. The search was limited to studies published in English. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) criteria were applied.

The screening of titles and abstracts was performed by the authors. The full texts were reviewed in a second screening. Papers were considered if the study design was a case report, cohort study, series of cases, ecological study, systematic review, meta-analysis, or clinical trial related to the cytokine storm effect and treatment of coronavirus infection.

Results

Our literature search identified 2616 abstracts, from which 89 potential studies were selected for detailed full-text review. After excluding the studies that were focused on direct antiviral therapy, 41 studies remained. The studies focused on the management of coronavirus infection and cytokine storm. No nonpeer reviewed publications were excluded from the study. Studies that were not available in full text were excluded. A total of 38 studies were included in the final analysis. Of these, 24 systematic reviews, 5 meta-analyses, 5 experimental model studies, 7 cohort studies, and 4 case reports were selected. Some studies contributed to more than one section in this review (Figure 1).

Figure 1. Flow diagram of search results.



Discussion

Principal Findings

The pathogenesis of cytokine storm in patients who are critically ill with COVID-19 involves a dysregulation of the immune response. Excessive release of cytokines causes damage on a local and systemic level. Elevated cytokine responses have been associated with a high rate of morbidity and mortality. Our review explored multiple treatment strategies, ranging from direct antiviral therapy to glucocorticoids to those that specifically target the cytokine storm effect, such as immunomodulators and plasmapheresis (Table 1). As there are currently no well-established treatment guidelines, research in targeting the cytokine storm is crucial in management at the earliest stage possible.

Table 1. Clinical management studies in viral outbreaks.

Treatment	SARS-CoV-2[a]	MERS[b]	H1N1[c]	SARS[d]
Glucocorticoids	• Wang et al[e] [5] • WHO[f] [6]	Russell et al[g] [7]	WHO [6]	Russell et al[g] [7]
IL[h]-6 antibody	• Xu et al[i] [8] • Sanofi [9] • Gritti[i] [10]	N/A[j]	N/A	N/A
IL-1 antibody	• Nicastri[i] [11] • Sobi [12]	N/A	N/A	N/A
CQ[k]/HCQ[l]	• Zhang et al[i] [13] • Dong et al[g] [14] • Gautret et al[i] [15]	N/A	N/A	Zhang et al[g] [16]
JAK2[m] inhibitors	• Wu and Yang[i] [17]	Wu and Yang[i] [17]	Wu and Yang[i] [17]	N/A
Antioxidants	• Zhang et al[g] [18] • Peng[i] [19]	Zhang et al[g] [18]	N/A	Zhang et al[g] [18]
Plasma therapy	• Tanne [20] • Roback and Guarner [21] • Shen et al[i] [22]	Chun et al[e] [23]	Hung et al[i] [24]	Cheng et al[i] [25]

[a]SARS-CoV-2: severe acute respiratory syndrome coronavirus 2.

[b]MERS: Middle East respiratory syndrome.

[c]H1N1: pandemic influenza A.

[d]SARS: severe acute respiratory syndrome.

[e]Case report.

[f]WHO: World Health Organization.

[g]Review.

[h]IL: interleukin.

[i]Series of cases/observational studies.

[j]N/A: not applicable.

[k]CQ: chloroquine.

[l]HCQ: hydroxychloroquine.

[m]JAK2: Janus kinase 2.

Cytokine Storm and the Impaired Immune System

The coronavirus pneumonia, such as SARS and MERS, has demonstrated that a combination of severe inflammation, oxidative stress, and an exaggerated immune response contributes to the COVID-19 pathology. This uncontrolled and excessive load of proinflammatory cytokines, referred to as a cytokine storm, can lead to acute lung injury, acute respiratory distress syndrome (ARDS), and even death. Huang et al [26] reported that patients who are critically ill with COVID-19 have an elevated cytokine profile that resembles that of cytokine storms reported in SARS and MERS. The prospective cohort study from Wuhan, China measured the cytokine levels in 41 patients. Out of 41 patients, 13 (32%) were admitted to the ICU for high-flow nasal cannula or higher-level oxygen support requirements. The majority of the patients in this cohort presented with fever, dry cough, and dyspnea. Chest computed tomography (CT) scans showed bilateral ground-glass opacities, and some patients even progressed to ARDS in just 2 days.

The pathophysiology of COVID-19 patients with respiratory compromise is attributed to the diffuse alveolar damage that SARS-COV-2 causes, similar to what was reported in prior patients with SARS. Although the viral infection targets the alveolar epithelial cells, the coronavirus is known to have a long incubation period. During this time, the virus can cause multi-organ damage, specifically affecting the spleen, lymph nodes, small blood vessels, heart, liver, and kidney. Consequently, the impaired organs that make up the immune system lead to lymphocytopenia, which interferes with clearance of the virus. Autopsies from patients with COVID-19 and pathology from those who had pneumonectomy for lung cancer have shown varying degrees of diffuse alveolar damage, including type II epithelial cell hyperplasia, alveolar cavity fibrosis, and microvascular thrombus [27].

In addition to the virus itself causing alveolar damage, an immunocompromised system poses further challenge to treating patients who are critically ill with COVID-19 that have elevated levels of IL-1β, interferon (IFN)–γ, IFN-inducible protein 10 (IP-10), and monocyte chemoattractant protein 1 (MCP-1), which activate T-helper (Th)1 cell responses. Patients who required ICU admission had higher concentrations of granulocyte colony-stimulating factor, IP-10, MCP-1, macrophage inflammatory protein 1A, and tumor necrosis factor alpha (TNF-α) than did those not requiring ICU admission. Overall, Huang et al [26] suggested that the cytokine storm was associated with disease severity. Despite supportive treatment, mortality was as high as 6 out of the 41 (15%) patients in this cohort [26].

Early studies on patients with SARS showed that elevation of proinflammatory cytokines such as IL-1β, IL-6, IL-12, IFN-γ, IP-10, and MCP-1 was associated with pulmonary inflammation and extensive lung damage [26]. A prospective cohort study by Wong et al [28] found that patients with SARS exhibited a significant increase in the antiviral cytokine IFN-γ and proinflammatory cytokines (IL-1β, IL-6 and IL-12) during a cytokine storm, with a moderate increase in anti-inflammatory IL-10 in only some patients. This selective IL-10 response was not further analyzed. The early elevation of inflammatory cytokines was attributed to the SARS-induced activation of Th1 cells and natural killer (NK) cells, resulting in pulmonary inflammation. In a former clinical study by Wong et al [29], 128 of 157 (82%) patients with SARS developed neutrophilia associated with ICU admission and mortality from pneumonia [29]. Therefore, studies indicate that polymorphonuclear leukocyte–induced acute pneumonitis in the early stages of SARS-CoV-2 infection contributes to the later development of ARDS and ultimately the pulmonary destruction in patients with SARS.

A similar elevation of inflammatory cytokines has been reported among patients with MERS, although the theory of cytokine storm has been much less studied during the MERS outbreak. A study by Mahallawi et al [30] compared the profile of cytokine responses in plasma samples from patients with MERS (n=7) versus those of the healthy controls (n=13). A strong proinflammatory Th1 and Th17 response was seen in patients with MERS, with markedly increased concentrations of proinflammatory cytokines (IFN-γ, TNF-α, IL-15, and IL-17), that elicited a type II IFN response with innate and acquired immunity interfering with viral replication [27,31]. Patients with MERS also had a marked increase in IFN-α2 concentration compared to the healthy controls (P<.01). Patients who were infected had a significant increase in IFN-α2 ranging from 26- to 71-fold increase when compared to healthy controls, indicating a type I IFN response as first line defense against the viral infection [27].

Lymphocytopenia and Inflammatory Response

Lymphocytopenia is a prominent marker of COVID-19 that is a diagnostic criteria for COVID-19 in China. Both T cell and NK cell counts are depressed in patients with COVID-19, with pronounced reduction noted among the more severe cases [32]. A study by Zhang et al [16] examined patients with COVID-19 in Wuhan, China. Patients were categorized as having mild versus severe infection. Patients who were severe were those with extrapulmonary systemic hyperinflammation syndrome and required ICU care [16]. Among these patients, CD8+ T cells were reported to be reduced by 28.43% and 61.9% in mild and severe groups, respectively, while NK cells were reduced by 34.31% and 47.62% in mild and severe groups, respectively [16]. According to prior studies, the more severe cases of COVID-19 were associated with acute lymphocytopenia with destruction of lymphoid tissues including the spleen and lymph node. Immunohistochemical staining showed that CD4+ T cells and CD8+ T cells were decreased, which was associated with an overexpression of proinflammatory cytokines and chemokines. Although the mechanism is not conclusive, studies ultimately attribute lymphocytopenia in patients who were critically ill with COVID-19 to a dysregulation of the immune system [33].

The overall mechanism that leads to rapid respiratory failure in the setting of excessive inflammatory response remains unclear. Prominent theories propose a direct invasion via angiotensin converting enzyme 2 (ACE 2) versus a systemic inflammatory response via the cytokine storm effect. Given that there have been no findings of ACE 2 expression on lymphocytes, it is largely theorized that the lymphocyte destruction is due to cytokine storm alone. A study by Zhang et al [16] showed that in type II alveolar epithelia and macrophages, virus inclusion bodies were detected, indicating a primary cytokine storm effect. Regardless of the underlying mechanism, anti-inflammatory treatment specifically targeting the cytokine storm effect seems vital to the survival of patients who are critically ill with COVID-19, though there is a lack of evidence at this time supporting effective treatments.

Glucocorticoid Therapy

The effect of glucocorticoids has been extensively studied with SARS and MERS in the past. During the SARS outbreak in 2003, it became the primary immunomodulatory therapy with mixed results. Although glucocorticoids improved fever and oxygenation in many cases, other studies showed adverse reactions, including delayed virus clearance or even further worsening of the disease. Currently, glucocorticoids are used to suppress cytokine storm symptoms and improve ARDS.

A retrospective review of 138 patients in Wuhan, China by Wang et al [5] examined characteristics of hospital-related transmission of the illness. The review also analyzed a variety of treatments including antivirals, antibiotics, and glucocorticoids. Although it did not control for any variables, the study found significant treatment differences between patients in the ICU and those not in the ICU, with a trend toward patients in the ICU receiving glucocorticoids. Out of 138 patients, 62 (44.9%) received glucocorticoids. Of the 36 patients admitted to the hospital and transferred to the ICU, 26 (72%) patients were receiving glucocorticoids.

Russel et al [7] reported in a literature review that patients during the SARS and MERS outbreaks who received glucocorticoids were more likely to require mechanical ventilation, vasopressor support, and renal replacement therapy. It also did not improve 90-day mortality (adjusted odds ratio 0.8, 95% CI 0.5-1.1;

P=.12). Rather, the results indicated delayed clearance of viral RNA from respiratory tract secretions (adjusted hazard ratio 0.4, 95% CI 0.2-0.7; *P*<.001) [7]. Because of the overall poor clinical support for glucocorticoid therapy, the current WHO guideline does not support glucocorticoid use for treatment of viral pneumonia and ARDS for COVID-19 cases without concurrent conditions such as chronic obstructive pulmonary disease or asthma that would independently require glucocorticoid use [6].

IL-6 Antibodies

Tocilizumab

Tocilizumab (TCZ) is a recombinant human IL-6 monoclonal antibody that inhibits IL-6 signaling and mediation of inflammatory response. Xu et al [8] conducted a retrospective study to determine the efficacy of TCZ in addition to the typical treatment with Lopinavir, Methylprednisolone, and supplemental oxygen to treat patients who are severely ill with COVID-19 [8]. Severity was defined by the Diagnosis and Treatment Protocol for Novel Coronavirus Criteria sponsored by the National Health Commission of the People's Republic of China and was defined if one of the following conditions were met: respiratory rate≥30 breaths/minute, SpO2≤93% while breathing room air, or PaO2/FiO2 ratio≤300 mmHg. TCZ 400 mg was intravenously administered once to 21 patients. Prior to treatment with TCZ, inflammatory markers including c-reactive protein (CRP) and IL-6 were measured, with a mean CRP value of 75.06 mg/L and mean IL-6 level of 132.38 pg/ml [8]. Patients' fevers improved within a few days, and 15 of 20 (75%) patients that required oxygen had improved oxygenation, with opacity of lung lesions on CT scans showing improvement in 19 of 21 (91%) patients. The percentage of peripheral lymphocytes returned to normal in 10 of 19 (53%) patients with measured levels. Additionally, CRP levels returned to normal in 16 of 19 (84%) patients by the fifth day of treatment. IL-6 levels, unfortunately, were not measured after treatment [8]. Overall, the data suggested that TCZ may be an effective treatment in severe COVID-19 cases. Based on promising results from the Xu et al [8] study, Sanofi and Regeneron pharmaceuticals received approval to open a clinical trial on Sarilumab, another anti–IL-6 monoclonal antibody in mid-March 2020 [9].

Siltuximab

Siltuximab is an IL-6 targeted monoclonal antibody that has been approved by numerous regulatory bodies including the Food and Drug Administration (FDA) in the treatment of Castleman's disease who are negative for HIV and human herpesvirus-8. Gritti et al [10] began enrollment into the Siltuximab in Serious COVID-19 study at the Papa Giovanni XXIII Hospital in Bergamo, Italy in March of 2020 as a retrospective cohort study examining patients receiving siltuximab versus matched controls [10]. Their primary endpoints were the need for invasive ventilation, time spent in ICU, and 30-day mortality. Preliminary results included 21 patients that received a dose of 11 mg/kg siltuximab given over 1 hour with a second dose per investigator discretion. There were 5 patients that received the second dose 48-72 hours after the initial infusion. Clinical improvement was observed in 7 of the 21 (33%) patients, 9 (43%) patients stabilized clinically (no change or worsening in condition), and 5 (24%) patients experienced worsening in condition. Of the 5 that worsened, 1 died and 1 experienced a stroke. CRP levels were measured in 16 patients, with improvement in the days after administration. The study is currently ongoing, and the currently released data is promising in that it offers a potential therapy for COVID-19.

Sarilumab

Sarilumab is an additional IL-6 targeted monoclonal antibody that has been approved in the treatment of rheumatoid arthritis [9]. TThere is currently no data for the use of sarilumab in viral syndromes. Given available data for TCZ, additional IL-6 therapies are being investigated. There is currently a phase II/III randomized, double-blind, placebo-controlled trial to evaluate response in patients who are severely ill with COVID-19. The primary endpoint is reduction of fever, and the secondary endpoint is decreased demand for supplemental oxygen.

IL-1 Antibodies

Anakinra is a targeted monoclonal antibody IL-1 receptor antagonist that is being investigated. Due to promising current data on TCZ, anakinra is being investigated in a phase II/III open label, controlled, parallel, three arm study versus Emapalumab or standard of care [11]. Emapalumab is a monoclonal antibody against IFN-γ, which is currently being used in the treatment of hemophagocytic lymphohistiocytosis, a life-threatening disease caused by excessive immune activation [12].

Chloroquine and Hydroxychloroquine

Chloroquine (CQ) and hydroxychloroquine (HCQ) have been used for decades as the primary choice for prophylaxis and treatment of malaria but also have had effect on the Marburg virus, Zika virus, dengue virus, Ebola virus, and SARS. CQ blocks viral infection by altering the endosomal pH that is required for viral particles to bind to the cell surface receptor. CQ also interferes with the glycosylation of SARS cellular receptors, specifically ACE 2. Its immunomodulatory, anti-inflammatory, and antiviral mechanism works synergistically, making it a choice for their efficacy and side-effect profile [16].

An early clinical trial conducted in patients with COVID-19 in China showed better clinical outcome and earlier viral clearance in patients treated with CQ compared to control groups [13,14]. Based on these positive findings, Chinese experts recommended that patients with mild, moderate, and severe cases of COVID-19 pneumonia be treated with 500 mg CQ twice per day for 10 days [14].

A clinical trial by Gautret et al [15] studied the effect of HCQ in patients with SARS-CoV-2 from hospitals in South France. A total of 26 patients received oral HCQ 200 mg three times per day for 10 days, and 16 were control patients. Among the HCQ-treated patients, 6 patients also received azithromycin (500 mg on day 1, followed by 250 mg per day for the next four days) to prevent bacterial superinfection. Each day, patients received a standardized clinical examination along with a nasopharyngeal sample collection. Additionally, 6 patients

treated with HCQ were lost in follow-up by survey due to early cessation of treatment. Despite a small sample size, the study showed that the HCQ treatment was associated with clinical improvement, faster viral load reduction, and effect reinforcement with azithromycin [13]. At day 6 postinclusion, 14 of the 20 (70%) HCQ-treated patients were virologically cured compared with the 2 of 16 (13%) in the control group (P=.001). All patients treated with HCQ and azithromycin combination were virologically cured compared to the 8 of 14 (57.1%) in patients treated with HCQ and 2 of 16 (13%) in the control group (P<.001). One patient who was still polymerase chain reaction (PCR)–positive on day six postinclusion under HCQ treatment only received azithromycin in addition to the HCQ on day eight postinclusion; the viral load was cleared on day nine postinfection. The clearing of the viral nasopharyngeal carriage of SARS-CoV-2 is critical, as studies have shown that the mean duration of viral shedding in patients with COVID-19 in China was 20 days [13]. Given the prolonged viral shedding period, transmissibility is a critical issue, and further investigation and trials are needed to determine the efficacy of these agents.

Janus Kinase 2 Inhibitors

Patients with severe COVID-19 syndrome have elevated counts of Th17 cells contributing to the development of cytokine storm. Wu and Yang [17] proposed the use of the Janus kinase 2 (JAK2) inhibitor fedratinib for the suppression of Th17 cytokine production. JAK2 works through signal transducer and activator of transcription (STAT)3 to mediate IL-6 and IL-23 signals for Th17 cells. JAK1 and tyrosine kinase 2 receptors work through STAT1 and STAT2, which are important to the function of antiviral immunity. Wu and Yang [17], thus, proposed selective targeting of JAK2 inhibition to mitigate or even prevent the cytokine storm in COVID-19. In their previous testing in vitro, they found that Fedratinib decreased expression of IL-17 and IL-22 expression [17]. Fedratinib is currently approved by the FDA for use in myeloproliferative neoplasms that take advantage of the JAK2 pathway, such as polycythemia vera; however, it has not been tested in the setting of viral syndromes yet.

Targeting the Antioxidant Effect

Other therapies such as melatonin and vitamin C have been implicated in anti-inflammatory and antioxidant effects, and could potentially prove beneficial in the COVID-19 setting. There is currently a clinical trial underway for vitamin C; however, melatonin has not been examined yet at this time.

Melatonin

Although many therapies have been proposed based on previous use in viral pandemics, others are proposed simply due to beneficial effects seen in other nonviral conditions. Melatonin, a drug typically used for insomnia, has been found to have beneficial effects in other diseases such as atherosclerosis and respiratory distress. Per Zhang et al [18], in contrast to other coronaviruses, SARS-CoV-2 infection leads to increased levels of IL-1β, IFN-γ, IP-10, and MCP-1, in addition to IL-4 and IL-10, which was diminished in many patients with SARS. The study suggested that melatonin mediates anti-inflammatory effects through sirtuin 1, which inhibits proteins, down regulates macrophages toward proinflammation, and attenuates lung injury and inflammation. It has also been shown to suppress proinflammatory cytokines and regulate nucleotide-binding oligomerization domain-like receptors 3 in radiation-induced lung injury.

Melatonin may also have antioxidative effects in the setting of ARDS. Previous studies by Gitto et al [34] used melatonin in newborns with respiratory distress with success due to antioxidant and anti-inflammatory effects. In a study of patients with multiple sclerosis, melatonin has been shown to reduce serum concentrations of TNF-α, IL-6, IL-1β, and lipoperoxides, further leading to evidence that it may reduce inflammatory cytokines [35]. Testing melatonin in ICU patients in the past has not shown significant adverse events. Given the high safety profile, Zhang et al [18] proposed the use of melatonin for study in patients with COVID-19 [36].

Vitamin C

Vitamin C, or ascorbic acid, is a water-soluble vitamin that aids in the synthesis of collagen in connective tissue and may also work as an antioxidant. Studies in chicken embryo cultures demonstrated that vitamin C provided resistance to coronavirus infection. Furthermore, there have been human trials that are suggestive of a lower incidence of pneumonia in groups on vitamin C treatment [19,37].

Given previous data on vitamin C, there is currently an ongoing prospective randomized clinical trial in Wuhan, China examining 7 days of high dose intravenous (IV) vitamin C therapy versus placebo with endpoints being ventilator requirements, vasopressor requirements, ICU length of stay, and 28-day mortality. The full results of the study are expected to be released by September 2020 [19].

Convalescent Plasma Therapy

The search for COVID-19 treatment has extended to plasma therapy, which has been used in previous viral outbreaks such as SARS, Ebola, and H1N1. Convalescent plasma therapy relies on the purified plasma obtained usually by apheresis from patients that have recovered from the particular infection and subsequently transfused to ABO-compatible patients with the active disease. Given previous success, this therapy is currently being proposed in patients who are critically ill who have failed more conservative management. In fact, as of March 2020, the FDA approved the use of convalescent plasma therapy to treat patients who are critically ill with COVID-19, provided that the treating physicians apply via the process for an investigational new drug application [20]. Here, we aim to discuss plasma therapy in patients who are critically ill, as it has been used in prior respiratory viral outbreaks and potential applications in the current pandemic, including efficacy in inflammatory syndromes.

SARS Treatment in Hong Kong in 2003

During the height of the SARS epidemic in 2003, Cheng et al [25] explored the efficacy of convalescent plasma therapy. The study included 339 patients that were admitted at the Prince of Wales Hospital in Hong Kong between March 20 and May 26

of 2003 for SARS. Initial treatment was with Cefotaxime and Levofloxacin (or clarithromycin) to treat community-acquired pneumonia. If fever persisted, patients were treated with Ribavirin at 1200 mg per os *ter in die* or IV 400 mg every 8 hours and Prednisolone (0.5-1 mg/kg) on day three. Patients with radiographic progression and hypoxemia were given pulsed methylprednisolone 500 mg IV daily for 2-3 doses. For patients that continued to deteriorate per SaO2<90% on 0.5 FiO2, 200-400 ml (4-5 ml/kg) of ABO-compatible convalescent plasma obtained from recovered patients with SARS was administered. Out of 339 patients, 80 (24%) received convalescent plasma treatment.

Of the 80 patients selected for convalescent plasma treatment, 33 (41%) had good outcomes (discharge by day 22), and 47 (59%) had poor outcomes (hospitalization beyond 22 days or death). The median age of patients with good outcomes was 37.9, and the median age of patients with poor outcomes was 50.2 (P<.001). The median day of plasma infusion was 11.7 for patients with good outcomes and 16 for patients with poor outcomes (P<.001). Patients given convalescent plasma before day 14 tended toward a better outcome, with a mortality of 6.3% (3/48) versus 22% (7/32) in the after day 14 group (P=.08). Out of 33 patients with good outcomes, 20 (61%) were PCR-positive and seronegative for coronavirus compared to 10 of 47 (21%) with poor outcome (P<.001). The 30 patients that were PCR-positive and seronegative at the time of convalescent plasma therapy had better outcomes than patients already seropositive, 67% (20/30) versus 20% (10/50) (P=.001). There were no reported adverse effects after administration of plasma infusions; however, the primary limitation of plasma convalescent therapy was availability. The mortality rate recorded in the study was 13% (10/80), while the SARS mortality rate in Hong Kong during that same time period was 17% (299/1755) from March 6 to May 24, 2003 [25].

H1N1 Treatment in Hong Kong in 2009

During the H1N1 virus pandemic of 2009, Hung et al [24] examined the effects of convalescent plasma treatment in patients who were critically ill in Hong Kong via a prospective cohort study. From September 1, 2009, to June 30, 2010, patients older than 18 years with H1N1 infection requiring intensive care were offered treatment with convalescent plasma harvested via apheresis from suitable donors. A total of 93 patients were recruited with 20 patients receiving plasma therapy. The remaining patients declined plasma treatment and were included as untreated controls. Donors were selected by reverse transcriptase-PCR (RT-PCR)–positive influenza A virus M and pandemic H1 genes and negative RT-PCR testing of seasonal influenza A virus H1 and H3 genes with clinical recovery from infection at least 2 weeks in addition to meeting standard plasma donation criteria. All patients received treatment with oseltamivir and inhaled zanamivir for influenza.

Of the patients that received convalescent plasma therapy versus patients who were untreated, mortality was found to be 20% (4/20) versus 55% (40/73) with P=.01 [24]. Overall, the mortality was 47% (44/93). Other measured factors included viral load, IL-6, IL-10, and TNF-α, which were all found to be significantly lower in the treatment group on subsequent measurements. No reported adverse events were recorded from plasma therapy.

SARS-CoV-2 Treatment in Shenzhen, China in 2020

One of the earliest studies of convalescent plasma therapy for SARS-CoV-2 was done by Shen et al [22] in Shenzhen, China. From January 20, 2020, through March 25, 2020, at Shenzhen Third People's Hospital, 5 patients who were critically ill with COVID-19 were treated with convalescent plasma transfusion [22]. Selection criteria included COVID-19 diagnosis via RT-PCR, severe pneumonia with rapid progression (P_aO_2/F_iO_2<300 with mechanical ventilatory support), and continuously high viral load despite antiviral treatment.

Treatment prior to plasma therapy included methylprednisolone, lopinavir/ritonavir, IFN-α-1b, arbidol, and favipiravir. One day prior to receiving treatment, the patient's blood was collected and enzyme-linked immunosorbent assay (ELISA) and neutralizing antibody titers were tested. On the day of treatment, ABO blood type and compatibility with convalescent plasma donors were measured. Patients were then treated with two consecutive transfusions of 200-250 mL of ABO-compatible convalescent plasma for 400 mL in total with ongoing treatment for antiviral agents until SARS-CoV-2 viral loads were negative [22].

Convalescent donors were between 18 and 60 years of age, and had recovered from SARS-CoV-2 infection. They were previously diagnosed with COVID-19 by a quantitative RT-PCR and tested negative for SARS-CoV-2, respiratory viruses, hepatitis B, hepatitis C, HIV, and syphilis. They were also required to be asymptomatic for at least 10 days and have SARS-CoV-2 specific ELISA antibody titer higher than 1:1000 and a neutralizing antibody titer greater than 40. Donation of 400mL of convalescent plasma was then obtained by apheresis and immediately transfused to recipients [22].

Administration of plasma was between 10 and 22 days with the median being day 20. Patients were between 30 and 70 years of age. Of the 5 patients, only 1 had pre-existing conditions of hypertension and mitral insufficiency, and none had a history of smoking. The measured characteristics before and after transfusion included body temperature, cycle threshold value, procalcitonin, CRP, and IL-6 levels [22].

The results noted improvement of cycle threshold value 1 day posttransfusion, with negative testing achieved between days 1 and 12 posttransfusion. IL-6, CRP, and procalcitonin trended downward. Of the 5 patients, 3 were able to be weaned from ventilation, and one was able to be weaned from extracorporeal membrane oxygenation to mechanical ventilation. Donor's plasma receptor binding domain immunoglobulin (Ig)G and IgM, and neutralizing antibody titers were measured via a cell culture to determine inhibition activity against SARS-CoV-2 [22].

Risks and Challenges to Plasma Therapy

Convalescent plasma therapy involves strict quality control to ensure that there is adequate neutralizing antibodies present in the obtained plasma. However, even with quality control, it is not without risk. The risks of convalescent plasma include the

same risks as those with transfusion of any blood products: anaphylactic shock, transfusion-associated circulatory overload, and transfusion-related acute lung injury (TRALI).

During the MERS outbreak in May 2015, convalescent plasma therapy was used in many patients. Chun et al [23] described 1 such patient that experienced an unfortunate complication from this therapy. The patient was a 32-year-old male who was admitted and diagnosed with MERS and subsequently treated with ribavirin and lopinavir/ritonavir, as well as a single dose of IFN-α-2a. ABO/RhD-compatible convalescent plasma was then obtained from a patient who had previously recovered from MERS. Within 2 hours of transfusion, the patient developed respiratory distress that was concerning for TRALI despite apparent compatibility. Viral load was found to be reduced in the patient, and retrospective analysis was not able to trace anti–human leukocyte antigen (HLA) or anti-HNA antibodies.

Investigation into these issues has shown that the additive risk of HLA in female donors with one childbirth can be as high as 14.5%, arguing for a preference toward male donors to reduce this risk [38-40]. Since then, many studies and authors have suggested the use of only male donors citing a previous review [39].

In addition, the limitations in the case series published by Shen et al [22] and Roback and Guarner [21] in their editorial to JAMA reported difficulties in finding standard antibody doses from donors of convalescent plasma [21,22]. Therefore, they proposed combining convalescent plasma with hyperimmune globulin (H-Ig). Unlike convalescent plasma, H-Ig has standardized antibody doses from collecting plasma samples from patients with high antibody titers and filtering out IgM antibodies. H-Ig and convalescent plasma may both be stored for years. Monitoring patient response in comparison to objective lab values such as antibody titers and neutralizing activity could help identify which patients can be donors or recipients [21].

Conclusion

SARS-CoV-2 is a global pandemic that has challenged the world of medicine, research, and health care. The purpose of this paper is to highlight the important current and developing therapies in the management of severe COVID-19, in particular, therapies that can aid in the syndrome known as cytokine storm.

Immunomodulators such as siltuximab, anakinra, sarilumab, fedratinib, and TCZ target different components of the inflammatory cascade involved in cytokine storm and are in clinical trials. Data is currently strongest for TCZ; however, the results of these studies are not all ready, and investigators may wish to evaluate if patients could be candidates for any one of these trials.

Although convalescent plasma therapy works by using passively produced antibodies from patients recovered from the illness to treat patients in active disease states, a preliminary case series from China has found that there may also be activity that reduces the markers of cytokine storm such as IL-6. Further use of this therapy is underway in multiple nations around the world in various clinical trials, and it is a promising avenue of treatment for patients who are critically ill.

The treatment for COVID-19 has been met with dire need and urgency during this global pandemic. Current treatments are essentially limited to supportive care and treatment of secondary infections without directly targeted therapies to either the virus or the hypothesized cytokine storm. Although there are a number of clinical trials underway with constantly new therapies being proposed for patients who are critically ill, further study and concrete data are needed to establish treatment guidelines.

Conflicts of Interest

None declared.

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020
2. Wilson N, Kvalsvig A, Barnard LT, Baker MG. Case-fatality risk estimates for COVID-19 calculated by using a lag time for fatality. Emerg Infect Dis 2020 Jun;26(6):1339-1441. [doi: 10.3201/eid2606.200320] [Medline: 32168463]
3. CDC COVID-19 Response Team. Severe outcomes among patients with coronavirus disease 2019 (COVID-19) - United States, February 12-March 16, 2020. MMWR Morb Mortal Wkly Rep 2020 Mar 27;69(12):343-346. [doi: 10.15585/mmwr.mm6912e2] [Medline: 32214079]
4. Zhang L, Zhu F, Xie L, Wang C, Wang J, Chen R, et al. Clinical characteristics of COVID-19-infected cancer patients: a retrospective case study in three hospitals within Wuhan, China. Ann Oncol 2020 Jul;31(7):894-901 [FREE Full text] [doi: 10.1016/j.annonc.2020.03.296] [Medline: 32224151]
5. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA 2020 Feb 07:1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
6. World Health Organization. Clinical management of severe acute respiratory infection (SARI) when COVID-19 disease is suspected. Interim guidance. Pediatr Med Rodz 2020 Jul 14;16(1):9-26. [doi: 10.15557/pimr.2020.0003]
7. Russell CD, Millar JE, Baillie JK. Clinical evidence does not support corticosteroid treatment for 2019-nCoV lung injury. Lancet 2020 Feb;395(10223):473-475. [doi: 10.1016/s0140-6736(20)30317-2]

8. Xu X, Han M, Li T, Sun W, Wang D, Fu B, et al. Effective treatment of severe COVID-19 patients with tocilizumab. Proc Natl Acad Sci U S A 2020 May 19;117(20):10970-10975 [FREE Full text] [doi: 10.1073/pnas.2005615117] [Medline: 32350134]
9. Sanofi and Regeneron begin global Kevzara® (sarilumab) clinical trial program in patients with severe COVID-19. Sanofi. 2020 Mar 16. URL: http://www.news.sanofi.us/2020-03-16-Sanofi-and-Regeneron-begin-global-Kevzara-R-sarilumab-clinical-trial-program-in-patients-with-severe-COVID-19 [accessed 2020-03-19]
10. Gritti G, Raimondi F, Ripamonti D, et al. IL-6 signalling pathway inactivation with siltuximab in patients with COVID-19 respiratory failure: an observational cohort study. medRxiv 2020 Jun 20:A. [doi: 10.1101/2020.04.01.20048561]
11. Nicastri E. Efficacy and safety of Emapalumab and Anakinra in reducing hyperinflammation and respiratory distress in patients with COVID-19 infection. ClinicalTrials.gov. URL: https://clinicaltrials.gov/ct2/show/study/NCT04324021 [accessed 2020-03-30]
12. Sobi to initiate a clinical study to evaluate whether anakinra and emapalumab may relieve complications associated with severe COVID-19 disease. Sobi. 2020 Mar 18. URL: https://www.sobi.com/sites/default/files/pr/202003183346-1.pdf [accessed 2020-03-30]
13. Zhang Q, Wang Y, Qi C, Shen L, Li J. Clinical trial analysis of 2019-nCoV therapy registered in China. J Med Virol 2020 Jun;92(6):540-545 [FREE Full text] [doi: 10.1002/jmv.25733] [Medline: 32108352]
14. Dong L, Hu S, Gao J. Discovering drugs to treat coronavirus disease 2019 (COVID-19). Drug Discov Ther 2020;14(1):58-60. [doi: 10.5582/ddt.2020.01012] [Medline: 32147628]
15. Gautret P, Lagier J, Parola P, et al. Hydroxychloroquine and Azithromycin as a treatment of COVID-19: preliminary results of an open-label non-randomized clinical trial. medRxiv 2020 Jul:A. [doi: 10.1101/2020.03.16.20037135]
16. Zhang W, Zhao Y, Zhang F, Wang Q, Li T, Liu Z, et al. The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (COVID-19): the perspectives of clinical immunologists from China. Clin Immunol 2020 May;214:108393 [FREE Full text] [doi: 10.1016/j.clim.2020.108393] [Medline: 32222466]
17. Wu D, Yang XO. TH17 responses in cytokine storm of COVID-19: an emerging target of JAK2 inhibitor Fedratinib. J Microbiol Immunol Infect 2020 Jun;53(3):368-370 [FREE Full text] [doi: 10.1016/j.jmii.2020.03.005] [Medline: 32205092]
18. Zhang R, Wang X, Ni L, Di X, Ma B, Niu S, et al. COVID-19: Melatonin as a potential adjuvant treatment. Life Sci 2020 Jun 01;250:117583 [FREE Full text] [doi: 10.1016/j.lfs.2020.117583] [Medline: 32217117]
19. Peng Z. Vitamin C infusion for the treatment of severe 2019-nCoV infected pneumonia. ClinicalTrials.gov. 2020 Feb 11. URL: https://clinicaltrials.gov/ct2/show/NCT04264533 [accessed 2020-04-03]
20. Tanne JH. Covid-19: FDA approves use of convalescent plasma to treat critically ill patients. BMJ 2020 Mar 26;368:m1256. [doi: 10.1136/bmj.m1256] [Medline: 32217555]
21. Roback JD, Guarner J. Convalescent plasma to treat COVID-19: possibilities and Challenges. JAMA 2020 Mar 27:A. [doi: 10.1001/jama.2020.4940] [Medline: 32219429]
22. Shen C, Wang Z, Zhao F, Yang Y, Li J, Yuan J, et al. Treatment of 5 critically ill patients with COVID-19 with convalescent plasma. JAMA 2020 Mar 27:1582-1589 [FREE Full text] [doi: 10.1001/jama.2020.4783] [Medline: 32219428]
23. Chun S, Chung CR, Ha YE, Han TH, Ki C, Kang E, et al. Possible transfusion-related acute lung injury following convalescent plasma transfusion in a patient with Middle East respiratory syndrome. Ann Lab Med 2016 Jul;36(4):393-395 [FREE Full text] [doi: 10.3343/alm.2016.36.4.393] [Medline: 27139619]
24. Hung IF, To KK, Lee C, Lee K, Chan K, Yan W, et al. Convalescent plasma treatment reduced mortality in patients with severe pandemic influenza A (H1N1) 2009 virus infection. Clin Infect Dis 2011 Feb 15;52(4):447-456. [doi: 10.1093/cid/ciq106] [Medline: 21248066]
25. Cheng Y, Wong R, Soo YOY, Wong WS, Lee CK, Ng MHL, et al. Use of convalescent plasma therapy in SARS patients in Hong Kong. Eur J Clin Microbiol Infect Dis 2005 Jan;24(1):44-46 [FREE Full text] [doi: 10.1007/s10096-004-1271-9] [Medline: 15616839]
26. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
27. Zhang T, Sun LX, Feng RE. [Comparison of clinical and pathological features between severe acute respiratory syndrome and coronavirus disease 2019]. Zhonghua Jie He He Hu Xi Za Zhi 2020 Jun 12;43(6):496-502. [doi: 10.3760/cma.j.cn112147-20200311-00312] [Medline: 32241072]
28. Wong R, Wu A, To K, Lee N, Lam CWK, Wong CK, et al. Haematological manifestations in patients with severe acute respiratory syndrome: retrospective analysis. BMJ 2003 Jun 21;326(7403):1358-1362 [FREE Full text] [doi: 10.1136/bmj.326.7403.1358] [Medline: 12816821]
29. Wong CK, Lam CWK, Wu AKL, Ip WK, Lee NLS, Chan IHS, et al. Plasma inflammatory cytokines and chemokines in severe acute respiratory syndrome. Clin Exp Immunol 2004 Apr;136(1):95-103 [FREE Full text] [doi: 10.1111/j.1365-2249.2004.02415.x] [Medline: 15030519]
30. Mahallawi WH, Khabour OF, Zhang Q, Makhdoum HM, Suliman BA. MERS-CoV infection in humans is associated with a pro-inflammatory Th1 and Th17 cytokine profile. Cytokine 2018 Apr;104:8-13. [doi: 10.1016/j.cyto.2018.01.025] [Medline: 29414327]

31. Gattoni A, Parlato A, Vangieri B, Bresciani M, Derna R. Interferon-gamma: biologic functions and HCV terapy (type I/II) (2 of 2 parts). Clin Ter 2006;157(5):457-468. [Medline: 17147054]
32. The diagnosis and treatment plan for 2019-nCoV (the Seventh Trial Edition). National Health Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml [accessed 2020-03-04]
33. Qin C, Zhou L, Hu Z, Zhang S, Yang S, Tao Y, et al. Dysregulation of immune response in patients with COVID-19 in Wuhan, China. SSRN J 2020 Mar 02:A. [doi: 10.2139/ssrn.3541136]
34. Gitto E, Reiter RJ, Sabatino G, Buonocore G, Romeo C, Gitto P, et al. Correlation among cytokines, bronchopulmonary dysplasia and modality of ventilation in preterm newborns: improvement with melatonin treatment. J Pineal Res 2005 Oct;39(3):287-293. [doi: 10.1111/j.1600-079x.2005.00251.x]
35. Sánchez-López AL, Ortiz GG, Pacheco-Moises FP, Mireles-Ramírez MA, Bitzer-Quintero OK, Delgado-Lara DL, et al. Efficacy of melatonin on serum pro-inflammatory cytokines and oxidative stress markers in relapsing remitting multiple sclerosis. Arch Med Res 2018 Aug;49(6):391-398. [doi: 10.1016/j.arcmed.2018.12.004] [Medline: 30595364]
36. Mistraletti G, Umbrello M, Sabbatini G, Miori S, Taverna M, Cerri B, et al. Melatonin reduces the need for sedation in ICU patients: a randomized controlled trial. Minerva Anestesiol 2015 Dec;81(12):1298-1310 [FREE Full text] [Medline: 25969139]
37. Kashiouris MG, L'Heureux M, Cable CA, Fisher BJ, Leichtle SW, Fowler AA. The emerging role of vitamin c as a treatment for sepsis. Nutrients 2020 Jan 22;12(2):292 [FREE Full text] [doi: 10.3390/nu12020292] [Medline: 31978969]
38. MacLennan S, Barbara JA. Risks and side effects of therapy with plasma and plasma fractions. Best Pract Res Clin Haematol 2006;19(1):169-189. [doi: 10.1016/j.beha.2005.01.033] [Medline: 16377549]
39. Schmickl CN, Mastrobuoni S, Filippidis FT, Shah S, Radic J, Murad MH, et al. Male-predominant plasma transfusion strategy for preventing transfusion-related acute lung injury. Crit Care Med 2015;43(1):205-225. [doi: 10.1097/ccm.0000000000000675]
40. Handbook of Transfusion Medicine. Norwich: The Stationery Office; 2001.

Abbreviations

ACE 2: angiotensin converting enzyme 2
ARDS: acute respiratory distress syndrome
COVID-19: coronavirus disease
CQ: chloroquine
CRP: c-reactive protein
CT: computed tomography
ELISA: enzyme-linked immunosorbent assay
FDA: Food and Drug Administration
H1N1: pandemic influenza A
HCQ: hydroxychloroquine
HLA: human leukocyte antigen
H-Ig: hyperimmune globulin
ICU: intensive care unit
IFN: interferon
Ig: immunoglobulin
IL: interleukin
IP-10: interferon-inducible protein 10
IV: intravenous
JAK2: Janus kinase 2
MCP-1: monocyte chemoattractant protein 1
MERS: Middle East respiratory syndrome
NK: natural killer
PCR: polymerase chain reaction
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RT-PCR: reverse transcriptase-polymerase chain reaction
SARS: severe acute respiratory syndrome
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
STAT: signal transducer and activator of transcription
TCZ: tocilizumab
Th1: T-helper
TNF-α: tumor necrosis factor alpha
TRALI: transfusion-related acute lung injury

WHO: World Health Organization

Edited by G Eysenbach; submitted 13.05.20; peer-reviewed by X Yang, D Saleh; comments to author 19.06.20; revised version received 03.07.20; accepted 22.07.20; published 13.08.20.

Please cite as:
Quirch M, Lee J, Rehman S
Hazards of the Cytokine Storm and Cytokine-Targeted Therapy in Patients With COVID-19: Review
J Med Internet Res 2020;22(8):e20193
URL: http://www.jmir.org/2020/8/e20193/
doi:10.2196/20193
PMID:32707537

Review

Social, Behavioral, and Cultural factors of HIV in Malawi: Semi-Automated Systematic Review

Amaury Thiabaud[1*], PhD; Isotta Triulzi[1,2*], MSc; Erol Orel[1], DEA; Kali Tal[1,3], PhD; Olivia Keiser[1], PhD

[1]Institut de Santé Globale, Université de Genève, Genève, Switzerland

[2]Institute of Management, Scuola Superiore Sant'Anna, Pisa, Italy

[3]Institute of Primary Health Care (BIHAM), University of Bern, Bern, Switzerland

[*]these authors contributed equally

Corresponding Author:
Amaury Thiabaud, PhD
Institut de Santé Globale
Université de Genève
Chemin des mines 9
Genève, 1202
Switzerland
Phone: 41 22 379 81 30
Email: amaury.thiabaud@unige.ch

Abstract

Background: Demographic and sociobehavioral factors are strong drivers of HIV infection rates in sub-Saharan Africa. These factors are often studied in qualitative research but ignored in quantitative analyses. However, they provide in-depth insight into the local behavior and may help to improve HIV prevention.

Objective: To obtain a comprehensive overview of the sociobehavioral factors influencing HIV prevalence and incidence in Malawi, we systematically reviewed the literature using a newly programmed tool for automatizing part of the systematic review process.

Methods: Due to the choice of broad search terms ("HIV AND Malawi"), our preliminary search revealed many thousands of articles. We, therefore, developed a Python tool to automatically extract, process, and categorize open-access articles published from January 1, 1987 to October 1, 2019 in the PubMed, PubMed Central, JSTOR, Paperity, and arXiV databases. We then used a topic modelling algorithm to classify and identify publications of interest.

Results: Our tool extracted 22,709 unique articles; 16,942 could be further processed. After topic modelling, 519 of these were clustered into relevant topics, of which 20 were kept after manual screening. We retrieved 7 more publications after examining the references so that 27 publications were finally included in the review. Reducing the 16,942 articles to 519 potentially relevant articles using the software took 5 days. Several factors contributing to the risk of HIV infection were identified, including religion, gender and relationship dynamics, beliefs, and sociobehavioral attitudes.

Conclusions: Our software does not replace traditional systematic reviews, but it returns useful results to broad queries of open-access literature in under a week, without a priori knowledge. This produces a "seed dataset" of relevance that could be further developed. It identified known factors and factors that may be specific to Malawi. In the future, we aim to expand the tool by adding more social science databases and applying it to other sub-Saharan African countries.

(J Med Internet Res 2020;22(8):e18747) doi:10.2196/18747

KEYWORDS

HIV/AIDS; topic modelling; text mining; Malawi; risk factors; machine learning

Introduction

Demographic and sociobehavioral factors are strong drivers of HIV in sub-Saharan Africa, but the interactions between these factors — the way their influence shifts over time and space and influences HIV prevalence and incidence — are poorly understood. Some epidemiological studies have reported on the spatial variability of the HIV epidemic, using statistical analyses to assess the association between the spatial distribution of HIV prevalence and potential risk factors [1-7]. They found, for

example, that high population density [6] or a short distance to a road or a clinic [7] were associated with a high HIV prevalence. Tomita et al [8] showed that behavior (sexual debut, uptake of contraception, and circumcision) and social determinants strongly influenced the risk of HIV acquisition. An analysis of 29 sub-Saharan African countries found associations between 12 demographic and sociobehavioral factors, including variables related to age, literacy, HIV knowledge, domestic violence, women's empowerment, and sexual activity [9]. The patterns of associations were complex and varied by sex and country, but the study did not include many potentially significant factors because they were absent from the data or were only available for some countries and the study did not consider subnational variation. These epidemiological studies did not draw on qualitative research, and they rarely contextualized the associations they identified.

Social scientists of various disciplines have performed qualitative studies of social and cultural factors related to HIV, providing rich detail on the perceptions and behaviors of people in specific localities. For example, medical anthropologists have examined maternal care-seeking behavior in different geographic regions and groups [10]. Cultural studies have analyzed connections between belief in witchcraft and folk epidemiological wisdom about HIV [11] and the connection between women's educational level and attendance at antenatal care [12]. Sociologists have studied the perceptions of HIV testing in rural Malawi [13] or have analyzed power structures [14] and the vulnerability of adolescent girls [15].

Qualitative research gives us an in-depth understanding of local situations and may help identify factors that quantitative analyses have not considered. Qualitative analyses often focus on individual knowledge, opinions, attitudes, and challenges, while quantitative analyses quantify relationships between various factors and between factors and outcome variables. Combining qualitative and quantitative studies may reveal how and why various factors interact across time and space in a complex and widespread epidemic.

We used the topic of HIV in Malawi as a case study for an in-depth literature review of quantitative and qualitative literature on social and behavioral factors that may influence the HIV epidemic. We designed the review to be broad and inclusive to capture all possible relevant factors and expected it to identify known factors, neglected factors, and some factors that may never have been identified or analyzed in quantitative studies. We chose Malawi for several reasons: (1) Malawi has a relatively high HIV prevalence that varies substantially between regions [16], (2) the country is socioculturally diverse because it is home to many ethnic groups (eg, Chewa, Nyanja, Tumbuka, Yao, Lomwe, Sena, Tonga, Ngoni, Ngonde, Asians, and Europeans), and (3) a preliminary search revealed that Malawi was the focus of many scientific studies on HIV, including some by our group [17-19].

Because we knew that the breadth of our topic would return too many studies to read, we developed a semi-automated literature search engine and software that automatically downloads and analyzes open-access, full-text articles; this software can be used for searches on any broad topic across any region.

Methods

Search Strategy

We searched all English language articles published from the inception of the databases from January 1, 1987 to October 1, 2019 using the query "HIV AND Malawi". We performed an automated search of PubMed, PubMed Central (PMC), Paperity, and arXiV using a custom Python script and the corresponding application programming interfaces (APIs). We also sent a request for the same type of data to JSTOR since it offered no API to directly access the database.

Inclusion and Exclusion Criteria

We used a broad query designed to capture all articles about HIV and Malawi, not just those focused on health, and considered all studies that discussed social, behavioral, and cultural factors that might be associated with HIV infection in Malawi. The selection process is described in more detail in the following sections. We included original peer-reviewed articles, both quantitative and qualitative studies, and preprints. We also analyzed systematic reviews but preferred to include original publications when possible. We discarded articles that investigated the effects of HIV/AIDS (eg, [20]).

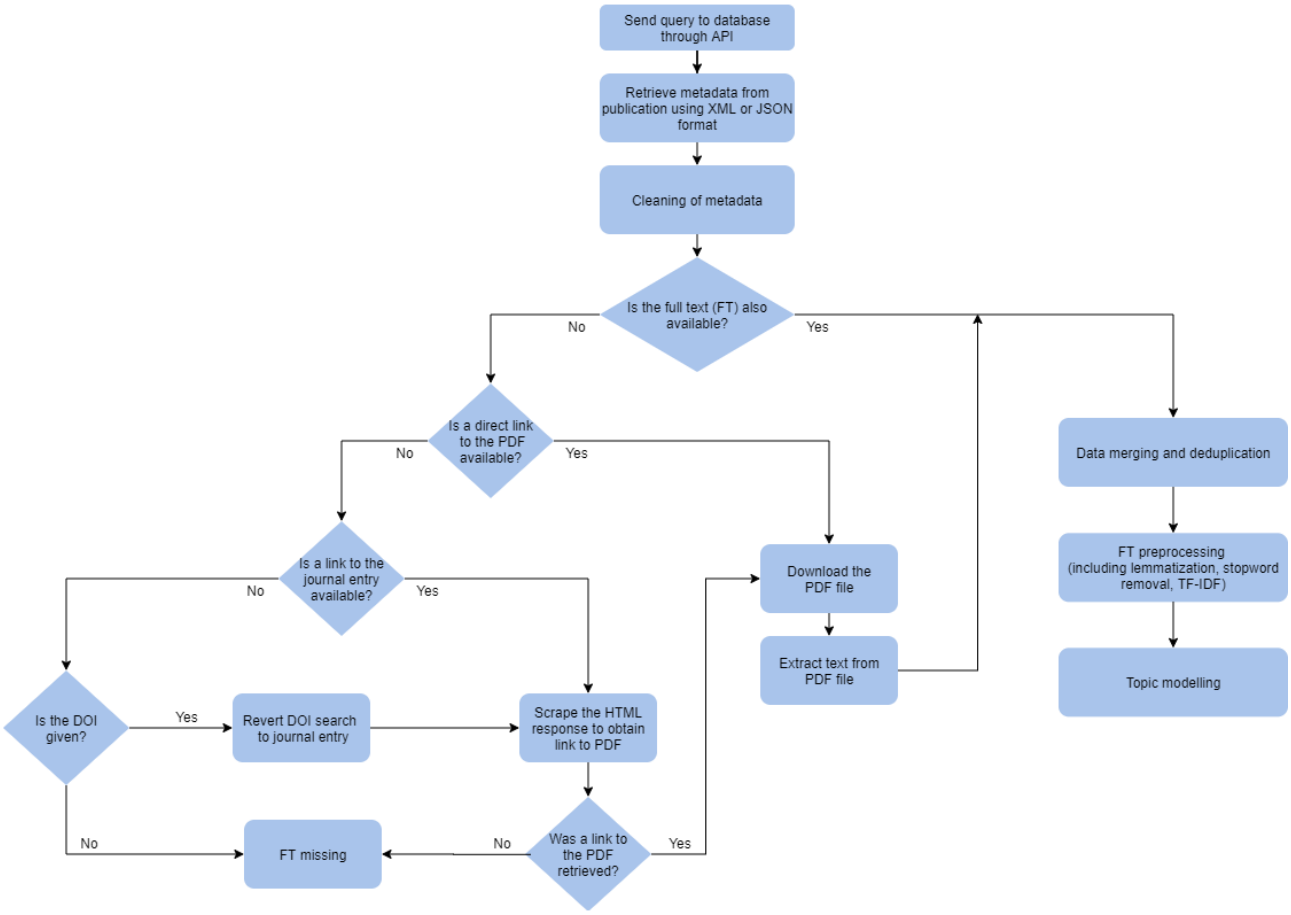
Data Collection

Figure 1 shows the flow of data collection and procession. We extracted the following information from articles: list of authors and their affiliations and MeSH keywords, Digital Object Identifier (DOI), title, abstract, full text, publication date, journal provider, and URL of the PDF version of the full-text article. Mandatory fields were title and availability of the full text; other items were retrieved if possible. If full-text articles could not be directly retrieved from the database, they were extracted from either an automatically downloaded PDF when the URL of the PDF was available or retrieved through an automatic reversed search with the DOI resolver [21]. The DOI resolver identified PDFs by scraping the internet. In some cases, the information in databases was incomplete (eg, missing DOI), or we could not access the PDF (eg, access restrictions, PDF contained only images, unavailable PDF) so we could not obtain the full text. We then checked the data for duplicates and merged them.

Each extracted full-text article was tokenized and processed further. We deleted non-ASCII characters, numbers, words under 4 characters (except some acronyms of interest like "HIV" and the abbreviation for antiretroviral therapy, "ART"), and a list of stopwords from the NLTK Python toolkit [22] with relevant additions (eg, URL, "volume," "journal"). We then lemmatized the text to avoid duplicating words with different inflectional endings.

Once cleaned, we ran term frequency-inverse document frequency from the scikit-learn Python package to extract relevant keywords [23]. Because the formatting of author affiliations was so heterogeneous, we retained only the city. We then stored original data and the generated data in a local SQLite database.

Figure 1. Flow for gathering and preprocessing data. API: application programming interface; DOI: Digital Object Identifier; TF-IDF: term frequency-inverse document frequency.



Topic Modelling

Finally, we used topic modelling to classify documents into topics based on their similarity. The process allowed us to broaden our search terms to identify relevant publications and to extract keywords relevant to a topic, providing an overview of salient and relevant terms that may best represent the data. We scored each document for a set number of topics, using the latent Dirichlet allocation (LDA) method [24]. Based on their highest topic score, publications were allocated to a topic.

To optimize computational efficiency and quickly identify potentially relevant articles, we initially chose 5 topics; for each, we used the same approach to identify 5 subtopics. We repeated the process 4 times and identified 625 topics. The resulting "tree of topics" increased its specificity at each repetition, which helped us identify our topics of interest faster than if we had started with many topics. We manually selected the number of iterations, but our selection was analysis-driven. Although more iterations would have reduced the number of potentially relevant articles that we needed to check manually, it would have raised the risk of missing relevant articles, and the performance of the algorithm would have decreased.

The code used for this paper was frozen and is publicly available [25]. Note that this code is not up-to-date as we are improving the software, adding databases, and trying different methods of classification.

Results

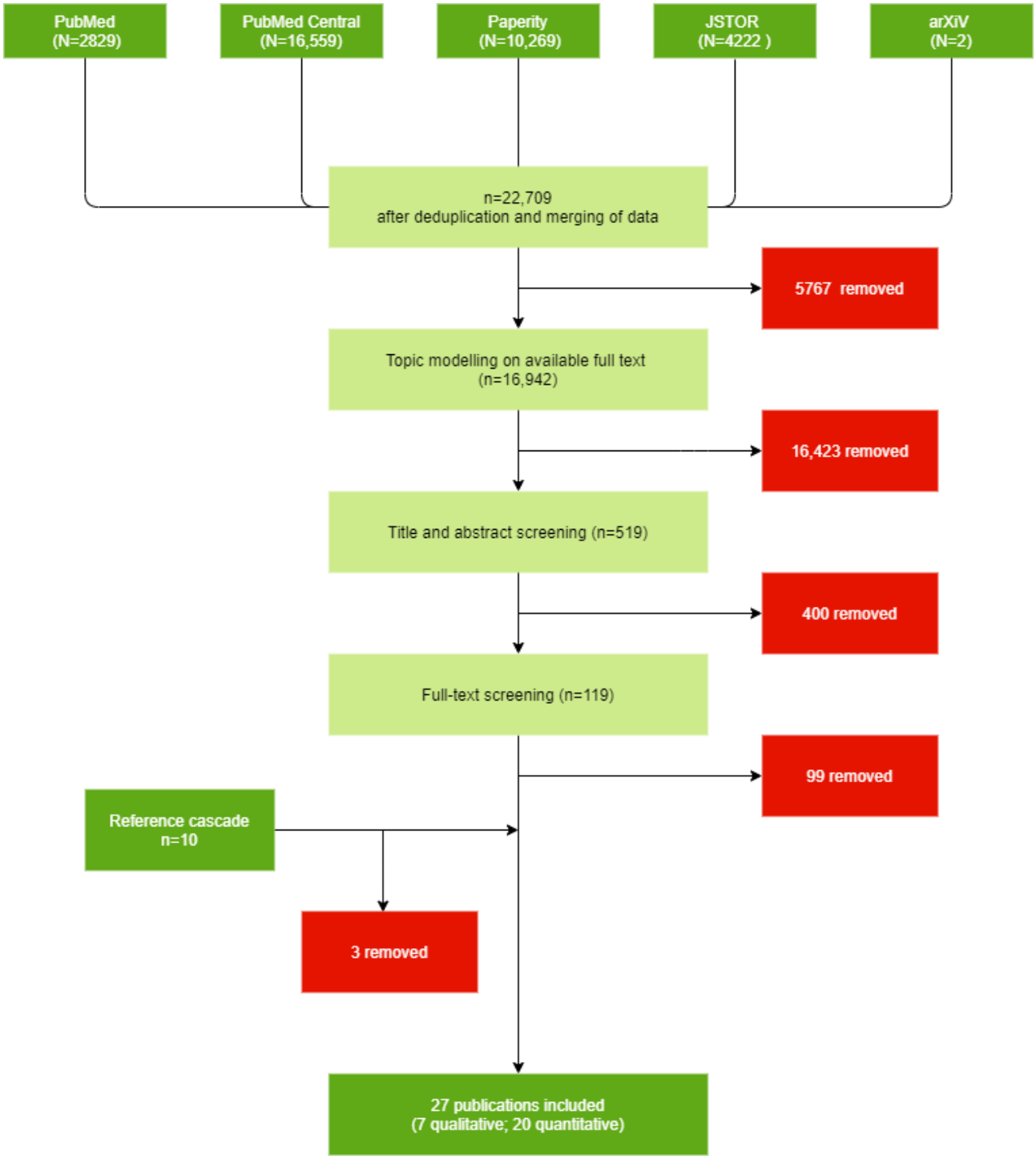
The PRISMA diagram (Figure 2) summarizes the selection process for relevant articles. Based on our search term, the software extracted a total of 22,709 unique articles, of which 16,942 full-text articles (74.6%) were retrieved directly from the database or extracted from PDFs. Topic modelling automatically screened these articles, reducing our selection to a subset of 519 relevant articles that included 14 topics related to behaviors, beliefs, culture, and religion. Of these, we manually selected 119 based on titles and abstracts and after applying our inclusion and exclusion criteria. After full-text screening, 20 articles were included in the systematic review.

We identified 10 more articles from references and included 7 of them.

The systematic review finally included 27 publications: 20 (74.1%) were quantitative, and 7 (25.9%) were qualitative. It took about 5 days to retrieve and preprocess the data for topic modelling. The iterative topic modelling took about a day.

In the following sections, we summarize the main findings of the included articles by topic area.

Figure 2. Study selection process.



Religion

We identified 3 publications that found religion was associated with the risk of HIV transmission or decreased risk behavior. All noted the effect of the religious leader on family planning, sexual morality, positive attitude towards condom use [26], and reducing stigma [27,28]. The studies also showed that congregations in rural Malawi responded to the HIV epidemic by taking care of orphans, helping the sick, or sponsoring HIV-related knowledge programs [27,28]. In this way, religion shaped the HIV experience of congregation members. Yeatman and Trinitapoli [26] noticed an association between religious socialization and contraceptive use. The literature search also highlighted that Muslim men were more likely to be circumcised

and less likely to be HIV-infected than men of other faiths [24], though they were also less likely to use condoms. Being religious did not necessarily help prevent HIV, although Lau et al [29] noticed a positive association between weekly attendance of religious services and women's use of modern contraceptives. Faith was often associated with condom rejection. Anglewicz and Clark [30] found that Muslim men tended to find condom use less acceptable than did those of other faiths, while a change to Christianity increased condom use in men.

Partnerships and Gender

Reniers and Tfaily [31] investigated the role of polygyny in Rumphi, Mchinji, and Balaka. They reported more extramarital sex in polygynous marriages than in monogamous marriages. They further found evidence that in polygynous marriages, the latest wives were more likely HIV-positive than the first wives. Other studies also showed that having more than one sexual partner increased HIV prevalence, led to differences in adherence counselling and testing, or led to differences in male circumcision practices [29,32,33]. Stephenson [34] found a discrepancy between different types of relationships: Cohabiting men were less likely to have risky extramarital sex than married men.

Age asymmetry in relationships plays a role in HIV transmission [35]. One article examined the effect of an age difference between women and their partners on HIV transmission on Likoma Island. When male partners were 2-12 years older than their female partners, the women's risk of being HIV-positive was higher than for women whose partners were more than 12 years older. When female partners were more than 5 years older than their male partner, they were more likely to be HIV-positive than women whose partners were less than 5 years older. Never using condoms and being married were associated with larger age differences between men and women.

Condom use helps prevent HIV infection. Anglewicz and Clark [30] showed that marital status and women's and men's risk perceptions were associated with condom use. Getting married reduced the acceptability of using condoms during sexual intercourse. A woman's perception of her HIV status was generally more important than her real HIV status for the acceptability of condom use within marriage. Known HIV status was a more important determinant of condom use in men than in women [30].

Power structures and closeness in partnerships may strongly influence HIV transmission. Becker et al [32] showed that in women, prior HIV testing and emotional closeness to a partner were associated with acceptance of home-based HIV testing and counselling services. Schatz [36] showed that married women may be able to protect themselves from infection by communicating with their husband about HIV, by confronting their husband's sexual partners, or by refusing polygyny. Women may also seek support and advice from their friends and relatives or even ask for a divorce. This suggests that women are not always vulnerable; some can self-advocate for their protection in marriage to reduce their risk of acquiring HIV, especially in the matrilocal southern part of Malawi where a woman can tell her husband to "take your mat and go" [36]. Even in the patrilocal northern region, Schatz [36] highlighted the support of women by their kin if the husband's risky behaviors cause women to return to their family.

Beliefs

We found several articles that studied beliefs and found associations between misconceptions and ignorance about HIV transmission factors and risky behavior. Authors labelled these "false beliefs" and found they stigmatized people with HIV [26,37,38]. One study mentioned beliefs about women's cleansing rituals. Some congregations believe women will be cleansed by having unprotected sex (eg, after the death of their husbands, after giving birth, or after miscarriage) [39]. The belief that HIV-infected men can be cured by having sex with a virgin woman also spreads the disease.

Personal beliefs play an important role in HIV prevention and risk of HIV infection. Three studies highlighted an important role of perceived HIV status between partners. Anglewicz and Clark [40] examined the accuracy of perceived HIV status in 768 monogamous couples, finding that partners tended to overestimate the risk of being HIV-positive; overestimation was associated with marital infidelity. But knowing one's or one's spouse's actual HIV status significantly reduced HIV risk. Fedor et al [41] showed that once HIV-negative women and HIV-positive men learned their status, they reduced risky behavior by increasing condom use and having sex with fewer partners.

Social and Behavioral Characteristics

We identified 3 studies on the effect of behavior-change interventions on HIV [42-44]. In 2 studies [42,44], the authors found that the intervention seemed to affect HIV risk behaviors and knowledge; the third study found no effect [43]. Crittenden et al [42] studied the spread of behavioral and psychological factors with peer group interventions in adults living in rural areas of central Malawi. The behavioral changes that were promoted in the intervention group (eg, partner communication, use of condoms, recent HIV test) spread to other persons in the same community. The second study [44] assessed the effect of a cash transfer program (lottery ranging from US $1 to US $5) in adolescents and women aged 13-22 years who attended school. The primary outcome was HIV prevalence 18 months after study enrollment; it was 1.2% in the intervention group and 3.0% in the control group. Women who received the cash transfer were less sexually active than women in the control group. In contrast, a study in northern Malawi [43] showed that behavior change interventions did not reduce the risk of HIV infection in Malawian adolescents, possibly because these interventions send contradictory messages or because adolescents are more influenced by their living environment (culture, religion, peers).

Several studies investigated the association between migration and HIV. Helleringer et al [45] reported an association between concurrent partnerships and HIV serodiscordance among couples on Likoma Island. HIV positivity was associated with migration out of the country (circular out-migration) and sexual contact with temporary in-migrants to the island. Anglewicz and Clark [46] concluded that migrants were more at risk for HIV

infection, but migration was not the reason for the higher risk; people with HIV were more likely to migrate, thereby reversing the causality.

Low socioeconomic status often drives HIV in Malawi [39,44]. It was associated with early sexual relationships, transactional sex, and a higher probability of having sex with older men [39]. In some studies, a low socioeconomic status was also associated with coercive heterosexual relationships. Coercive sex was a strong predictor for HIV infection in male victims [47]; the likelihood of being HIV positive was 7.2 times higher among men who had been sexually coerced than among those who had not been. One publication studied the association of coercive sexual behavior with social and economic status [48] and found that unemployment was strongly associated with coercive sex in young men in Blantyre, whereas material deprivation only was strongly associated with coercive sex in young women.

HIV infection was associated with intravaginal practices and products applied by women to manage their sexual relationship, manage menstruation, and improve their health. Women used cloth or paper to wipe out their vagina, inserted products to dry or tighten the vagina, and used intravaginal cleaning soap [39,49]. The use of the injectable hormonal birth control drug medroxyprogesterone acetate was associated with HIV seroconversion in HIV-negative women during a clinical trial [33]. Among men who have sex with men, the use of water-based lubricants could lower HIV risk [50]. Being older than 25 years, not being married, and age at first sexual intercourse were associated with HIV infection [50]. Some of these variables were also identified in other studies [29,32,47].

Although studies from other sub-Saharan African countries showed an association between alcohol and drug consumption with HIV testing, HIV infection, or uptake of preventive methods, evidence was limited in our study. Lau et al [29] found no association between tobacco use and male circumcision. Conroy and Chilungo [47] found some association between alcohol use and being HIV positive, but the association was not statistically significant (odds ratio 1.56). They did find a significant association between sexual coercion of women and alcohol use among men.

Discussion

Strengths

Using a broadly inclusive search phrase and repeated topic modelling, we quickly identified a small number of highly relevant articles about sociobehavioral factors and HIV in Malawi among 16,942 open access articles from 5 different databases. Our Python tool quickly reduced the number of potentially relevant articles to 519 in a few hours. It took us 5 days to screen titles and abstracts of these 519 articles, identify 119 potential full-text articles, and include 20 remaining articles in the review. We then added 7 more articles from references.

Our software allowed us to omit the time-consuming step of devising and tuning a specific search query combining logic keywords (« AND », « OR », and « NOT ») and modifying the search to suit the different requirements of each database. Traditional systematic searches require prior knowledge of the topic of interest (the deeper the better). While deep knowledge and tailored search strings offer benefits, they also risk missing relevant articles on topics one did not think to include and may limit the possibility of discoveries. Our software also allowed more exhaustive searching since it relied on full-text articles instead of only abstracts.

Limitations

This version of the software is limited to databases that provide free APIs for open access to full-text articles. Databases often used for systematic reviews (eg, Web of Science or Scopus), databases commonly used by social scientists (eg, SocIndex, CINAHL, ATLA Religion, ProQuest services), and preprint servers for accessing non-peer-reviewed literature (eg, medRxiV, bioRxiV) do not provide suitable APIs. For example, Web of Science has a basic API but does not provide access to full-text articles. Some databases like Scopus or ProQuest provide a free basic API from which we could retrieve basic metadata but accessing more detailed information and full-text data require a subscription. Medical preprint servers like medRxiv provide basic RSS feeds to obtain some metadata. Thus, our software is best for searching peer-reviewed, open-access medical literature, especially since most publications behind a paywall forbid text mining.

The LDA algorithm, like machine learning methods, has inherent limitations. The user must choose the number of topics; 10 is the default, but the ideal number of topics differs from corpus to corpus [24]. Our stepwise topic modelling approach mitigated this limitation. The LDA algorithm also performs badly on small corpora [24], so as our corpus was reduced, risk of incoherent topics increased. Therefore, we halted our iterations when the corpus shrank too much and analyzed the parent corpus instead.

Our software only works on PDFs, but not yet on image-based PDFs. It also does not parse HTML-only publications. We are working on integrating optical character recognition to translate images to texts. This limitation was reduced by our finding that image-based PDFs were usually less-relevant, older articles.

Our broad search strategy (“HIV AND Malawi” anywhere in the text) retrieved many irrelevant articles that mentioned Malawi only in the references, so we intend to add the option to exclude references from the search, We also plan to allow searches to be performed in specific parts of the publications, similar to a traditional systematic review (title, abstract, full text). However, some databases do not allow this, and each provider has a different approach to those searches.

Because the topics identified by the software sometimes overlapped, each article might fall under several topics, and each article was attributed only to the highest scoring topic, while ambiguous or multiple attributions were ignored. Consequently, a relevant publication might have been classified in a topic of no interest and thus was not incorporated in the final result. We checked to see if the 7 additional publications identified from references were collected by our tool but not included and, therefore, misclassified. Only one such publication was identified [36]; the other 6 were either not extracted or were

retrieved but not used for classification as the full text was missing.

Our software is intended to complement rather than replace systematic reviews. We have not yet compared our approach to a systematic review, and we expect we missed relevant articles unavailable through open access. We also expect that our software missed some topics, while also finding new topics. As the number of open access articles and preprints increases [51], and as journals and preprint sites add APIs, we expect our software to become more useful. We plan to add more databases as we gain access.

We may have missed factors that influence HIV transmission in Malawi, and it could be that broadening the search to "Malawi" would overcome this limitation and reveal possible interactions between social, political, economic, and other factors that influence the course of the HIV epidemic but have never been studied in the context of HIV.

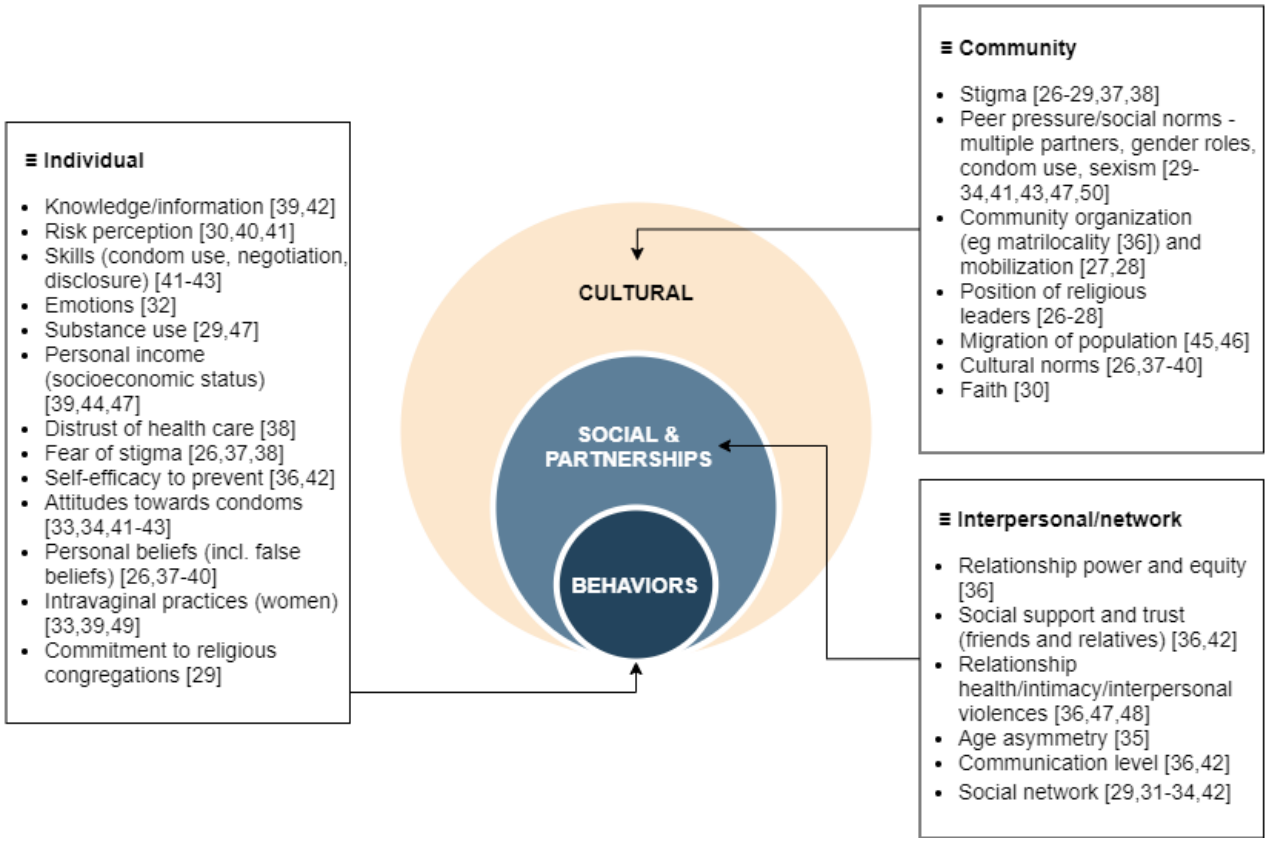
HIV-Specific Discussion

Figure 3 shows the identified factors in a 3-level diagram, modeled after that of Kaufman et al [52]. The 27 identified articles include 24 behavioral, social, and cultural factors of HIV infection. Compared to previous literature reviews on HIV-related factors observed worldwide [52-55], a few behavioral-related factors were not retrieved in our study. These include, for example, denial of HIV status, motivation and intention to change sexual behavior, reactions to stress, physical and mental health status, outcome expectancies (ie, anticipated consequences as a result of engaging in a specific behavior), and empowerment. For social factors, relationship satisfaction and level of relationship commitment were also missing. Finally, we found no article discussing racism.

The 2015-2020 National HIV Prevention Strategy plan from the National AIDS Commission of Malawi targets specific HIV-related interventions at multiple levels for different population groups [56]. We compared the many behavioral, social, and cultural factors of HIV infection mentioned by the Prevention Strategy plan with the factors identified by our software. For the key populations of men having sex with men and sex workers, reducing the number of partners, consistent use of condoms, targeted campaigns on HIV testing and risk reduction, alcohol and substance abuse, positive health, and gender-based violence prevention programs are all behavioral interventions targeting risk factors that are present in our systematic review. Regarding the priority populations, additional interventions such as comprehensive sexuality and messages on intergenerational sex for young women at risk, stop early marriage campaigns, female support for voluntary medical male circumcision, HIV testing and counselling, and communication for couples are also addressing part of our behavioral list of HIV risk factors.

Behavioral factors, at an individual or a community level, that are not targeted by the Prevention Strategy plan for 2015-2020 are intravaginal practices. For the general population, distrust of health care, commitment to religious congregations, and the position of religious leaders are not addressed either. These are factors that should be considered when elaborating the new Malawi Prevention Strategy Plan for 2021-2026.

Figure 3. Summary of factors found within publications identified by the review.



Conclusions

From a set of articles limited by the existence of journal paywalls, our Python software quickly narrowed a set of over 16,000 articles to a small set of relevant articles. We identified sociobehavioral factors, including factors related to society and culture, such as folk beliefs, theology, and moral standards, that may influence the course of the HIV epidemic, yet are rarely considered in the quantitative literature. Extending our approach to other countries could give researchers a more complete picture of the different drivers of the HIV epidemic in different settings and clarify the reasons for the spatial variability of HIV across sub-Saharan Africa.

Acknowledgments

The authors would like to thank Aziza Merzouki for fruitful discussions. This work was supported by the Swiss National Foundation (SNF) under grant #163878.

Authors' Contributions

AT wrote the software and obtained and analyzed the results it produced. AT further read the remaining articles after software screening and wrote the first draft of this article. IT helped read the remaining articles after software cleaning and identify relevant risk factors. IT also helped write the results section of this article. EO wrote the HIV-specific subsection of the discussion. KT provided English edits and helped identify and correct unclear sentences in the article. OK reviewed the paper and helped write the introduction.

Conflicts of Interest

None declared.

References

1. HIV/AIDs in Namibia: Behavioural and Contextual Factors Driving the Epidemic. Government of the Republic of Namibia. Ministry of Health and Social Services. 2009. URL: http://www.infocenter.nercha.org.sz/sites/default/files/NamibiaHIVDrivers.pdf [accessed 2020-03-23]
2. Zulu LC, Kalipeni E, Johannes E. Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994-2010. BMC Infect Dis 2014 May 23;14:285 [FREE Full text] [doi: 10.1186/1471-2334-14-285] [Medline: 24886573]
3. Ngesa O, Mwambi H, Achia T. Bayesian spatial semi-parametric modeling of HIV variation in Kenya. PLoS One 2014;9(7):e103299 [FREE Full text] [doi: 10.1371/journal.pone.0103299] [Medline: 25061669]
4. Poulin M, Muula AS. An inquiry into the uneven distribution of women's HIV infection in rural Malawi. Demogr Res 2011;25:869-902 [FREE Full text] [doi: 10.4054/DemRes.2011.25.28] [Medline: 28936119]
5. Zulu LC, Kalipeni E, Johannes E. Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994-2010. BMC Infect Dis 2014 May 23;14:285 [FREE Full text] [doi: 10.1186/1471-2334-14-285] [Medline: 24886573]
6. Manda S, Masenyetse L, Cai B, Meyer R. Mapping HIV prevalence using population and antenatal sentinel-based HIV surveys: a multi-stage approach. Popul Health Metr 2015;13:22 [FREE Full text] [doi: 10.1186/s12963-015-0055-z] [Medline: 26336361]
7. Feldacker C, Emch M, Ennett S. The who and where of HIV in rural Malawi: Exploring the effects of person and place on individual HIV status. Health Place 2010 Sep;16(5):996-1006 [FREE Full text] [doi: 10.1016/j.healthplace.2010.06.004] [Medline: 20598623]
8. Tomita A, Vandormael AM, Bärnighausen T, de Oliveira T, Tanser F. Social Disequilibrium and the Risk of HIV Acquisition: A Multilevel Study in Rural KwaZulu-Natal Province, South Africa. J Acquir Immune Defic Syndr 2017 Jun 01;75(2):164-174 [FREE Full text] [doi: 10.1097/QAI.0000000000001349] [Medline: 28291049]
9. Baranczuk Z, Estill J, Blough S, Meier S, Merzouki A, Maathuis MH, et al. Socio-behavioural characteristics and HIV: findings from a graphical modelling analysis of 29 sub-Saharan African countries. J Int AIDS Soc 2019 Dec;22(12):e25437 [FREE Full text] [doi: 10.1002/jia2.25437] [Medline: 31854506]
10. Zamawe C. Factors that Affect Maternal Care Seeking Behaviour and the Choice of Practitioner(s) during Complications: the Case of Mang'anja Tribe in Malawi. Research on Humanities and Social Sciences 2013;3(18):18-26 [FREE Full text]
11. Ashforth A. When the Vampires Come for You: A True Story of Ordinary Horror. Social Research 2014;81:851-882 [FREE Full text]
12. Babalola S. Women's education level, antenatal visits and the quality of skilled antenatal care: a study of three African countries. J Health Care Poor Underserved 2014 Feb;25(1):161-179. [doi: 10.1353/hpu.2014.0049] [Medline: 24509018]
13. Conroy AA. 'It means there is doubt in the house': perceptions and experiences of HIV testing in rural Malawi. Cult Health Sex 2014;16(4):397-411 [FREE Full text] [doi: 10.1080/13691058.2014.883645] [Medline: 24580127]

14. Jacques G. “Coming Out” or Coming in? Social Exclusion of Sexual Minorities in Africa: Challenges for Social Work Education and Practice. Journal of Gay & Lesbian Social Services 2014 Feb 24;26(1):91-110. [doi: 10.1080/10538720.2013.829396]
15. Mkandawire P, Luginaah I, Baxter J. Growing up an orphan: vulnerability of adolescent girls to HIV in Malawi. Trans Inst Br Geogr 2013 Mar 08;39(1):128-139. [doi: 10.1111/tran.12002]
16. Palk L, Blower S. Geographic variation in sexual behavior can explain geospatial heterogeneity in the severity of the HIV epidemic in Malawi. BMC Med 2018 Feb 09;16:22 [FREE Full text] [doi: 10.1186/s12916-018-1006-x] [Medline: 29422096]
17. Phiri N, Haas AD, Msukwa MT, Tenthani L, Keiser O, Tal K. "I found that I was well and strong": Women's motivations for remaining on ART under Option B+ in Malawi. PLoS One 2018;13(6):e0197854 [FREE Full text] [doi: 10.1371/journal.pone.0197854] [Medline: 29874247]
18. Merzouki A, Styles A, Estill J, Baranczuk Z, Petrie K, Keiser O. Identifying groups of people with similar socio-behavioural characteristics in Malawi to inform HIV interventions: a Latent Class Analysis. medRxiv 2019 Dec 27 (forthcoming)(forthcoming) [FREE Full text] [doi: 10.1101/2019.12.26.19015560]
19. Tenthani L, Haas AD, Tweya H, Jahn A, van OJJ, Chimbwandira F, et al. Retention in care under universal antiretroviral therapy for HIV-infected pregnant and breastfeeding women ('Option B+') in Malawi. AIDS 2014 Feb 20;28(4):589-598 [FREE Full text] [doi: 10.1097/QAD.0000000000000143] [Medline: 24468999]
20. Hellandendu J. Contributory factors to the spread of HIV/AIDS and its impacts in sub-Saharan African countries. European Scientific Journal 2012 Jun 29;8(14):144-156 [FREE Full text] [doi: 10.19044/esj.2012.v8n14p%25p]
21. DOI solver. URL: https://dx.doi.org [accessed 2020-07-24]
22. Natural Language Toolkit. NLTK 3.5 documentation. 2020 Apr 13. URL: https://www.nltk.org [accessed 2020-07-16]
23. Machine Learning in Python. scikit-learn. URL: https://scikit-learn.org/stable/ [accessed 2020-07-16]
24. Blei D, Ng A, Jordan, MI. Latent Dirichlet Allocation. Journal of Machine Learning Research 2003;3:993-1022. [doi: 10.1162/jmlr.2003.3.4-5.993]
25. A freeze of the ASR as it was used for the paper "Social, behavioural, and cultural factors of HIV in Malawi: a semi-automated systematic review" submitted to JMIR. ASR_freeze_HIVMalawiJMIR. URL: https://gitlab.com/Pezam/asr_freeze_hivmalawijmir [accessed 2020-07-16]
26. Yeatman SE, Trinitapoli J. Beyond Denomination: The Relationship between Religion and Family Planning in Rural Malawi. Demogr Res 2008 Oct 24;19(55):1851-1882 [FREE Full text] [doi: 10.4054/DemRes.2008.19.55] [Medline: 20463916]
27. Trinitapoli J. Religious Responses to AIDS in Sub-Saharan Africa: An Examination of Religious Congregations in Rural Malawi. Review of Religious Research 2006;47(3):253.
28. Adams J, Trinitapoli J. The Malawi Religion Project: Data collection and selected analyses. Demogr Res 2009 Sep 01;21(4):255-288 [FREE Full text] [doi: 10.4054/DemRes.2009.21.10] [Medline: 20072664]
29. Lau FK, Jayakumar S, Sgaier SK. Understanding the socio-economic and sexual behavioural correlates of male circumcision across eleven voluntary medical male circumcision priority countries in southeastern Africa. BMC Public Health 2015 Aug 22;15:813 [FREE Full text] [doi: 10.1186/s12889-015-2135-1] [Medline: 26297202]
30. Anglewicz P, Clark S. The effect of marriage and HIV risks on condom use acceptability in rural Malawi. Soc Sci Med 2013 Nov;97:29-40 [FREE Full text] [doi: 10.1016/j.socscimed.2013.06.024] [Medline: 24161086]
31. Reniers G, Tfaily R. Polygyny and HIV in Malawi. Demogr Res 2008 Oct 10;19(53):1811-1830 [FREE Full text] [doi: 10.4054/DemRes.2008.19.53] [Medline: 21949484]
32. Becker S, Taulo FO, Hindin MJ, Chipeta EK, Loll D, Tsui A. Pilot study of home-based delivery of HIV testing and counseling and contraceptive services to couples in Malawi. BMC Public Health 2014 Dec 20;14(1):1309 [FREE Full text] [doi: 10.1186/1471-2458-14-1309] [Medline: 25526799]
33. Kumwenda J, Makanani B, Taulo F, Nkhoma C, Kafulafula G, Li Q, et al. Natural History and Risk Factors Associated with Early and Established HIV Type 1 Infection among Reproductive‐Age Women in Malawi. Clin Infect Dis 2008 Jun 15;46(12):1913-1920. [doi: 10.1086/588478]
34. Stephenson R. Community-level gender equity and extramarital sexual risk-taking among married men in eight African countries. Int Perspect Sex Reprod Health 2010 Dec;36(4):178-188 [FREE Full text] [doi: 10.1363/3617810] [Medline: 21245024]
35. Beauclair R, Helleringer S, Hens N, Delva W. Age differences between sexual partners, behavioural and demographic correlates, and HIV infection on Likoma Island, Malawi. Sci Rep 2016 Nov 02;6:36121 [FREE Full text] [doi: 10.1038/srep36121] [Medline: 27805053]
36. Schatz E. 'Take your mat and go!': rural Malawian women's strategies in the HIV/AIDS era. Cult Health Sex 2005;7(5):479-492. [doi: 10.1080/13691050500151255] [Medline: 16864217]
37. Regnerus MD, Salinas V. Religious Affiliation and Aids-Based Discrimination in Sub-Saharan Africa. Review of Religious Research 2007;48(4):385-400 [FREE Full text]
38. Westercamp N, Bailey RC. Acceptability of male circumcision for prevention of HIV/AIDS in sub-Saharan Africa: a review. AIDS Behav 2007 May;11(3):341-355 [FREE Full text] [doi: 10.1007/s10461-006-9169-4] [Medline: 17053855]

39. Ramjee G, Daniels B. Women and HIV in Sub-Saharan Africa. AIDS Res Ther 2013 Dec 13;10(1):30 [FREE Full text] [doi: 10.1186/1742-6405-10-30] [Medline: 24330537]
40. Anglewicz PA, Bignami-Van Assche S, Clark S, Mkandawire J. HIV risk among currently married couples in rural Malawi: what do spouses know about each other? AIDS Behav 2010 Feb;14(1):103-112 [FREE Full text] [doi: 10.1007/s10461-008-9497-7] [Medline: 19051003]
41. Fedor TM, Kohler H, Behrman JR. The Impact of Married Individuals Learning HIV Status in Malawi: Divorce, Number of Sexual Partners, and Condom Use With Spouses. Demography 2015 Feb;52(1):259-280 [FREE Full text] [doi: 10.1007/s13524-014-0364-z] [Medline: 25582891]
42. Crittenden KS, Kaponda CPN, Jere DL, McCreary LL, Norr KF. Participation and diffusion effects of a peer-intervention for HIV prevention among adults in rural Malawi. Soc Sci Med 2015 May;133:136-144 [FREE Full text] [doi: 10.1016/j.socscimed.2015.03.055] [Medline: 25864150]
43. Mwale M, Muula AS. Effects of adolescent exposure to behaviour change interventions on their HIV risk reduction in Northern Malawi: a situation analysis. SAHARA J 2018 Dec;15(1):146-154 [FREE Full text] [doi: 10.1080/17290376.2018.1529612] [Medline: 30278823]
44. Baird SJ, Garfein RS, McIntosh CT, Ozler B. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. Lancet 2012 Apr 07;379(9823):1320-1329. [doi: 10.1016/S0140-6736(11)61709-1] [Medline: 22341825]
45. Helleringer S, Mkandawire J, Kalilani-Phiri L, Kohler H. Cohort Profile: The Likoma Network Study (LNS). Int J Epidemiol 2014 Apr;43(2):545-557 [FREE Full text] [doi: 10.1093/ije/dyt001] [Medline: 23543589]
46. Anglewicz P. Migration, marital change, and HIV infection in Malawi. Demography 2012 Feb;49(1):239-265 [FREE Full text] [doi: 10.1007/s13524-011-0072-x] [Medline: 22109083]
47. Conroy AA, Chilungo A. Male victims of sexual violence in rural Malawi: the overlooked association with HIV infection. AIDS Care 2014;26(12):1576-1580 [FREE Full text] [doi: 10.1080/09540121.2014.931562] [Medline: 24992179]
48. Kamndaya M, Kazembe LN, Vearey J, Kabiru CW, Thomas L. Material deprivation and unemployment affect coercive sex among young people in the urban slums of Blantyre, Malawi: A multi-level approach. Health Place 2015 May;33:90-100 [FREE Full text] [doi: 10.1016/j.healthplace.2015.03.001] [Medline: 25814337]
49. Low N, Chersich MF, Schmidlin K, Egger M, Francis SC, van de Wijgert JHHM, et al. Intravaginal practices, bacterial vaginosis, and HIV infection in women: individual participant data meta-analysis. PLoS Med 2011 Feb 15;8(2):e1000416 [FREE Full text] [doi: 10.1371/journal.pmed.1000416] [Medline: 21358808]
50. Wirtz AL, Jumbe V, Trapence G, Kamba D, Umar E, Ketende S, et al. HIV among men who have sex with men in Malawi: elucidating HIV prevalence and correlates of infection to inform HIV prevention. J Int AIDS Soc 2013;16 Suppl 3:18742 [FREE Full text] [Medline: 24321110]
51. Trends for open access to publications. European Commission. URL: https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/trends-open-access-publications_en [accessed 2020-03-04]
52. Kaufman MR, Cornish F, Zimmerman RS, Johnson BT. Health behavior change models for HIV prevention and AIDS care: practical recommendations for a multi-level approach. J Acquir Immune Defic Syndr 2014 Aug 15;66 Suppl 3:S250-S258 [FREE Full text] [doi: 10.1097/QAI.0000000000000236] [Medline: 25007194]
53. Coates TJ, Richter L, Caceres C. Behavioural strategies to reduce HIV transmission: how to make them work better. Lancet 2008 Aug 23;372(9639):669-684 [FREE Full text] [doi: 10.1016/S0140-6736(08)60886-7] [Medline: 18687459]
54. Faust L, Yaya S. The effect of HIV educational interventions on HIV-related knowledge, condom use, and HIV incidence in sub-Saharan Africa: a systematic review and meta-analysis. BMC Public Health 2018 Nov 13;18(1):1254 [FREE Full text] [doi: 10.1186/s12889-018-6178-y] [Medline: 30424761]
55. Bekker L, Beyrer C, Quinn TC. Behavioral and biomedical combination strategies for HIV prevention. Cold Spring Harb Perspect Med 2012 Aug 01;2(8):a007435 [FREE Full text] [doi: 10.1101/cshperspect.a007435] [Medline: 22908192]
56. National HIV Prevention Strategy 2015-2020. National AIDS Commission. 2014. URL: http://www.aidsmalawi.org.mw/viewhiv/Malawi%20HIV%20Prevention%20%20Strategy%202015%20-%202020_Final.pdf [accessed 2020-03-04]

Abbreviations

API: application programming interface
DOI: Digital Object Identifier
LDA: latent Dirichlet allocation
PMC: PubMed Central
TF-IDF: term frequency-inverse document frequency

Edited by G Eysenbach; submitted 23.03.20; peer-reviewed by C Nweneka, C Xu; comments to author 09.05.20; revised version received 20.05.20; accepted 04.06.20; published 14.08.20.

Please cite as:
Thiabaud A, Triulzi I, Orel E, Tal K, Keiser O
Social, Behavioral, and Cultural factors of HIV in Malawi: Semi-Automated Systematic Review
J Med Internet Res 2020;22(8):e18747
URL: http://www.jmir.org/2020/8/e18747/
doi:10.2196/18747
PMID:32795992

Review

The Academic Viewpoint on Patient Data Ownership in the Context of Big Data: Scoping Review

Martin Mirchev[1*], MPhil, LLM, MPH, PhD; Iskra Mircheva[1*], MIS, Prof Dr; Albena Kerekovska[1*], MD, Prof Dr

Department of Social Medicine and Healthcare Organization, Faculty of Public Health, Medical University of Varna, Varna, Bulgaria

[*]all authors contributed equally

Corresponding Author:
Iskra Mircheva, MIS, Prof Dr
Department of Social Medicine and Healthcare Organization
Faculty of Public Health
Medical University of Varna
55 Marin Drinov Str.
Varna,
Bulgaria
Phone: 359 887 618 180
Email: mircheva_is@abv.bg

Abstract

Background: The ownership of patient information in the context of big data is a relatively new problem, which is not yet fully recognized by the medical academic community. The problem is interdisciplinary, incorporating legal, ethical, medical, and aspects of information and communication technologies, requiring a sophisticated analysis. However, no previous scoping review has mapped existing studies on the subject.

Objective: This study aims to map and assess published studies on patient data ownership in the context of big data as viewed by the academic community.

Methods: A scoping review was conducted based on the 5-stage framework outlined by Arksey and O'Malley and further developed by Levac, Colquhoun, and O'Brien. The organization and reporting of results of the scoping review were conducted according to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses and its extensions for Scoping Reviews). A systematic and comprehensive search of 4 scientific information databases, PubMed, ScienceDirect, Scopus, and Springer, was performed for studies published between January 2000 and October 2019. Two authors independently assessed the eligibility of the studies and the extracted data.

Results: The review included 32 eligible articles authored by academicians that correspond to 3 focus areas: problem (ownership), area (health care), and context (big data). Five major aspects were studied: the scientific area of publications, aspects and academicians' perception of ownership in the context of big data, proposed solutions, and practical applications for data ownership issues in the context of big data. The aspects in which publications consider ownership of medical data are not clearly distinguished but can be summarized as ethical, legal, political, and managerial. The ownership of patient data is perceived primarily as a challenge fundamental to conducting medical research, including data sales and sharing, and to a lesser degree as a means of control, problem, threat, and opportunity also in view of medical research. Although numerous solutions falling into 3 categories, technology, law, and policy, were proposed, only 3 real applications were discussed.

Conclusions: The issue of ownership of patient information in the context of big data is poorly researched; it is not addressed consistently and in its integrity, and there is no consensus on policy decisions and the necessary legal regulations. Future research should investigate the issue of ownership as a core research question and not as a minor fragment among other topics. More research is needed to increase the body of knowledge regarding the development of adequate policies and relevant legal frameworks in compliance with ethical standards. The combined efforts of multidisciplinary academic teams are needed to overcome existing gaps in the perception of ownership, the aspects of ownership, and the possible solutions to patient data ownership issues in the reality of big data.

(J Med Internet Res 2020;22(8):e22214) doi:10.2196/22214

KEYWORDS

big data; ethics; legal aspects; ownership; patient-generated health data

Introduction

Background

At present, global health systems are undergoing a fundamental change. We have reached the stage of a major transition to ways in which we can improve the generation of and access unimaginable amounts of information. With the emergence and development of big data as a result of the information revolution, we can now manage and transform the approaches by which we control this information and, in health care, as a consequence, the ability to control and cure diseases. On the other hand, the emergence of big data in health care poses additional challenges, especially with regard to the privacy of individuals' personal data, security, ownership, management, and control.

Personal data, which some call the "21st Century New Oil" [1] or "the new currency," [2], are generated at an extremely high rate owing to the achievements of modern information and communication technologies (ICTs). In this sense, it is particularly important to address some of the problematic issues directly affecting medical information: Who owns patient information or who would have the fairest claim? Hospitals, researchers, or the patients themselves? Can the information be publicly owned or does it belong to health care providers? Where lies the interest of data carrier developers (software programs, servers, clouds, or social networks)? In general, if patient information is a property of the individuals, will this help improve their health in the reality of big data?

Given the relevance of these issues, and the ensuing additional issues, the lack of proper discussion is more than weird. Isolated debates exist in different places around the world, but interest in improving health care locally, as well as globally, requires concrete and decisive approaches. This suggests going beyond the purely theoretical framework. Considering the importance of this issue, there is a great need to advance our understanding of the relationship between patient information ownership and big data. More importantly, understanding how big data can be used not only to deliver adequate health care and promote traditionally neglected initiatives such as disease prevention and health promotion but also to improve medical research without depriving patients of their potential right to own their medical data.

Ownership in the Specific Context of Patient Information

In 2009, Hall and Schulman [3] asked, "Who owns medical information? The one who gives care, receives care, or pays for care? All of the above? None of the above? Does it really matter? In the emerging era of electronic health informatics, few other medicolegal questions are more critical, more contested, or more poorly understood."

Patient information ownership is an issue that requires justifying *self-ownership* [4,5], a modern term of what John Locke considers in 1680 as a *property of his own person* ("every man has a property in his own person: thus nobody has any right to but himself") [6], that is, patient information or personal health data belong to the characteristics that cannot be separated from the individual [7]. In the 1960s, Westin [8] proposed the idea that personal information should be formally recognized as an object of property rights.

Ownership is a complex, often transdisciplinary term that has eroded or has been fragmented depending on the observer's position and the purpose of the discussion [9]. It is often identified with *privacy*. However, the 2 terms are significantly different. The Oxford Dictionary of Law explains *privacy* as the qualified right to "be free from unwarranted intrusion and to keep certain matters from public view (or surveillance by the state), as recognized in Article 8 of the *European Convention on Human Rights and the *Human Rights Act 1998…." [10]. The term *ownership* is explained in the same dictionary as "the exclusive right to use, possess, and dispose of property, subject only to the rights of persons having a superior interest and to any restrictions on the owner's rights imposed by agreement with or by act of third parties, or by operation of law" [10]. Therefore, ownership implies certain legal rights over a property along with the explicit right to possess it, such as being able to control, enjoy, use, sell, rent, give away, make profit, or even destroy an item of property. This concept is clear from the perspective of corporeal ownership. It becomes much more complicated when considering incorporeal ownership, such as intellectual property or data and information [11].

Ownership is an important concept because it implies a level of control over the use of personal health data [12]. Other scholars have reviewed the interplay of property law and privacy law on health records and health data, with the bottom line being that neither property nor privacy law is completely applicable to health data or a patient's ability to control their health data [13].

The scarcity of academic discourse in this field is an interesting phenomenon, given the relevance of the topic [14]. What can be found as research on this subject usually comes from the field of jurisprudence, but there are authors who are known for their works in ethics and health care who also raise this issue [15-17].

Although we are currently witnessing a missing or at least limited academic debate on patient information ownership, the few academic authors who contribute to the debate seem to be grouped around 3 main points of view [14]. According to the first view, this information must be in the public domain [18]. The second view is that patients themselves must be the owners [19]. The third view is that property in itself is not a problem, and the issue can be regulated by other regulations, not specifically by property laws [20,21].

According to Rodwin [18], patient data should be privately owned by patients themselves as a means of protecting their privacy, but there are data that must necessarily be made public to ensure and protect the public interest, including key government initiatives that promote public health, individual patient safety, biomedical research, and not to forget economic interests.

Big Data in Health Care

Although the use of *huge amount of data* in health care is not a modern phenomenon [22-24], the term *big data* appeared in the 1990s and quickly became popular [25-27]. *Big* is a relative

term, especially when it comes to data, and big data usually include data sets that exceed the capabilities of commonly used software tools to store, manage, and process data within a reasonable period of time.

The volume of data is constantly growing, ranging from hundreds of terabytes until a few years ago to thousands of exabytes. Gantz and Reinsel [28] predicted an overall increase in health data by an average of 48% per year. Although there is no precise definition of big data [29], its attributes are well documented in the existing literature. These initially comprised the 3*V* model [30,31]—*Volume*, *Velocity*, and *Variety*—acknowledged by a number of authors [32,33]. Additional Vs were added by other authors to describe big data [34,35]—*Veracity*, *Value*, and *Variability* [36-38]. Currently, the term is defined by references from 3 to 15 attributes [39].

Although there are various common definitions of big data [40-42], none of them specifically focus on health, telemedicine, and health care. The European Commission developed the following definition [43] "Big Data in Health refers to large routinely or automatically collected datasets, which are electronically captured and stored. It is reusable in the sense of multipurpose data and comprises the fusion and connection of existing databases for the purpose of improving health and health system performance. It does not refer to data collected for a specific study." *Yet, not a word about personal patient data and ownership of patient data*.

Over the last few years, methods for aggregating and storing vast amounts of medical data have been improved through various applications [44-46]. Through analyses of big data, diseases can be detected earlier, with better chances and better outcomes for patients [47]. It is now widely believed that evidence-based medicine and, in the long run, personalized and precision medicine must constitute the gold standard of care [48-51]. Technological and informational conditions make ambiguities about the ownership of medical information even more interesting but also unclear. In this regard, the use of big data in health care has yet to prove its importance for clinical support, health insurance, monitoring of diseases and health determinants, treatment optimization, disease prevention and health promotion [52,53], health care improvement in general [54], and medical research, and it also poses serious ethical and regulatory challenges, ranging from risks to individual rights, privacy, and autonomy to transparency of initiatives and trust [55-57].

After the introduction of electronic health records (not to mention the variety of wearables), we are entering a new era of health monitoring. Regardless of its progress, it will be important to determine who will have access to what information, when and how, which ultimately raises the issue of ownership of patient information.

Considerations for This Study

In this study, we distanced ourselves from publications that refer to "data" as such, without specifying "patient data" explicitly; "property," argued mainly from the legal point of view as "intellectual property"; and "privacy" that is often identified as "ownership." It is *ownership* that is within the scope of our research.

Traditionally, *ownership* is debated from the legal standpoint [4,10,26,58-60], yet some authors say that clearly there is no specific data-related legislation that explicitly recognizes ownership in health data in the various EU Member States and in the United States [61-63]. Ultimately, a legal framework reflecting the rights of many stakeholders in the health information market is needed [64,65]. Another line of debate stems from the ethical concerns related to autonomy, privacy, confidentiality, and justice, but barely stressing on ownership as such [66,67].

In this paper, we did not focus on specific legal regulations or ethical considerations, as publications in these fields, to a large extent, do not explicitly address patient data in the context of today's information reality—big data. It is the perception of the authors that studying legal regulations or ethical considerations requires substantive and more specific research, which is not within the aim and scope of this study.

Our research showed that the peculiarities of *ownership* of patient information in the current data reality have not been the subject of research thus far, and even if they were related, they were not considered as the core issue but as part of studies aimed at something else.

Choice of Review Type

To map the available evidence in relation to the ownership of patient information in the context of big data, we conducted a scoping review, which by definition best suits the overall objectives of this study to clarify key concepts of an emerging scientific field and identify major sources, gaps, and innovative methods in the available evidence [68]. As we did not find any reviews on the question under investigation, the scoping review is the most appropriate type of review, because it can also be carried out as a stand-alone project, especially when the topic has not yet been extensively reviewed or is of a complex or heterogeneous nature [69-71].

The methodological framework for a scoping review was first outlined by Arksey and O'Malley [72] in 2005. It was further amended by Levac et al [73] in 2010 and by Peters et al [74] in 2015.

Although the terminology and methods for conducting a scoping review are still unclear and not well-described [75], they are constantly improving [76-78], and the use of scoping reviews keeps increasing. The average increase rate of indexed scoping reviews in PubMed for the last 5 years was 51.6% (compared with 16.9% for systematic reviews).

Our choice of scoping review was also determined by the fact that patient information ownership in the context of big data is a relatively new problem and apparently not yet fully recognized by the medical academic community as there are not many publications in this area. In addition, the problem is interdisciplinary, including legal, ethical, medical information, and communication technology aspects, which requires more complex searches and analyses of the available evidence.

Methods

Design

We conducted a scoping review, based on the 5-stage framework outlined by Arksey and O'Malley [72], and further developed by Levac et al [73]: (1) identifying the research questions, (2) identifying relevant studies, (3) study selection, (4) data items and data collection process, and (5) collating, summarizing, and reporting results. Specifically, Arksey and O'Malley's [72] optional sixth stage, *consultations*, was not incorporated in the scoping review, as study quality and evidence strength assessment fall beyond the aims of a scoping review.

The organization and reporting of the results of the scoping review was conducted according to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses and its extensions for Scoping Reviews) [79,80]. As the International Prospective Register of Systematic Reviews does not publish protocols for scoping reviews, the protocol for this scoping review has not been registered or published. A summary of the protocol used in this study is presented in the following subsections.

Identifying Research Questions

To our knowledge, no scoping review has mapped existing studies on patient information ownership in the context of big data. Therefore, this scoping review aimed to map and assess published studies on this issue. The specific *aim* of this scoping review was to determine how the medical academic community perceives the issue of ownership of patient information in the context of big data, its possible solutions, and implemented practical applications.

To achieve this goal, we identified and summarized 3 main focus areas of the study: (1) *problem*: ownership of patient information, (2) *area*: health care (medicine), and (3) *context*: big data.

On the basis of existing sources and during initial screening of the publications included in the study, the following research questions were defined:

- What is the *scientific area* in which the ownership of patient information in the context of big data is discussed by the academic community?
- What are the *main aspects* of the ownership of patient information in the context of big data as seen by the academic community?
- How does the academic community *perceive* the ownership of patient information in the context of big data?
- Are there any *solutions* for solving the problem of ownership of patient information in the context of big data *proposed* by the academic community? If there are such proposals, what kind of solutions are proposed?
- Are there any *practical applications* for solving the problem of ownership of patient information in the context of big data discussed by the academic community? If there are such applications, what kind of applications?

Identifying Relevant Studies

Owing to limited access options, we used the following scientific information databases: PubMed, Scopus, Science Direct, and Springer. Given the nature of the topic being investigated, gray literature was not included.

Eligibility Criteria

The eligibility criteria for inclusion and exclusion of articles in the scoping review are presented in Table 1.

Table 1. Eligibility criteria for inclusion and exclusion of publications in the study.

Criteria	Inclusion criteria	Exclusion criteria
Type of publication, availability	Full-text publication (article, review, commentary, and viewpoint), international research or report, published in a peer-reviewed journal or peer-reviewed congress proceedings	• The publication is not in a peer-reviewed journal or peer-reviewed congress proceedings • The publication is a study material or a book (book chapter) • The publication is not available in full text
Indexing	The publication is indexed in at least one of the scientific information databases under consideration	• The publication is not indexed in any of the scientific information databases under consideration
Period	January 2000-October 15, 2019	• Before January 2000 and after October 15, 2019
Language	English	• All others except English
Focus of the publication	The publication covers all 3 focus areas of the study (problem: ownership of patient information; area: health; and context: big data)	• The publication does not cover all 3 focus areas of the study
Relation to research questions	The publication discusses at least one of the research questions (scientific area of publication, aspects of ownership, perception of ownership, proposed solutions, and practical applications)	• The publication does not discuss any of the defined research questions • The publication discusses mainly human genomics (to our perception, this is a major scientific area that needs to be discussed in a separate study) • The publication discusses mainly major and/or specific legal regulations and ethical considerations (to our perception, these are major scientific areas that need to be discussed in separate studies)

Search Strategy

An initial standard (title, abstract, and keywords) search in July 2019, using the keywords "ownership," "health," and "Big Data," identifying the 3 focus areas of our study, proved to be inappropriate. To obtain more adequate information, a search strategy was developed. The search strategy was built initially for PubMed using additional keywords such as "owns" and "property" along with "ownership," and "patient" and "medic*" and "clinic*" along with "health." It was adopted for the requirements of the different databases. To obtain more relevant and adequate information and to save the screening of thousands, a priori unsuitable for the purposes of our study sources, we used additional filters in compliance with our eligibility criteria and the corresponding databases. The search strategy is described in Multimedia Appendix 1.

Study Selection

On the basis of our eligibility criteria, a systematic search within the chosen bibliographic databases (BDBs) using the conforming search strategy was conducted on October 15, 2019.

Further study selection was performed by assessing the suitability of the identified articles from the initial search to the eligibility criteria. Two authors independently screened the articles based on the eligibility criteria at the title and abstract levels, followed by a full-text screening. Any disagreements between the 2 authors were discussed on a case-by-case basis and resolved in consultation with the third author to ensure consensus. As required, the reference list of all identified articles was searched for additional studies meeting our inclusion criteria. No additional articles were included in our study. Mendeley and Zotero were used to filter duplicated articles and to facilitate the screening process.

Data Items and Data Collection Process (Charting the Data)

During data extraction, the following article summary information (standard categories) was charted:

- Author
- Title of publication
- Year of publication
- Country of author (in case of several authors, the institution and the country of the corresponding author were considered)
- Authors' affiliation (academic, research, or nonacademic)
- Type of publication (article or review, congress paper, etc)
- Publication media (journal, conference proceedings, etc)
- Journal title

On the basis of our research questions, we outlined the following *categories for classifying the eligible publications and mapping the available scientific evidence* in our study:

- *Scientific area* of publication in which the ownership of patient information in the context of big data is discussed by the academic community.
- *Aspects of ownership* (of patient information in the context of big data as seen by the academic community)
- *Perception of ownership* (of patient information in the context of big data as seen by the academic community)
- *Proposed solutions* for patient data ownership issues in the context of big data
- *Practical applications* for solving the problem of patient data ownership in the context of big data

All 13 categories were organized in an Excel (Microsoft Corporation) file. Pairs of authors extracted data from the included publications and filled them in an Excel sheet. Any disagreements were resolved by a third author.

As most of the publications eligible for our study were published in impact factor journals and broadly cited, and considering that such bibliometric indicators illustrate a higher quality of the publications, we included 3 additional categories in our study:

- Journal *impact factor*
- *Citation counter* of the publication
- BDB with the largest citation counter of the publication

All publications included in our study were additionally checked in the corresponding information databases and/or the official journal websites to determine the impact factor of the journal and the citation counter of the publications. The data were registered in our Excel file as additional categories (ie, 14-16).

Collating, Summarizing, and Reporting the Results

All authors discussed and agreed on the registered final results. To analyze the results of our research questions, we used both qualitative and quantitative methods. Most of the results were synthesized in narratives. For numerical summarizations, corresponding to the study categories, we used descriptive statistics (alternative analysis). The majority of the results are displayed in tables, graphs, and narratives.

Results

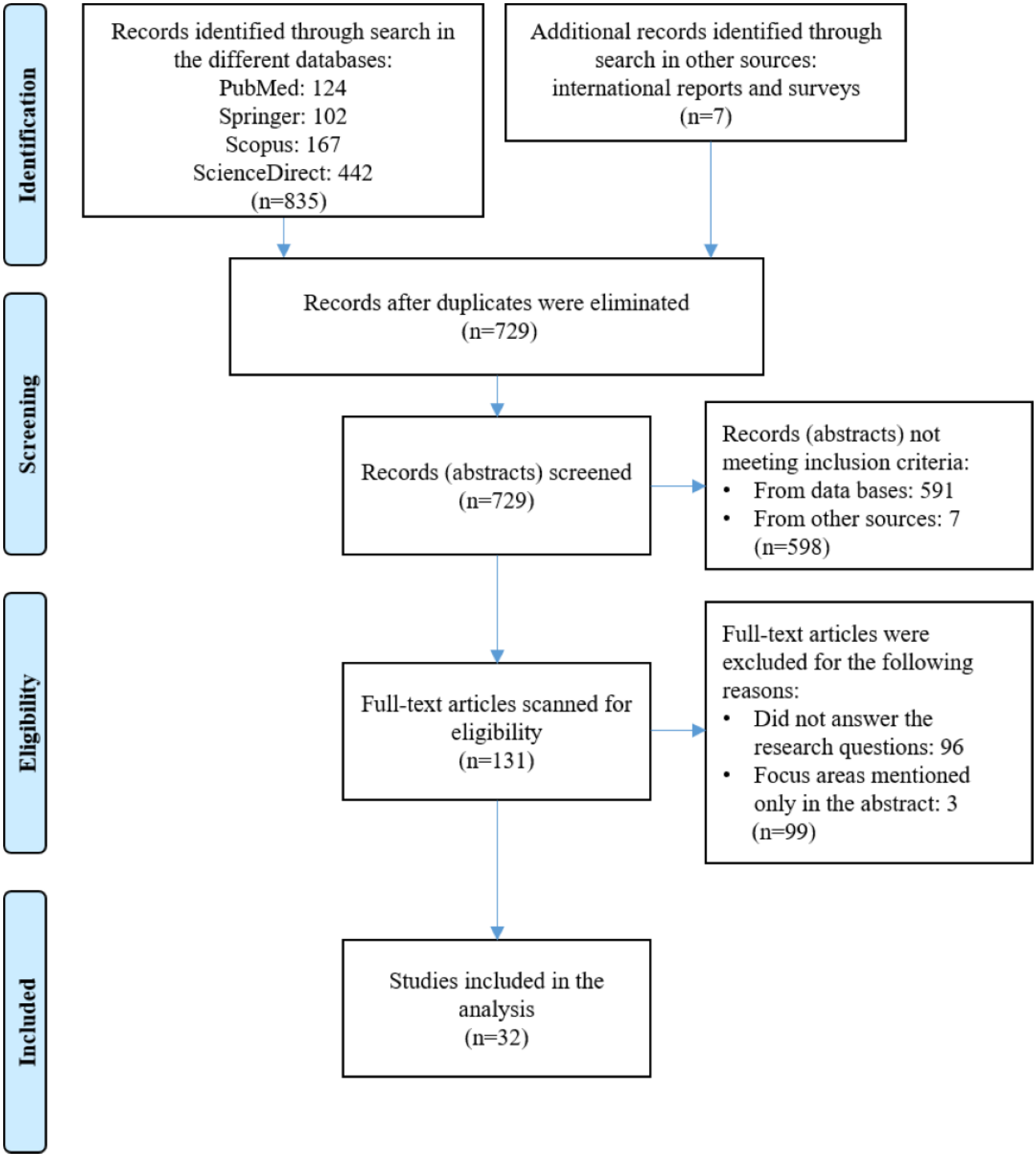
Search Results

A total of 835 publications met our eligibility criteria. Seven publications from other sources, mainly EU and international surveys and reports, were also considered. After filtering the duplicated articles, the total number of potentially appropriate articles was 729.

After screening the file with the bibliographic information and the abstracts of all potentially eligible publications (N=729), manually were excluded duplicated records (n=11), records without an abstract (n=19), records for which no full text was available or full text was not in English (n=85), publications that were not articles or reviews authored by academicians, published in full text in referenced and indexed journals (n=43 plus 7 from other sources), and publications that did not answer any of the research questions (n=433). For a detailed screening, we obtained 131 full-text publications in English. Each of the 131 publications was carefully read and 99 publications were excluded from the study, as they did not cover all 3 focus areas of the study (focus areas mentioned only in the abstract) or did not answer our research questions. The final number of publications included in our study remained to be 32. A PRISMA flow diagram for scoping reviews to visually report the search screening process is presented in Figure 1.

All characteristics of the included publications are presented in Multimedia Appendix 2 [81-112].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) flow diagram of scoping review results.



General and Bibliometric Characteristics of the Eligible Publications

Some of the main general and bibliometric characteristics of the sources included in the scoping review are presented in Table 2. The publications are arranged in alphabetical order of the first author's surname.

Table 2. Main general and bibliometric characteristics of the sources included in the scoping review.

References	Year of publication	Countries	Institutions	Types of publication	Journal titles
Andanda [81]	2019	South Africa	Academic	Article	IIC - International Review of Intellectual Property and Competition Law
Andreu-Perez et al [82]	2015	United Kingdom	Academic	Article	IEEE Journal of Biomedical and health informatics
Asche et al [83]	2017	United States	Academic	Article	Pharmacoeconomics
Ballantyne and Stewart [84]	2019	New Zealand	Academic	Article	Asian Bioethics Review
Balthazar et al [85]	2018	United States	Academic	Article	J Am Coll Radiol.
Bietz et al [86]	2018	United States	Academic	Article	J Am Med Inform Assoc
Cvrkel [87]	2019	United States	Academic	Article	Journal of Dentistry
Esmaeilzadeh and Mirzaei [88]	2018	United Kingdom	Academic	Review	J Med Internet Res.
Heitmueller et al [89]	2014	United Kingdom	Academic	Report	Health Affairs
Hölbl et al [90]	2018	Slovenia	Academic	Article	Symmetry
Hulsen [91]	2019	United Kingdom	Academic	Article	Front.Med.
Hunter [92]	2016	United Kingdom	Rethink Technology Research	Article	EMBO Rep.
Ienca et al [93]	2018	Switzerland	Academic	Article	PLoS ONE
Kaplan [94]	2016	United States	Academic	Article	Camb Q Healthc Ethics.
Kaplan [95]	2015	United States	Academic	Article	Camb Q Healthc Ethics.
Kish and Topol [96]	2015	United States	Scripps Research	Extended commentary	Nat Biotechnol.
Kostkova et al [97]	2016	United Kingdom	Academic	Article	Front. Public Health
Kruse et al [98]	2016	United States	Academic	Review	JMIR Med Inform
Kulynych and Greely [99]	2017	United States	Academic	Article	J Law Biosci.
Maher et al [100]	2019	United States	Academic	Article	International Journal of Medical Informatics
Mamoshina et al [101]	2017	United States	Academic	Article	Oncotarget
Mikk et al [102]	2017	United States	MITRE[a]-Research	Viewpoint	JAMA
Mittelstadt and Floridi [103]	2016	United Kingdom	Academic	Review	Sci Eng Ethics
Roehrs et al [104]	2017	Brazil	Academic	Article	J Med Internet Res.
Timmins et al [105]	2018	United Kingdom	Academic	Review	International Journal of Obesity
Vayena and Blasimme [106]	2017	Switzerland	Academic	Article	Journal of Bioethical Inquiry
Vazirani et al [107]	2019	United Kingdom	Academic	Article	J Med Internet Res
Viceconti et al [108]	2015	United Kingdom	Academic	Article	IEEE Journal of Biomedical and health informatics
Willems et al [109]	2019	The Netherlands	Academic	Article	Oral Oncology
Xafis and Labude [110]	2019	Singapore	Academic	Article	Asian Bioethics Review
Yaffe [111]	2019	Canada	Academic	Article	Seminars in Nuclear Medicine
Yue et al [112]	2016	China	Academic	Article	J Med Syst.

[a]MITRE: Massachusetts Institute of Technology Research and Engineering Research.

The 32 publications were from January 2014 to October 2019: 3% (1/32) in 2014 [89]; 13% (4/32) in 2015 [82,95,96,108];

19% (6/32) in 2016 [92,94,97,98,103,112]; 19% (6/32) in 2017 [83,99,101,102,104,106]; 19% (6/32) in 2018 [85,86,88, 90,93,105]; 28% (9/32) in 2019 [81,84,87,91,100,107,109-111].

The majority of the publications (12/32, 38%) were from the United States [83,85-87,91,94-96,98-102]; 31% (10/32) were from the United Kingdom [82,88,89,92,97,103,105,107,108]. Four (12%) publications were from other European countries (2 from Switzerland [93,106], 1 from Slovenia [90], and 1 from the Netherlands [109]). Six (19%) publications were from other countries: Brazil [104], Canada [111], China [112], New Zealand [84], Singapore [110], and South Africa [81].

The majority of the publications were articles (25/32, 78%). Only 4 (12%) of the publications were reviews [88,98,103,105]. There was also 1 report [89], 1 extended commentary [96], and 1 viewpoint [102].

Almost all of the publications are by authors employed in academic institutions (29/32, 91%). Three (9%) publications [92,96,102] are by academicians employed in research centers.

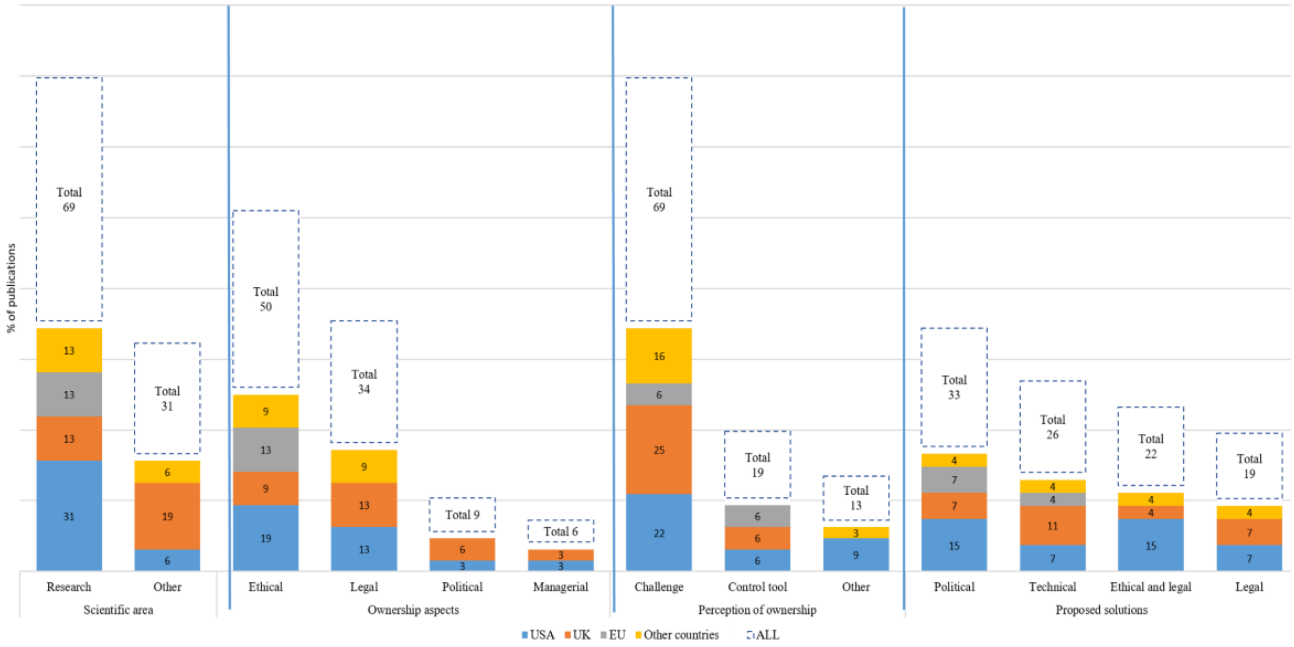
The majority of the publications are highly cited: 66% (21/32) of publications are cited 10 or more times. The other 10 publications are cited between 1 and 9 times, and only 1 publication from 2019 has not been cited yet.

The publications dealing with the problem of ownership of patient information in the context of big data are published by the academic community in 27 journals covering different scientific areas: (1) medicine, biomedical sciences, and public health—15 journals, 47% (15/32) of publications [83,85, 87,89-93,96,97,101,102,105,109,111]; (2) health and medical informatics—6 journals, 28% (9/32) of publications [82, 86,88,98,100,104,107,108,112]; and (3) medical ethics, law, and health policy—6 journals, 25% (8/32) of publications [81,84,94,95,99,103,106,110].

Categories for Outlining the Scope of Available Scientific Literature on the Research Questions Under Study

A summarized distribution of the publications included in the scoping review by countries and categories, outlining the scope of the available scientific literature on patient information ownership in the context of big data, is presented in Figure 2.

Figure 2. Distribution of the publications by countries and categories, outlining the scope of available scientific literatute on the patient information ownership in the context of big data.



Scientific Areas of the Publications

The distribution of the publications according to the *scientific area* in which the ownership of patient information in the context of big data as discussed by the academic community is presented in Figure 2. The following scientific fields were covered: *Research* (22/32, 69%) [81-86,88,91-95,97,99-101,105,106, 108-111]; *Other* (10/32, 31%), including (1) *ownership and control of medical data* [96,102]; (2) *medical records* [104,107]; (3) *big data use in health care* [89,98]; (4) *application of blockchains* [90,112]; and (5) *ownership as an ethical issue in the context of big data* [87,103].

Aspects of Ownership

The distribution of the articles according to the *aspects of ownership* of patient information in the context of big data as seen by the academic community is presented in Figure 2. The following aspects were distinguished: (1) *Ethical* (16/32, 50%) [83,85,87,88,90,93,94,100,102-106,109-111]; (2) *legal* (11/32, 34%), including *legal* [81,82,92,96,101,112] and *legal and ethical* [84,95,99,107,108]; (3) *political* (3/32, 9%), including *political* [89,97] and *political and legal* [86]; and (4) *managerial* (2/32, 6%) [91,98].

Perception of Ownership

The distribution of the articles according to the *perception of ownership* of patient information in the context of big data as seen by the academic community is presented in Figure 2. The following alternatives were distinguished: (1) *challenge* (23/32, 72%) [81-89,91-93,97-100,104,105,107-111]; (2) *control tool* (5/32, 16%) [90,101-103,106]; and (3) *other* (4/32, 13%), including: *problem* [95,112], *threat* [94], and *opportunity* [96].

Proposed Solutions

The distribution of the articles according to the *proposed solutions* for patient data ownership issues in the context of big data is presented in Figure 2. The following types of solutions were proposed: (1) *political* (9/32, 28%), including *political* [97,98,100,106,108], *political and legal* [86,99], and *political and technical* [109,111]; (2) *technical* (7/32, 22%), including *technical* [82,83,88,90,101,112] and *technical and legal* [107]; (3) *ethical and legal* (6/32, 19%) [85,87,94,95,103,110]; and *legal* (5/32, 16%) [81,89,92,96,102].

Five publications did not propose solutions [84,91,93,104,105].

Practical Applications

Only 3 (9%) publications presented *practical applications* for solving the problem of patient data ownership in the context of big data [106,107,112].

Discussion

General and Bibliometric Characteristics

The first publication discussing patient information ownership in the context of big data is from 2014, and the number of publications constantly increases. Considering that for the first 10 months of 2019, the number of publications is higher than the average of the previous years, we could expect an increase in the scientific developments in the future.

The authors are mainly from the United States and the United Kingdom. These are the countries where most work was done in the field of medical and health informatics, including the protection of privacy, confidentiality, and ownership of medical data, the use of ICTs, electronic personal records, and big data. However, a new trend was noticed starting in 2018—an increase in scientific developments from authors outside the United States and the United Kingdom, an indicator that academicians from different countries and continents became more involved in discussing and solving the problem of patient information ownership in the context of big data.

Only 3 publications are by academicians not currently employed in academic institutions. The authors are employed in research centers known for their developments in the field of medicine and health, respectively, Rethink Technology Research (UK) and the Scripps Research and the Massachusetts Institute of Technology Research and Engineering, both from the United States, and the publications are published in high impact factor journals.

Most of the publications are articles. The aim of the reviews included in this study is different from the aim of this review. The report, the extended commentary, and the viewpoint included in the study represent important aspects of ownership of patient information and have been repeatedly cited. All publications included in the study are published in scientific peer-reviewed journals with a high impact factor, indexed in one or more BDBs.

We may say that articles dealing with patient information ownership are published predominantly in medical journals (including health and medical informatics). This is reasonable considering that ownership of patient data is discussed mainly as a challenge to medical research, which in the era of big data seems impossible without the involvement of health and medical informatics instruments. Only one-fourth of the articles included in this scoping review are published in specialized journals covering medical ethics, law, and health policy. Hopefully, the publication activity in this area will increase, as medicine and health care, including medical research, are really in urgent need for an adequate legal framework and policies. There are no technical journals, as they do not address the issue of ownership of medical data at all as well as journals that are purely ethical or legal, as they are highly specialized, poorly cited, and in most cases not indexed in the BDBs under consideration.

Scientific Area of the Publications

Although the issue of ownership of patient information in the context of big data is extremely important, it appears that the academic community's interest in it is rather low. During PubMed, Scopus, ScienceDirect, and Springer searches, we encountered thousands of publications discussing the application of big data in health care and medicine, hundreds of publications discussing privacy and confidentiality, but only 32 reflecting the issue of ownership of medical data met our inclusion criteria and were included in our study. From the analysis, we identified the following scientific areas in which the publications included in this scoping review can be classified: (1) *research*; (2) *ownership and control of medical data*; (3) *medical records*; (4) *big data use in health care*; (5) *application of blockchains*; and (6) *ownership as an ethical issue in the context of big data*.

The majority of the publications discuss the issue of medical data ownership in the context of big data in the area of *research*: (1) *medical research*: clinical research [82], use of personal health data for medical research [86], precision medicine [91], biomedical research [101,106], obesity research [105], biomedical computing [108], head and neck cancer [109], and medical imaging [111]; (2) *access to and use of health or medical data for research*: impact of data ownership on data sharing and implementation of big data in health-related research [81], validation and data connection [83,85], public-private partnerships to share, analyze, and use biomedical big data [84], the use of personal medical data from wearables [84], data exchange [88], and open data for health care research [97]; (3) *secondary use of medical data for research:* medical data sale [92,94], marketing [95], genetic data from medical records [99], use of passive data [100], and creation and use of data depositories in view of reuse of health data [110]; and (4) *ethical challenges to big data in medical and biomedical research* [93].

Two publications are in the area of *medical records*: ownership discussed as a way of patient communication with personal

health records [104] and the use of blockchains for solving medical records problems [107].

Two other publications are in the area of using *big data in health care*: development of public policies to advance the use of big data in health care [89] and the challenges and opportunities of big data in health care [98].

The *application of blockchain* and other ICTs in health care is another research area: research on the use of blockchains in health care [90] and an application using blockchain architecture to assist patients in owning, controlling, and safely exchanging their own data [112].

The ownership of patient information as an *ethical issue in the context of big data* is discussed in 2 publications: data ownership as a complex concept, one of the areas of concern for ethical risk [103] and problems of data access, data ownership, and who has the rightful authority to authorize and profit from the use of the data in mobile health apps [87].

Only 2 of the publications discuss the ownership of medical data as a self-contained study and not as a part of another investigation. Both publications are in the area of *ownership and control of medical data*. The first one discusses the question of why patients should be the owners of their medical data [96], and the other explains why the patients deserve to be the owners and have control over their medical data [102].

Patient data ownership in the context of big data is discussed by the academic community in different scientific fields, each very important and up to date. Although not contradictory, there is a great discrepancy in the publication activity. In addition to the area of research, the other scientific areas are poorly represented and need to be better studied in the future. Future research should also investigate the issue of ownership as a core research question and not as a minor fragment among other topics.

Aspects of Ownership

The ownership aspects discussed in the publications are not clearly distinguished but can be summarized as follows: (1) *ethical*; (2) *legal*; (3) *political*; and (4) *managerial*, and in combinations such as *legal and ethical* and *political and legal*.

In most publications, ownership of patient data, whether primary or secondary, is considered *ethically, legally or ethically, and legally.* The legal and ethical are not necessarily the same, as Kaplan [95] points out, but their common ground must be found with regard to property.

Data have been created and used since the beginning of civilization [113]. What is changing is the speed at which we create and store data, and the fact that we already have not only methods but also processing capacities that allow us to extract useful information from this vast amount of data [114]. Hence, some of the main *ethical aspects* regarding patient-generated data—who owns it and how it can be used, controlled, exchanged or shared, and preserved—are considered in 50% (16/32) of the publications [83,85,87,88,90,93,94,100, 102-106,109-111], in most cases regarding the use of the data for research.

About one-fifth of the publications consider ownership of medical data in *legal terms.* There are no contradictions among the different opinions. They are predominantly focused on the need for the development of an adequate, alternative, harmonized legal framework, giving the individuals the right to own their health data and the adequate use of that data by the different stakeholders [81,82,92,96,101,112]. Ownership is a concept that is ill-suited for governing rights in big data, and the emergence of big data calls for an alternative normative framework with a view to ensuring fair access while minimizing legal and ethical risks [81]. This assumption is based on the United Nations Educational, Scientific, and Cultural Organization's (UNESCO) observation that the concept of ownership is no longer an adequate normative framework in the era of big data [115]. According to Hunter [92], as the volume and scope of collected personal health data increases, the greatest requirement is greater transparency regarding the use of that data, which should be harmonized. Kish and Topol [96] state that for the benefits of digital medicine to be fully realized, we not only need to find a shared home for personal health data but also give individuals the right to own them, and the issue of personal identity data is a historical challenge for lawyers. Mamoshina [101] argues that patients have no control over access privileges to their medical records and remain unaware of the true value of the data they have and the idea that they should own it. It cannot be determined whether a proprietary regime that allows total control of the data would actually be the best solution for patients, provided that medical information may be an enigma to them.

The other 5 publications view ownership in *legal and ethical aspects* [84,95,99,107,108]. Ballantyne and Stewart [84] discuss questions of ownership, both of the data and any resulting intellectual property or products. According to Kaplan [95], individuals should be aware of how data about them are collected and used, as using that data might be crucial. Moreover, it is not legally settled whether the data are merely *spoken words* or *property*. Consideration should be given to how these data are used and the ethical development of social norms and laws, as new technologies affect the integrity and protection of health data. On the other hand, according to Viceconti, all medical data are highly sensitive and, in many developed countries, are considered legally owned by the patient, and the health care provider is required to respect patient confidentiality. However, the need for individual confidentiality can be in conflict with the interests of society [108].

Two publications discuss ownership in a *political aspect*. Heitmueller [89] discusses ownership as a policy lever—the devolution of responsibility of data ownership to patients and how patients decide who they want to share their information with can improve health care and may also be a viable alternative to the extremely difficult task of making existing health care systems interoperable. Concerning the opening of patient-generated data to medical research, Kostkova [97] argues that user privacy and ownership of user-generated data remain an underexplored territory from policy and regulatory perspectives while becoming a booming business for the social media industry. In the absence of transparent data ownership regulation, 2 strikingly different approaches emerged for data

ownership, usage, and sharing: first, government-regulated clinical and research medical data and, second, private user-generated health data collected from social media, apps, online searches, and wearable devices.

Only one publication presents a survey, carried out among patients and researchers concerning barriers, including political and legal, to the use of personal health data in research, where medical data ownership is discussed in a *political and legal aspect* [86].

To be resolved, the problem of ownership, set initially as ethical, needs the cooperation of legal and political institutions as well as the capabilities of modern technologies. This illustrates a consistent structure from problem to solution—the first step is ethical analysis and the determination of right or wrong, useful or harmful, and fair or unjust; the second step is a subsequent public and political debate; and the third step is a law, regulation, or rulemaking act.

Two other publications discuss patient data ownership in a somewhat different aspect, that is, *managerial*. Hulsen [91] states that to fully realize the potential of big data, we must alter the way we work—forming collaborative networks to share samples, data, and methods as well as the legal and ethical frameworks necessary to build and maintain public trust and ensure equitable data access. According to Kruse [98], managerial issues such as governance and data ownership will need to move up on the priority list of organizations, and it should be treated as a primary asset instead of a byproduct of the business. This issue is supported by many authors. In fact, the link between medicine and business, most often expressed in the sale of medical data, both for medical and public health research and research in pharmacoepidemiology as well as for commercial purposes, is a growing trend [92,94,95,97,101].

Perception of Ownership

The ownership of patient data is considered as a (1) challenge, (2) control tool, (3) problem, and (4) threat and (5) an opportunity. According to our analysis, it is taken primarily as a *challenge.*

The ownership of medical data is generally considered a *challenge to perform medical research* [81,85,92,99], including research in the field of obesity [105], oral oncology, concerning patients having governance of their own data sets, especially when the data are linked to different sources [109], medical imaging [111], the use of passive medical data for research [100], the secondary use of medical data for research, especially in the field of the human genome and electronic medical records [98], and leveraging personal health data for medical research [86].

Data access challenges, such as data ownership, data security, and data value, are often considered as barriers to access [83,87], data sharing and exchange [88], access to electronic medical records [107], the use of personal health data [104], and the collection, storage, and reuse of research data [110]. Hulsen et al [91] state that although medical data appear to belong to medical institutions, “the data is the property of the patient and the access and use of that data outside of the clinical realm requires patient consent. This immediately puts a brake on the rapid exploitation of the large volume of data already held in clinical records for precision medicine” [91].

Data ownership, along with data privacy, privacy and security, and data management are serious *social and legal challenges* to big data [82].

In several articles, ownership of medical data is considered a *challenge to health policy* [89], an *ethical challenge* to public-private partnership [84] and modern technologies [108] as well as a *challenge to medical research and business* (sales and data sharing) [97]. An analysis by Ienca et al [93] reveals that the current ethical debate is being largely monopolized by issues of privacy and data protection. However, the issue of data ownership, although distinguished as an ethical challenge, does not actually appear as an ethical priority.

As seen, the challenges, although different in nature, are not contradicting. They concern diverse issues predominantly in performing medical research and activities in association with it as data access, data management, and data sharing. We may argue that these challenges emanate from poor ethical, legal, and policy implications. It will be difficult to overcome such challenges, and the combined efforts and expertise of multidisciplinary research teams are needed.

Apart from being a challenge, ownership of health data is also seen as a *control tool* [102,103,106], controlling access to data [90] and controlling data for biomedical research [101]. There is a slight discrepancy between data access as a challenge and data access as a control tool. This may be resolved if more research is performed.

Two publications consider the ownership of medical data as a *global problem* (along with the use of health data and patient and clinical data protection) for biomedical informatics, patient and physician confidentiality, and regulatory authorities [95] and as a problem (deficiency) for the use of modern blockchain technologies [112].

Kaplan considers the ownership of medical data as a *threat* to the secondary use of data, especially when selling health data [94]. On the contrary, Kish and Topol [96] assume the ownership of medical data as a civil right as an *opportunity* or a strategy for further digitalization of medicine.

The last perceptions of ownership need to be further explored, as it is difficult to judge their significance based on a few opinions.

Proposed Solutions

A solution to the issue of ownership of patient data in the context of big data is discussed in most publications. Only 5 publications do not offer such a solution [84,91,93,104,105]. Apart from the application of different, mainly new technologies, the solutions are rather proposals to the governments and the governing bodies of the health care institutions and are primarily concerned with enhancing the legal framework, developing adequate policies, finding consensus between ethical and legal aspects, in most cases, mainly related to the right to ownership and control of patients on their own data and the protection of data integrity. Considering that technologies are not a problem, the academic community should be more active in developing concrete

solutions and not just proposals, whether technical, legal, or political.

The proposed solutions can be summarized in the following categories: (1) *political;* (2) *technical*; (3) *ethical and legal*; and (4) *legal* and combinations between them: *political and legal*, *political and technical*, and *technical and legal.*

The majority of the proposed solutions concern *policy* decisions, although some of the solutions combine political, legal, or political and technical measures. In general, most of these solutions refer to medical research and the opportunities that big data provide for conducting international medical studies with the exclusive requirement to respect the right of ownership on patient data. Kostkova et al [97] proposed public and political discussions on ownership and responsibility for patient-generated data, as well as the development of a public policy to preserve personal information, which at the same time allows the use of such data to improve public health. According to Kruse et al [98], data ownership and management need to create new business roles that involve analyzing and organizing big data for universal accessibility and sharing and transparency between health care organizations. Maher et al [100] proposed to allow the active involvement of individuals in informed consent procedures and shared ownership of data across countries; researchers should use standardized and validated ways to securely share data, and survey participants should be aware of their data ownership. Extended control through participation management schemes was proposed by Vayena and Blasimme [106] to develop networks of regional cooperatives, potentially worldwide, and to offer open source software for the development of data analysis tools. In this way, "the idea that individuals have direct control over their data can be applied to different national characteristics as well as to international research projects aimed at analyzing data from different countries." Viceconti et al [108] proposed to fund domain-targeted research that allows specialized solutions to be developed for specific applications in biomedical computing and research.

Two publications combine *policy and law*: developing policies and legal norms for ownership between different countries [86] and offering patients some degree of control over their own data, especially when used for scientific research [99].

Political and technical solutions have also been proposed. According to Yaffe [111], instead of severely restrictive policies that do not benefit anyone, reasonable policies regarding the security and privacy of medical data are needed to allow more flexible access and safe exchange (including internationally) of health data between institutions, especially when it comes to medical research and access to various registries (cancer, mortality, rare diseases, etc). Willems et al [109] proposed that instead of bringing together all kinds of data sets in a central comprehensive database, a likely scenario might be that big data users will develop more organic, decentralized virtual networks.

About one-third of the proposed solutions are related to the use of different *technologies*, highlighting solutions related to the application of blockchains: using blockchains to preserve and protect data ownership [88] and to own and share data and health records, and access control [90,112]. The use of blockchains and artificial intelligence enables users to gain ownership and access privileges to their data as well as allow them to sell their data directly to consumers at a fair price [101]. Other technological solutions are also proposed: the use of an identifier for the data collected for the individual with the aim of preserving data security at all levels of the health system, including at every point where the data are collected [82]; the use of distributed networks to provide adequate access to the data, both in efficacy and pharmacoepidemiological studies [83]; and the use of secure multi-party computing—secure multilateral computing that allows third parties to perform calculations with patient data without compromising their integrity [112].

One publication combines *technologies* with *legal* norms. According to Vazirani et al [107], with appropriate regulatory documents and standards, blockchains can serve as a means of managing informed access to health data, as some of their most important features are security, confidentiality, and legal restrictions. This will increase operational interoperability without compromising security while respecting patients' data confidentiality [107].

With modern technologies, these solutions are fully adequate and feasible. Unfortunately, problems with ownership of medical information in the context of big data cannot have only a technical solution. To provide an adequate solution, it is necessary to develop an appropriate legal framework, backed by appropriate policy decisions and in compliance with the corresponding ethical standards.

There are no proposals for pure ethical solutions, as such solutions cannot be directly applied. They need to be incorporated within the corresponding legal regulations or policies. Six publications propose *ethical and legal* solutions. According to Balthazar et al [85], the community of radiologists, ethics professionals, and computer scientists must negotiate the appropriate way to deal with privacy, confidentiality, data ownership, informed consent, epistemology, and inequalities in the most equitable, ethical way. Cvrkel [87] proposed moving to a consent-focused framework: incorporating data ownership and access and profit agreements into well-developed informed consent. The combined efforts and expertise of lawyers, ethics professionals, and computer scientists on the legal and ethical collection and the use of data, together with the technical knowledge to combine and identify them, can contribute to the development of more informative policies [94,95]. According to Mittelstadt and Floridi [103], taking into account both forms of ownership, the right to *control* and the right to *profit from* data, to exercise adequate data access rights in the big data era, it is necessary to define the terms *commercial* and *scientific* value and to develop specifications for adequate rights and, where necessary, restrictions on access, as well as to modify data protection practices or legislation to oblige *data keepers* to provide data owners with reasonable access to them, so far as this is possible. A similar issue is proposed by Xafis and Labude [110]—state the conditions of access to data repositories—so that data owners can make appropriate decisions regarding the level of access they believe is appropriate for their data and research materials.

One *legal* solution concerns legal arrangements for appropriate *custodianship* of big data to ensure that data subjects maintain some control over access and future uses of their data, while delegating decision making in some matters to the data custodians. Such delegated decision making gives rise to custodial rights, not ownership of the data [81]. This proposal follows UNESCO's recommendation that a framework with new approaches to ownership and custodianship of personal data should be developed [115]. It may be expected that this idea will contribute to the development of adequate political and legal decisions in view of patient data ownership in the modern big data reality. Other proposals include developing public policy to advance the use of big data in health care, including delegating patient data responsibility to the patients and creating shared networks [89]; improving legal frameworks to protect patient anonymity, informed consent, and data quality assurance [92]; the individuals should be given the right to own their medical data, promoting the ownership of medical data as a civil right and as a major strategy for the further digitalization of medicine, providing new resources to potentially assist any individual who wishes to participate in that [96]; and legal arrangements for ownership as control of patient data (data use agreement) between patients and third parties (data managers, ie, health care and trade organizations), which will allow individuals to control their longitudinal digital records, can improve patient engagement, data accuracy, and health outcomes [102].

Applications

Despite the many solutions proposed, real applications related to ownership of patient information in the context of big data are only commented on in 3 publications.

One application is presented by Vayena [106]. This is the MIDATA cooperative model developed by MIDATA, Switzerland [116], that provides an example of how individuals can gain control over their own data through new type management mechanisms. The purpose of MIDATA is to store health-related data from a variety of sources and to provide it for scientific projects, while allowing data owners to make their own decisions about their data use. It is a nonprofit organization, but the potential profits generated by consumers will be reinvested in the maintenance of the cooperative or the funding of research [106]. Owners who are registered with MIDATA can actively contribute to medical research and clinical trials by allowing selective access to their own data. They may become members of the cooperative and participate in its management. The MIDATA model is designed for international application: MIDATA Switzerland supports the creation of regional and national MIDATA cooperatives that share the data platform and infrastructure.

The second application is cited by Vazirani et al [107], who described several applications of blockchain technologies for electronic health records, including MedRec [117], which uses blockchain technology and smart contracts to access patient data and manage access permissions. Other applications are mentioned in the review, but they do not affect the ownership of medical data.

The third application is presented by Yue et al [112], who presents an architecture of an application called the Healthcare Data Gateway based on blockchain, which, in addition to patients' access to their own clinical data and medical records, allows patients to own, control, share, and manage their own data easily and securely without compromising their privacy and provides a potentially new way to improve the health system's intelligence while maintaining patient data ownership. The data are stored in a private blockchain (centralized database with restricted access control, accessible only to authorized or specific users).

Study Limitations

This study has several major limitations. First, due to limited access options, only 4 scientific BDBs were searched: PubMed, ScienceDirect, Scopus, and Springer. We have not searched other BDBs such as ProQuest, EMBASE (Excerpta Medica dataBASE), and the Web of Knowledge as well as the databases that mainly index technical publications. The search covered the period from January 2000 to October 2019. Thus, the publications that met our inclusion criteria were limited to just 32. Second, our study used only publications in English, which also limited the number of eligible sources. Third, we used mostly sources published in scientific journals authored by academicians. No reports presented at congresses and conferences and published in the corresponding congress proceedings were used, as the full text of the potentially relevant publications was not available even on request. Moreover, reports from different institutions, mainly nongovernmental organizations, could not be used as in most cases they were not authored by academicians and/or were not indexed in the scientific databases.

Conclusions

Big data in medicine and health care is an issue broadly discussed at present. However, the problem of medical data ownership, especially in the context of big data, no matter its relevance, remains somewhat neglected. What we bear in mind is *ownership* and not *privacy* and *control* that are more broadly represented in the scientific literature by the academic community. The results of the scoping review indicate that the problem of ownership of patient information in the context of big data is poorly researched. Only 2 of the publications discuss the ownership of medical data as a self-contained study and not as a part of another investigation. The problem is not addressed consistently and in its integrity—in terms of ethical, political, and regulatory coherence. The issue of ownership has to be discussed in a more comprehensive way, including ownership problems as a core research question and not just mentioned among other issues.

There are 6 scientific fields in which the publications under review can be classified: research, medical records, use of big data in health care, blockchain applications, ownership and control, and ethics, while the area of research is predominant. The other fields are poorly represented and need to be better studied in the future.

The aspects in which publications consider ownership of medical data are not clearly distinguished, but can be summarized as

ethical, legal, political, and managerial, and in combinations such as legal and ethical and political and legal. The largest share is of the publications that consider the ownership of patient information in the context of big data in an ethical aspect. The other aspects need to be further researched.

The ownership of patient data is perceived primarily as a challenge, with this challenge being fundamental to conducting medical research, including access to and use of medical data, which is generally considered a matter of medical ethics. The ownership of medical data is also considered a challenge to policy, ethics, modern technologies, clinical research, and business (data sales and sharing). Apart from being a challenge, ownership of patient data is also seen as a means of control, a problem, a threat, and even an opportunity, also considering primarily medical research. It will be difficult to overcome such challenges, and the combined efforts and expertise of multidisciplinary research teams are needed.

The proposed solutions can be summarized in 3 broad areas: technology, law and ethics, and policy. All the 3 strands are extremely important, but they are not sufficiently represented in the publication activity, and to adequately address the ownership of patient information in the big data information context, these 3 strands need to be combined. To develop and implement an adequate technological solution, it must, in addition to complying with generally accepted ethical standards, comply with certain regulatory documents and policy decisions.

Despite the many solutions proposed, real applications related to the ownership of patient information in the context of big data are commented on in only 3 publications. It is well known that technologies do not prevent the creation of suitable applications. What is missing is adequate policy decisions expressed through the relevant legal frameworks.

The issue of patient information ownership in the context of big data must find its place in the scientific publishing field. It may receive more appropriate answers if special editions of renowned scientific journals are organized to address the issue of ownership of patient information in the context of big data, seminars and roundtables are organized during biomedical forums, and scoping reviews are regularly conducted.

In conclusion, this study may serve as a starting point for future research in this area. It is already evident that technologies are not an obstacle to the development of applications regarding patient data ownership. Cearley et al [118] stated that by 2023, the blockchain will be technically scalable and will support trusted private transactions with the necessary data confidentiality. Given the technological and scientific developments as well as the rapid commercialization of big data in health, the ethical, legal, and policy-making debate on patient data ownership is sure to become more important and widespread.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 13 KB - jmir_v22i8e22214_app1.docx]

Multimedia Appendix 2
Publications included in the scoping review - main characteristics.
[XLSX File (Microsoft Excel File), 16 KB - jmir_v22i8e22214_app2.xlsx]

References

1. Wainwright R, Donck F, Fertik M, Rake M, Savage SC, Cloppinger JH. Personal Data: The 'New Oil' Of The 21st Century. In: World Economic Forum on Europe and Central Asia. 2011 Presented at: WEF'11; June 8-9, 2011; Vienna, Austria URL: https://www.weforum.org/reports/world-economic-forum-europe-and-central-asia-2011
2. Eggers WD, Hammil R, Ali A. Data as the New Currency. Deloitte. 2013. URL: https://www2.deloitte.com/insights/us/en/deloitte-review/issue-13/data-as-the-new-currency.html [accessed 2019-07-15]
3. Hall MA, Schulman KA. Ownership of medical information. J Am Med Assoc 2009 Mar 25;301(12):1282-1284. [doi: 10.1001/jama.2009.389] [Medline: 19318657]
4. Tully J. An Approach to Political Philosophy. Cambridge, UK: University Press; 1993.
5. Macpherson CB. The Political Theory of Possessive Individualism: Hobbes to Locke. Oxord, England: Oxford University Press; 2011.
6. John L. Second Treatise on Civil Government. New York, USA: Blackmore Dennett; 2018.
7. Pateman C. Self-ownership and property in the person: democratization and a tale of two concepts. J Polit Philos 2002 Mar;10(1):20-53 [FREE Full text] [doi: 10.1111/1467-9760.00141]
8. Westin A. Privacy and Freedom. London, UK: The Bodley Head; 1967.
9. Waldron J. Property and ownership. In: Zalta E, editor. Stanford Encyclopedia of Philosophy. Stanford, USA: University of Stanford; 2012.
10. Law J, Martin EA. Dictionary of Law. Ninth Edition. Oxford, England: Oxford University Press; 2018.

11. Baron JB. Property as Control: The Case of Information. Michigan Telecommunications and Technology Law Review. 2012. URL: https://pdfs.semanticscholar.org/f57c/d0b8182fa15f8777dba481085f717d2c40a5.pdf [accessed 2020-08-11]
12. Beitz M. Personal Data for the Public Good: New Opportunities to Enrich Understanding of Individual and Population Health. Slide Share. 2014. URL: https://www.slideshare.net/mbietz/hdp-hde-researchbietzslideshare [accessed 2019-11-05]
13. Radin MJ. Property and personhood. Stanford Law Rev 1982 May;34(5):957. [doi: 10.2307/1228541]
14. Koskinen J, Kimppa KK. An unclear question: who owns patient information? In: Kreps D, Fletcher G, Griffiths M, editors. Technology and Intimacy: Choice or Coercion. New York, USA: Springer; 2016.
15. Koskinen JS, Heimo OI, Kimppa KK. A viewpoint for more ethical approach in healthcare information system development and procurement: the four principles. In: Eriksson-Backa K, Luoma A, Krook E, editors. xploring the Abyss of Inequalities: 4th International Conference on Well-Being in the Information Society. Berlin, Heidelberg: Springer; 2012.
16. Lessig L. Privacy as property. Soc Res 2002;69(1):247-269 [FREE Full text]
17. van Blarkom GW, Borking JJ, Olk JG. Handbook of Privacy and Privacy-Enhancing Technologies: The case of Intelligent Software Agents. Andrew Patrick. 2003. URL: http://www.andrewpatrick.ca/pisa/handbook/Handbook_Privacy_and_PET_final.pdf [accessed 2018-04-13]
18. Rodwin MA. Patient data: property, privacy & the public interest. Am J Law Med 2010;36(4):586-618. [doi: 10.1177/009885881003600403] [Medline: 21302847]
19. Koskinen J, Kainu V, Kimppa K. The concept of Datenherrschaft of patient information from a Lockean perspective. J Inf Comun Ethics Soc 2016 Mar 14;14(1):70-86 [FREE Full text] [doi: 10.1108/jices-06-2014-0029]
20. Evans BJ. Much ado about data ownership. Harvard J Law Technol 2011;25(1):62 [FREE Full text]
21. Evans BJ. Would patient ownership of health data improve confidentiality? Virtual Mentor 2012 Sep 1;14(9):724-732 [FREE Full text] [doi: 10.1001/virtualmentor.2012.14.9.pfor1-1209] [Medline: 23351352]
22. Graunt J. Natural and political observations mentioned in a following index, and made upon the bills of mortality. In: Mathematical Demography. Berlin, Heidelberg: Springer; 1975:11-20.
23. Cukier K, Mayer-Schönberger V. Big Data: A Revolution That Will Transform How We Live, Work, and Think. London, UK: John Murray Publishing; 2014.
24. Rothman KJ. Lessons from John Graunt. Lancet 1996 Jan 6;347(8993):37-39. [doi: 10.1016/s0140-6736(96)91562-7] [Medline: 8531550]
25. de Mauro A, Greco M, Grimaldi M. A formal definition of big data based on its essential features. Lib Rev 2016 Apr 4;65(3):122-135 [FREE Full text] [doi: 10.1108/lr-06-2015-0061]
26. Mashey JR. Big Data and the Next Wave of InfraStress. USENIX: The Advanced Computing Systems Association. 1998. URL: http://static.usenix.org/event/usenix99/invited_talks/mashey.pdf [accessed 2016-09-28]
27. Lohr S. The Origins of 'Big Data': An Etymological Detective Story. The New York Times. 2013. URL: https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/ [accessed 2016-09-28]
28. Gantz J, Reinsel D. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. EMC. 2012. URL: https://www.emc.com/leadership/digital-universe/2012iview/index.htm [accessed 2018-12-15]
29. Ward J, Barker A. Undefined by data, a survey of big data definitions. arXiv 2013:- epub ahead of print [FREE Full text]
30. Laney D. 3D Data Management: Controlling Data Volume, Velocity, and Variety. Gartner Blog Network. 2001. URL: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf [accessed 2019-12-21]
31. Big Data. Gartner Inc. URL: https://www.gartner.com/it-glossary/big-data/ [accessed 2020-01-04]
32. Sagiroglu S, Sinanc D. Big Data: A Review. In: International Conference on Collaboration Technologies and Systems. 2013 Presented at: CTS'13; May 20-24, 2013; San Diego, CA p. 42-47. [doi: 10.1109/cts.2013.6567202]
33. Gerhardt B, Griffin K, Klemann R. Cisco Data Center. Cisco. 2012. URL: http://www.unleashingit.com/docs/W13/Information-Infomediaries.pdf [accessed 2020-01-04]
34. Hilbert M. Big data for development: a review of promises and challenges. Dev Policy Rev 2015 Dec 13;34(1):135-174 [FREE Full text] [doi: 10.1111/dpr.12142]
35. Beyer MA, Laney D. The Importance of 'Big Data': A Definition. Gartner Inc. 2012. URL: https://www.gartner.com/en/documents/2057415/the-importance-of-big-data-a-definition [accessed 2020-01-05]
36. Chen M, Mao S, Zhang Y, Leung V. Big Data - Related Technologies, Challenges and Future Prospects. New York, USA: Springer International Publishing; 2014.
37. Dobre C, Xhafa F. Intelligent services for big data science. Future Gener Comput Syst 2014 Jul;37:267-281. [doi: 10.1016/j.future.2013.07.014]
38. Owais SS, Hussein NS. Extract five categories CPIVW from the 9V's characteristics of the big data. Int J Adv Comput Sci Appl 2016;7:254-258. [doi: 10.1109/CESYS.2016.7889917]
39. Kapil G, Agrawal A, Khan RA. A Study of Big Data Characteristics. In: International Conference on Communication and Electronics Systems. 2016 Presented at: ICCES'16; October 21-22, 2016; Coimbatore, India. [doi: 10.1109/cesys.2016.7889917]
40. de Mauro A, Greco M, Grimaldi M. What is Big Data? A Consensual Definition and a Review of Key Research Topics. In: AIP Conference Proceedings. 2015 Presented at: AIP'15; September 5-8, 2014; Madrid, Spain. [doi: 10.1063/1.4907823]

41. Dutcher J. What is Big Data? UC Berkeley School of Information. 2014. URL: https://datascience.berkeley.edu/what-is-bigdata/2014 [accessed 2016-02-01]
42. Manyika MC, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute. 2011. URL: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation [accessed 2018-02-14]
43. Study on Big Data in Public Health, Telemedicine and Healthcare. European Commision. 2016. URL: https://ec.europa.eu/digital-single-market/en/news/study-big-data-public-health-telemedicine-and-healthcare [accessed 2020-08-11]
44. O'Driscoll A, Daugelaite J, Sleator RD. 'Big data', Hadoop and cloud computing in genomics. J Biomed Inform 2013 Oct;46(5):774-781 [FREE Full text] [doi: 10.1016/j.jbi.2013.07.001] [Medline: 23872175]
45. Costa FF. Big data in biomedicine. Drug Discov Today 2014 Apr;19(4):433-440. [doi: 10.1016/j.drudis.2013.10.012] [Medline: 24183925]
46. Luo J, Wu M, Gopukumar D, Zhao Y. Big data application in biomedical research and health care: a literature review. Biomed Inform Insights 2016;8:1-10 [FREE Full text] [doi: 10.4137/BII.S31559] [Medline: 26843812]
47. Jee K, Kim G. Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. Healthc Inform Res 2013 Jun;19(2):79-85 [FREE Full text] [doi: 10.4258/hir.2013.19.2.79] [Medline: 23882412]
48. Chawla NV, Davis DA. Bringing big data to personalized healthcare: a patient-centered framework. J Gen Intern Med 2013 Sep;28(Suppl 3):S660-S665 [FREE Full text] [doi: 10.1007/s11606-013-2455-8] [Medline: 23797912]
49. Ashley EA. Towards precision medicine. Nat Rev Genet 2016 Aug 16;17(9):507-522. [doi: 10.1038/nrg.2016.86] [Medline: 27528417]
50. Huser V, Cimino JJ. Impending challenges for the use of big data. Int J Radiat Oncol Biol Phys 2016 Jul 1;95(3):890-894 [FREE Full text] [doi: 10.1016/j.ijrobp.2015.10.060] [Medline: 26797535]
51. Murdoch TB, Detsky AS. The inevitable application of big data to health care. J Am Med Assoc 2013 Apr 3;309(13):1351-1352. [doi: 10.1001/jama.2013.393] [Medline: 23549579]
52. Mancini M. Exploiting big data for improving healthcare services. J e-Learn Knowl Soc 2014;10(2):23-33. [doi: 10.20368/1971-8829/929]
53. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
54. Fernandes L, O'Connor M, Weaver V. Big data, bigger outcomes: healthcare is embracing the big data movement, hoping to revolutionize HIM by distilling vast collection of data for specific analysis. J AHIMA 2012 Oct;83(10):38-43; quiz 44. [Medline: 23061351]
55. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb;11(2):e1003904. [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]
56. Naito M. Utilization and application of public health data in descriptive epidemiology. J Epidemiol 2014;24(6):435-436 [FREE Full text] [doi: 10.2188/jea.je20140182] [Medline: 25327184]
57. Yuen-Reed G, Mojsilovic A. The role of big data analytics in health payer transformation to consumer-centricity. In: Weaver C, Ball M, Kim G, Kiel J, editors. Healthcare Information Management Systems: Cases, Strategies, and Solutions. Switzerland: Springer; 2016:399-420.
58. Hoeren T. Big data and the ownership in data: recent developments in Europe. Eur Intell Prop Rev 2014;36(12):751-754 [FREE Full text]
59. Mayer A. Regulating data as property: a new construct for moving forward. Duke Law Technol Rev 2018;16:220-277 [FREE Full text]
60. Hall M. Property, privacy and the pursuit of integrated electronic medical records. SSRN J 2009 Jan;95(2):26. [doi: 10.2139/ssrn.1334963]
61. van Asbroeck B, Debussche J, César J. Big Data & Issues & Opportunities: Data Ownership. Bird & Bird. 2019. URL: https://www.twobirds.com/en/news/articles/2019/global/big-data-and-issues-and-opportunities-data-ownership, [accessed 2020-06-01]
62. Atchinson BK, Fox DM. The politics of the health insurance portability and accountability act. Health Aff (Millwood) 1997;16(3):146-150. [doi: 10.1377/hlthaff.16.3.146] [Medline: 9141331]
63. Tovino SA. The HIPAA privacy rule and the EU GDPR: illustrative comparisons. Seton Hall Law Rev 2017;47(4):973-993. [doi: 10.1201/b17548-8]
64. Dove ES, Thompson B, Knoppers BM. A step forward for data protection and biomedical research. Lancet 2016 Apr;387(10026):1374-1375. [doi: 10.1016/s0140-6736(16)30078-2]
65. Adibuzzaman M, DeLaurentis P, Hill J, Benneyworth BD. Big data in healthcare - the promises, challenges and opportunities from a research perspective: a case study with a model database. AMIA Annu Symp Proc 2017;2017:384-392 [FREE Full text] [Medline: 29854102]
66. Beauchamp T, Childress J. Principles of Biomedical Ethics. Seventh Edition. New York, United States: Oxford University Press; 2015.

67. Kainu V, Koskinen JM. Between public and personal information - not prohibited, therefore permitted? In: Canellopoulou-Bottis M, editor. Privacy and Surveilance. Current Aspects and Future Perspectives. New York, USA: SSRN; 2013.
68. Munn Z, Peters M, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
69. Tricco A, Zarin W, Ghassemi M, Nincic V, Lillie E, Page M, et al. Same family, different species: methodological conduct and quality varies according to purpose for five types of knowledge synthesis. J Clin Epidemiol 2018 Apr;96:133-142. [doi: 10.1016/j.jclinepi.2017.10.014] [Medline: 29103958]
70. Mays N, Roberts E, Popay J. Synthesising research evidence. In: Fulop N, Allen P, Clarke A, Black N, editors. Studying the Organization and Delivery of Health Services: Research Methods. London, UK: Routledge; Aug 23, 2001:188-220.
71. Ehrich K, Freeman G, Richards S, Robinson I, Shepperd S. How to do a scoping exercise: continuity of care. Res Pol Plan 2002;20(1):25-29 [FREE Full text]
72. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
73. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
74. Peters M, Godfrey C, McInerney P, Baldini SC, Khalil H, Parker D. The Joanna Briggs Institute Reviewers' Manual 2015: Methodology for JBI Scoping Reviews. School of Nursing - LSU Health New Orleans. 2015. URL: https://nursing.lsuhsc.edu/JBI/docs/ReviewersManuals/Scoping-.pdf [accessed 2020-08-11]
75. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 2014 Dec;67(12):1291-1294. [doi: 10.1016/j.jclinepi.2014.03.013] [Medline: 25034198]
76. Moher D, Stewart L, Shekelle P. All in the family: systematic reviews, rapid reviews, scoping reviews, realist reviews, and more. Syst Rev 2015 Dec 22;4:183 [FREE Full text] [doi: 10.1186/s13643-015-0163-7] [Medline: 26693720]
77. Tricco AC, Lillie E, Zarin W, O'Brien K, Colquhoun H, Kastner M, et al. A scoping review on the conduct and reporting of scoping reviews. BMC Med Res Methodol 2016 Feb 9;16:15 [FREE Full text] [doi: 10.1186/s12874-016-0116-4] [Medline: 26857112]
78. Tricco A, Antony J, Zarin W, Strifler L, Ghassemi M, Ivory J, et al. A scoping review of rapid review methods. BMC Med 2015 Sep 16;13:224 [FREE Full text] [doi: 10.1186/s12916-015-0465-6] [Medline: 26377409]
79. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
80. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
81. Andanda P. Towards a paradigm shift in governing data access and related intellectual property rights in big data and health-related research. Int Rev Ind Prop Copyr Law 2019 Oct 15;50(9):1052-1081 [FREE Full text] [doi: 10.1007/s40319-019-00873-2]
82. Andreu-Perez J, Poon C, Merrifield R, Wong S, Yang G. Big data for health. IEEE J Biomed Health Inform 2015 Jul;19(4):1193-1208. [doi: 10.1109/JBHI.2015.2450362] [Medline: 26173222]
83. Asche C, Seal B, Kahler K, Oehrlein E, Baumgartner M. Evaluation of healthcare interventions and big data: review of associated data issues. Pharmacoeconomics 2017 Aug;35(8):759-765. [doi: 10.1007/s40273-017-0513-5] [Medline: 28474299]
84. Ballantyne A, Stewart C. Big data and public-private partnerships in healthcare and research. Asian Bioethics Rev 2019 Sep 30;11(3):315-326 [FREE Full text] [doi: 10.1007/s41649-019-00100-7]
85. Balthazar P, Harri P, Prater A, Safdar N. Protecting your patients' interests in the era of big data, artificial intelligence, and predictive analytics. J Am Coll Radiol 2018 Mar;15(3 Pt B):580-586. [doi: 10.1016/j.jacr.2017.11.035] [Medline: 29402532]
86. Bietz M, Bloss C, Calvert S, Godino J, Gregory J, Claffey M, et al. Opportunities and challenges in the use of personal health data for health research. J Am Med Inform Assoc 2016 Apr;23(e1):e42-e48 [FREE Full text] [doi: 10.1093/jamia/ocv118] [Medline: 26335984]
87. Cvrkel T. The ethics of mHealth: moving forward. J Dent 2018 Jul;74(Suppl 1):S15-S20 [FREE Full text] [doi: 10.1016/j.jdent.2018.04.024] [Medline: 29929583]
88. Esmaeilzadeh P, Mirzaei T. The potential of blockchain technology for health information exchange: experimental study from patients' perspectives. J Med Internet Res 2019 Jun 20;21(6):e14184 [FREE Full text] [doi: 10.2196/14184] [Medline: 31223119]
89. Heitmueller A, Henderson S, Warburton W, Elmagarmid A, Pentland A, Darzi A. Developing public policy to advance the use of big data in health care. Health Aff (Millwood) 2014 Sep;33(9):1523-1530. [doi: 10.1377/hlthaff.2014.0771] [Medline: 25201656]

90. Hölbl M, Kompara M, Kamišalić A, Zlatolas LN. A systematic review of the use of blockchain in healthcare. Symmetry 2018 Oct 10;10(10):470. [doi: 10.3390/sym10100470]
91. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From big data to precision medicine. Front Med (Lausanne) 2019;6:34 [FREE Full text] [doi: 10.3389/fmed.2019.00034] [Medline: 30881956]
92. Hunter P. The big health data sale: as the trade of personal health and medical data expands, it becomes necessary to improve legal frameworks for protecting patient anonymity, handling consent and ensuring the quality of data. EMBO Rep 2016 Aug;17(8):1103-1105 [FREE Full text] [doi: 10.15252/embr.201642917] [Medline: 27402546]
93. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: a scoping review. PLoS One 2018;13(10):e0204937 [FREE Full text] [doi: 10.1371/journal.pone.0204937] [Medline: 30308031]
94. Kaplan B. Selling health data: de-identification, privacy, and speech. Camb Q Healthc Ethics 2015 Jul;24(3):256-271. [doi: 10.1017/S0963180114000589] [Medline: 26059952]
95. Kaplan B. How should health data be used? Camb Q Healthc Ethics 2016 Apr;25(2):312-329. [doi: 10.1017/S0963180115000614] [Medline: 26957456]
96. Kish LJ, Topol EJ. Unpatients-why patients should own their medical data. Nat Biotechnol 2015 Sep;33(9):921-924. [doi: 10.1038/nbt.3340] [Medline: 26348958]
97. Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, et al. Who owns the data? Open data for healthcare. Front Public Health 2016;4:7 [FREE Full text] [doi: 10.3389/fpubh.2016.00007] [Medline: 26925395]
98. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and opportunities of big data in health care: a systematic review. JMIR Med Inform 2016 Nov 21;4(4):e38 [FREE Full text] [doi: 10.2196/medinform.5359] [Medline: 27872036]
99. Kulynych J, Greely H. Clinical genomics, big data, and electronic medical records: reconciling patient rights with research when privacy and science collide. J Law Biosci 2017 Apr;4(1):94-132 [FREE Full text] [doi: 10.1093/jlb/lsw061] [Medline: 28852559]
100. Maher NA, Senders JT, Hulsbergen AF, Lamba N, Parker M, Onnela J, et al. Passive data collection and use in healthcare: a systematic review of ethical issues. Int J Med Inform 2019 Sep;129:242-247. [doi: 10.1016/j.ijmedinf.2019.06.015] [Medline: 31445262]
101. Mamoshina P, Ojomoko L, Yanovich Y, Ostrovski A, Botezatu A, Prikhodko P, et al. Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare. Oncotarget 2018 Jan 19;9(5):5665-5690 [FREE Full text] [doi: 10.18632/oncotarget.22345] [Medline: 29464026]
102. Mikk KA, Sleeper HA, Topol EJ. The pathway to patient data ownership and better health. J Am Med Assoc 2017 Oct 17;318(15):1433-1434. [doi: 10.1001/jama.2017.12145] [Medline: 28973063]
103. Mittelstadt B, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. Sci Eng Ethics 2016 Apr;22(2):303-341. [doi: 10.1007/s11948-015-9652-2] [Medline: 26002496]
104. Roehrs A, da Costa CA, Righi RD, de Oliveira KS. Personal health records: a systematic literature review. J Med Internet Res 2017 Jan 6;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
105. Timmins K, Green M, Radley D, Morris M, Pearce J. How has big data contributed to obesity research? A review of the literature. Int J Obes (Lond) 2018 Dec;42(12):1951-1962 [FREE Full text] [doi: 10.1038/s41366-018-0153-7] [Medline: 30022056]
106. Vayena E, Blasimme A. Biomedical big data: new models of control over access, use and governance. J Bioeth Inq 2017 Dec;14(4):501-513 [FREE Full text] [doi: 10.1007/s11673-017-9809-6] [Medline: 28983835]
107. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing blockchains for efficient health care: systematic review. J Med Internet Res 2019 Feb 12;21(2):e12439 [FREE Full text] [doi: 10.2196/12439] [Medline: 30747714]
108. Viceconti M, Hunter P, Hose R. Big data, big knowledge: big data for personalized healthcare. IEEE J Biomed Health Inform 2015 Jul;19(4):1209-1215. [doi: 10.1109/JBHI.2015.2406883] [Medline: 26218867]
109. Willems S, Abeln S, Feenstra K, de Bree R, van der Poel EF, de Jong RJ, et al. The potential use of big data in oncology. Oral Oncol 2019 Nov;98:8-12 [FREE Full text] [doi: 10.1016/j.oraloncology.2019.09.003] [Medline: 31521885]
110. Xafis V, Labude M. Openness in big data and data repositories. Asian Bioethics Rev 2019 Oct 1;11(3):255-273 [FREE Full text] [doi: 10.1007/s41649-019-00097-z]
111. Yaffe MJ. Emergence of 'big data' and its potential and current limitations in medical imaging. Semin Nucl Med 2019 Mar;49(2):94-104. [doi: 10.1053/j.semnuclmed.2018.11.010] [Medline: 30819400]
112. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. J Med Syst 2016 Oct;40(10):218. [doi: 10.1007/s10916-016-0574-6] [Medline: 27565509]
113. Beam A, Kohane I. Big data and machine learning in health care. J Am Med Assoc 2018 Apr 3;319(13):1317-1318. [doi: 10.1001/jama.2017.18391] [Medline: 29532063]
114. Librenza-Garcia D. Ethics in the era of big data. In: Passos I, Mwangi B, Kapczinski F, editors. Personalized Psychiatry. New York, USA: Springer Nature; 2019:161-172.
115. Report of the International Bioethics Committee of UNESCO (IBC) on Big Data and Health. UNESCO Digital Library. 2017. URL: https://unesdoc.unesco.org/ark:/48223/pf0000248724 [accessed 2019-01-10]
116. MIDATA. 2015. URL: https://www.midata.coop/ [accessed 2019-10-10]

117. Azaria A, Ekblaw A, Vieira T, Lippman A. MedRec: Using Blockchain for Medical Data Access and Permission Management. In: 2nd International Conference on Open and Big Data. 2016 Presented at: OBD'16; August 22-24, 2016; Vienna, Austria URL: https://ieeexplore.ieee.org/document/7573685 [doi: 10.1109/OBD.2016.11]
118. Cearley D, Jones N, Smith D, Burke B, Chandrasekaran A, Lu CK. Top 10 Strategic Technology Trends for 2020. Gartner Inc. 2019. URL: https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2020/ [accessed 2020-02-10]

Abbreviations

BDB: bibliographic database
ICT: information and communication technology
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses and its extensions for Scoping Review
UNESCO: United Nations Educational, Scientific, and Cultural Organization

Edited by G Eysenbach; submitted 06.07.20; peer-reviewed by J Li, A Benis, L Genaro; comments to author 19.07.20; revised version received 24.07.20; accepted 26.07.20; published 18.08.20.

Please cite as:
Mirchev M, Mircheva I, Kerekovska A
The Academic Viewpoint on Patient Data Ownership in the Context of Big Data: Scoping Review
J Med Internet Res 2020;22(8):e22214
URL: http://www.jmir.org/2020/8/e22214/
doi:10.2196/22214
PMID:32808934

Review

Improving Patients' Medication Adherence and Outcomes in Nonhospital Settings Through eHealth: Systematic Review of Randomized Controlled Trials

Zoie SY Wong[1], PhD; Braylien Siy[1], MD, MPH; Katharina Da Silva Lopes[1], MSc, MPH, PhD; Andrew Georgiou[2], PhD

[1]Graduate School of Public Health, St. Luke's International University, Tokyo, Japan

[2]Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

Corresponding Author:
Zoie SY Wong, PhD
Graduate School of Public Health
St. Luke's International University
5/F, 3-6-2 Tsukiji, Chuo-ku
Tokyo, 104-0045
Japan
Phone: 81 07041295996
Email: zoiesywong@gmail.com

Abstract

Background: Electronic health (eHealth) refers to the use of information and communication technologies for health. It plays an increasingly important role in patients' medication management.

Objective: To assess evidence on (1) whether eHealth for patients' medication management can improve drug adherence and health outcomes in nonhospital settings and (2) which eHealth functions are commonly used and are effective in improving drug adherence.

Methods: We searched for randomized controlled trials (RCTs) on PubMed, MEDLINE, CINAHL, EMBASE, EmCare, ProQuest, Scopus, Web of Science, ScienceDirect, and IEEE Xplore, in addition to other published sources between 2000 and 2018. We evaluated the studies against the primary outcome measure of medication adherence and multiple secondary health care outcome measures relating to adverse events, quality of life, patient satisfaction, and health expenditure. The quality of the studies included was assessed using the Cochrane Collaboration's Risk of Bias (RoB) tool.

Results: Our initial search yielded 9909 records, and 24 studies met the selection criteria. Of these, 13 indicated improvement in medication adherence at the significance level of $P<.1$ and 2 indicated an improved quality of life at the significance level of $P<.01$. The top 3 functions that were employed included mechanisms to communicate with caregivers, monitoring health features, and reminders and alerts. eHealth functions of providing information and education, and dispensing treatment and administration support tended to favor improved medication adherence outcomes (Fisher exact test, $P=.02$). There were differences in the characteristics of the study population, intervention design, functionality provided, reporting adherence, and outcome measures among the included studies. RoB assessment items, including blinding of outcome assessment and method for allocation concealment, were not explicitly reported in a large number of studies.

Conclusions: All the studies included were designed for patient home-based care application and provided a mechanism to communicate with caregivers. A targeted study population such as older patients should be considered to maximize the public health impact on the self-management of diseases.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO) CRD42018096627; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=96627

(J Med Internet Res 2020;22(8):e17015) doi:10.2196/17015

KEYWORDS

eHealth; self-administered drug; self-management; medication adherence; nonhospital settings; randomized controlled trial

Introduction

Integrating electronic health (eHealth) into medication prescription, dispensing, and administration processes is a promising step in the direction of achieving better medication safety, treatment, and health outcomes [1]. Health technology that supports patients' medication management can be integrated into different media including mobile health (mHealth) [2], telehealth [3], SMS, and wearable devices [1]. Offering a range of functionalities such as remote consultation and monitoring essential health indicators [1], eHealth plays essential roles in informing, educating, connecting, monitoring, and motivating patients [4].

Noncompliance with medication regimes on the part of patients is common, particularly among those who suffer from chronic diseases such as diabetes, hypertension [5], and cardiovascular conditions [6]. Failure to adhere to medication regimes can lead to poor health outcomes and increased health care costs [7]. Studying improvements in medication adherence has become an important area of focus in eHealth [6,8]. While evaluating eHealth, it is also common to consider several essential health care outcomes [9-12]. The recent literature has examined the impact of eHealth on patient safety [9,13], quality of life [14], and satisfaction [15] as well as health care spending [16].

This study is a systematic review that investigates how eHealth impacts the outcomes of patients' self-medication management. Based on the definition offered by the World Health Organization, this study defines eHealth [17] as referring to advancements in information and communication technologies that support care delivery and patient health management [18]. Instead of prescribing electronic medication in hospitals [9,13], we are interested in how eHealth contributes to the change in medication-taking behaviors in the nonacute disease management and recovery phases. We focus on drug-taking events in nonhospital and nonacute settings. These settings include home care, long-term care for older people, rehabilitation care, and outpatient facilities [19,20].

Prior studies have incorporated various methods of evaluation, such as rating systems and scales, user testing, and content analysis, to assess eHealth targeting for medication adherence [21]. This study focuses only on randomized controlled trials (RCTs), as this approach is the gold standard for evaluating digital medical intervention studies [22,23]. Drawing upon the current body of RCT studies, this review aims to assess the best available evidence on how eHealth interventions for self-management of medication improves drug adherence and health care outcomes. At the same time it characterizes the eHealth functions that are most commonly incorporated and those that favor improved medication adherence. In doing so, the study will contribute to the design, application, and sustainable development of eHealth in patient self-medication management.

Methods

Literature Search

The search was carried out in August 2018. To ensure exhaustive search results, cross-sectional databases in the fields of medicine, nursing care, public health, science, engineering, and social science were covered. The following databases were searched from 2000 to August 2018: PubMed, MEDLINE, CINAHL, EMBASE, EmCare, ProQuest, Scopus, Web of Science, ScienceDirect, and IEEE Xplore. We also included the Cochrane Library and gray literature sources, including Google Scholar and Open Access Theses and Dissertations. The snowball method was used to manually search citations within the studies included. We also hand-searched all the RCTs from the Journal of Medical Internet Research (JMIR) journals. Owing to the variation in terminology used to describe the topic of interest, we employed a broad, inclusive search strategy that covered the concepts of medication administration, eHealth, and nonhospital settings. Considering the appropriate Medical Subject Headings (MeSH) terms, we developed a set of master search terms that were applied to electronic data sets (enclosed in the Multimedia Appendix 1).

The inclusion criteria are presented as follows. We included all RCT studies that examined the effect of an eHealth intervention involving medication administration in a nonhospital setting. The periods for intervention needed to be at least six months. eHealth covers a broad range of mHealth, digital health, telehealth, electronic messaging, and electronic reminder interventions. Purely telephone-based outreach was not considered as health technology. We included studies in which participants needed to take medication regularly under nonacute settings. Only studies that focused solely on oral drug administration were selected. If the study did not specify a medication route or the intervention was administrated through a variety of routes (eg, [24]), it was not included. Studies were evaluated only when they reported, either directly or indirectly, on drug medication adherence, health care outcomes in adverse drug events, patient satisfaction, costs, or quality of life. Cluster and pragmatic RCTs were also included. No limit was applied to the databases in terms of article language. Studies that focused on participants with mental health problems (including depression, stress disorder, psychosis, and schizophrenia), or those who may have suffered from complex psychological issues often associated with serious illnesses (such as HIV/AIDS), were excluded because of their potential to weaken the study population's representativeness and affect the generalizability of the results.

Two reviewers (one with a medical and public health background [BS] and another one with health technology and informatics background [ZSYW]) independently carried out title and abstract reviews in the screening phase. Both were well familiar with the aforementioned inclusion and exclusion criteria and evaluated the full texts of articles independently. Their results were compared, and Cohen κ was measured in the assessment of eligibility stage in order to evaluate the inter-rater reliability between the reviewers. For conflicting decisions, a consensus was arrived at between reviewers by discussing

rationales and concerns and reexamining how each compromised decision satisfied the principles of inclusion and exclusion as outlined. The Rayyan web application [25] was used to facilitate the double-blind evaluations and to maintain review records. We developed a data extraction sheet for the studies included (Multimedia Appendix 2) that complied with the minimum standards of the Data Extraction Template for Included Studies, as developed by the Cochrane Consumers and Communication Review Group [26].

The details recorded (wherever available) for each included paper were as follows: general review information, study population, study characteristics, outcome measures, and results. We reported the primary outcomes and all the available secondary ones. Both qualitative and quantitative materials were extracted. We were interested in the frequency of the measurement outcomes of the intervention group (health technology) when compared with control (usual care practice). Risk difference, which is one of the most useful ways to present RCT research results, was used as a quantitative reporting measure. The outcome measures with the largest possible intervention timeframe for results were reported. The study protocol was published in the PROSPERO registry on August 7, 2018 (Registration number: CRD42018096627).

Outcome Measures

Medication adherence [27] refers to the degree to which patients' medication-taking behavior accords with appropriate medical advice [28]; it was set as the primary outcome measure in this study. For the secondary outcome measures, we included indications of adverse event (or safety outcome), quality of life, patient satisfaction, and health expenditure/spending, as eHealth studies often assess these health care measures and considering them would allow us to evaluate eHealth impact more comprehensively.

Health Technology Functions

It is common for eHealth applications to have multiple features. Referring to [4] and the range of capabilities of drug application [21], we compiled a list of commonly used health technology functions, namely, mechanisms to communicate with caregivers, monitoring health features, reminding and alerting, providing information and education, dispensing treatment and administration support, personalized feedback, reporting and trending, dynamic treatment adjustment, social support, and setting goals and planning. Based on the eHealth intervention described in each study included, we tallied the occurrence of these functions to determine how often the functions were applied in eHealth interventions. Considering the presence of each health technology function and improved medication adherence at $P<.1$, $P<.01$, and $P<.001$ as variables, we constructed 2×2 contingency tables to examine whether the proportions for different health technology functions were different. Fisher exact test of independence was employed.

Assessment of Methodological Quality

Two reviewers (BS and ZSYW) also performed quality assessment for the studies included using the Cochrane Collaboration's Risk of Bias (RoB) assessment tool via RevMan version 5.3 software [29]. Following the RoB assessment guidelines [30], the reviewers assessed the RoB signaling questions as either *Low risk*, *High risk*, or *Unclear* RoB, based on the evidence accessible from each of the studies included. A third-party opinion was sought from another coauthor (KDSL), as and when needed.

Results

Literature Search

Figure 1 presents the flow diagram of the literature search (JMIR hand-searched result is appended in Multimedia Appendix 3). After removing duplicates, our initial search yielded 9909 records, of which 92 were reviewed for full-text assessment and 24 satisfied the inclusion criteria and were included in this review. Cohen κ between the reviewers was 0.846, which is equivalent to a strong level of agreement.

Figure 1. Flow diagram. RCT: randomized controlled study.

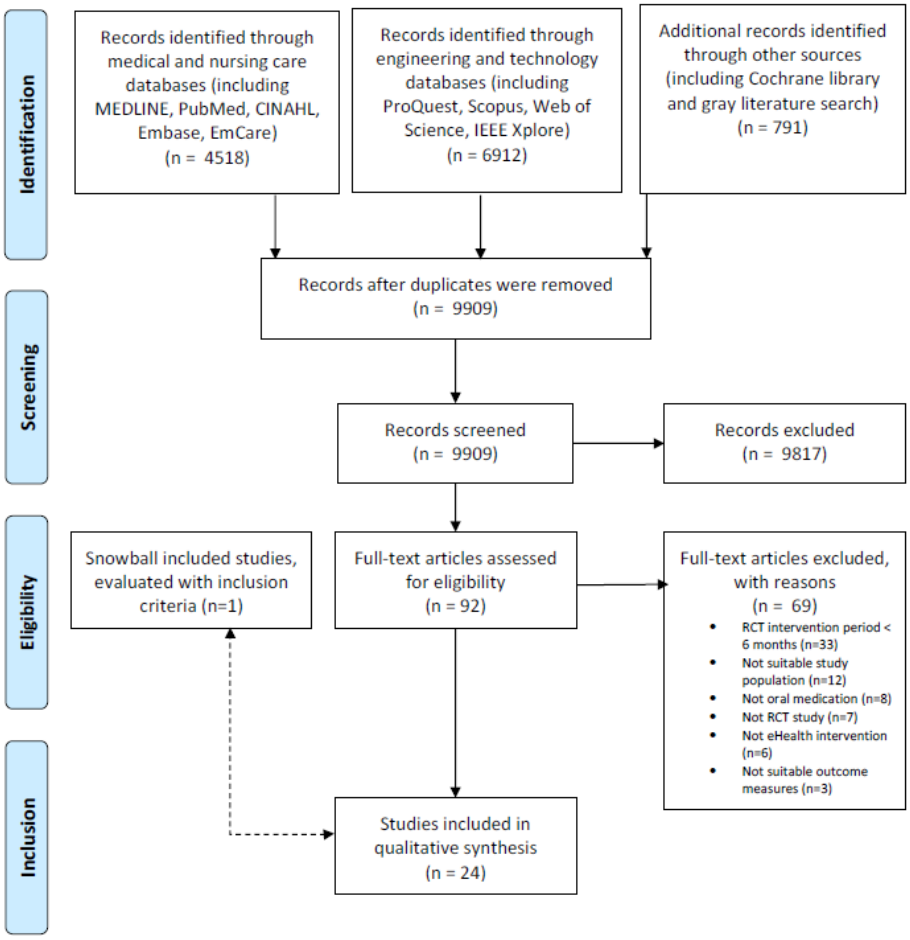


Table 1 illustrates the characteristics of the studies included. All the studies included were published after 2008. Of the 24 studies included, 11 (46%) [2,31-40] were published in the last 5 years (since 2014) and 20 (83%) [2,3,31-48] were published in the last 10 years (since 2009). All the studies included were published in English. As many as 10 [2,3,33-35,39,42,45,46,49] out of 24 studies were undertaken in the United States and 8 [31,36,40,41,47,50-52] were carried out in European countries. Other countries studied included Canada [43,44], South Africa [32], Iran [37], South Korea [38], and Taiwan [48]. Further, 19 [2,3,31,32,34,35,37-41,43-46,48,49,51,52] studies focused on the adult population, and only 3 [33,42,47] specifically targeted older patients (aged 65 years and older). All the health technology interventions were primarily designed for home-based usage. The relevant clinical conditions covered were asthma [41,48,50], heart disease [31,34,39,40,43,45,47,49], hypertension [32,35,37,44,46,51,52], diabetes mellitus [38], post-transplantation (renal [36] and lung [2]), comorbidities of diabetes mellitus and hypertension [3], and other unspecified diseases [33,42]. Eight studies [2,3,32,38,41,43,45,46] reported incorporation of clinical decision support or advanced computational algorithms to aid the self-management of diseases, while five [2,3,32,43,45] offered technology-led (instead of caregiver-led) dynamic feedback tailored to suit the patients' conditions.

Table 1. Study characteristics.

Author, year of publication	Description of health technology intervention (intervention group size)
Jerant et al. 2003 [49]	Video-based Telecare: Aviva personal telecare unit installed at home that allows real-time videoconferencing with nurse caregiver, equipped with electronic stethoscope for lung auscultation (N=13).
DeVito Dabbs et al. 2016 [2]	Pocket PATH with a smartphone platform: Custom programs allow patient input of daily measurements (spirometry, vital signs, symptoms). Also includes decision-support feature that automatically sends reminders to patient, and to call the transplant coordinator, whenever measures reached immediate report level (N=99).
Hashimoto et al. 2011 [41]	Internet-based management tool: Included an electronic diary and treatment decision support (dose adjustment of oral corticosteroids) for patients (N=51).
Marek et al. 2013 [42]	MD.2: Medication-dispensing machine that stores and preloaded 60 plastic reusable cups in a locked compartment. Generate online compliance reports to monitor missed doses. Nurse care coordinated with physician(s) and pharmacist(s), visiting the participants at least every two weeks and performing care plan activities (N=152).
Willems et al. 2008 [50]	Electronic asthma monitor: Portable handheld device with a matching modem that can register lung function values and symptoms on the monitor (N=55).
Boyne et al. 2014 [31]	Health Buddy (telemonitoring device): Equipped with a liquid crystal display and 4 keys connected to a landline phone. Patients received daily preset dialogues and questions about their symptoms, knowledge, and behavior, which had to be answered by touching the keys. Subsequently the answers were transmitted to the nurses' desktop (N=197).
Sherrard et al. 2009 [43]	Interactive voice response (IVR): Developed an algorithm of 11 questions addressing medication compliance, reporting of adverse events, providing information on common medications, and offering general medication safety tips. The IVR system recorded patients' voiced responses (yes or no) into a central database (N=164).
Bobrow et al. 2016 [32]	Personalized short messaging service text messages were sent to (1) information-only message (N=457) and (2) interactive message (N=458) group participants at weekly intervals, at a time and in a language selected by the participant. Messages focused on the techniques of goals and planning, repetition and substitution, social support, and natural consequences.
Marek et al. 2014 [33]	Medication-dispensing machine + nurse care coordination (every 2 weeks), preloaded with medications in reusable plastic cups. (N=150).
Volpp et al. 2017 [34]	Vitality GlowCaps: 4 electronic pill bottles used for cardiovascular medications (including β-blockers, statins, aspirin, antiplatelet agents), which electronically monitored openings. Transmitted information to health organizations (N=682).
Kim et al. 2016 [35]	Withings Blood Pressure Monitor with iPhone with apps: Provided portals and a dashboard to link with families, caregivers, and health care professionals. Equipped with an online disease management program featuring educational materials (N=52).
Rinfret et al. 2009 [44]	IT-supported program: Consisted of educational booklet, digital home blood pressure monitor, logbook, and access to a telephone-linked management program. The system collected self-recorded blood pressure and self-assessed adherence data and integrated these data with actual pharmacy medication refill. Able to generate reports (N=111).
Santschi et al. 2008 [51]	Participants received drug with electronic monitoring devices: Medication Event Monitoring System (MEMS, AARDEX Ltd) used to obtain accurate, detailed dynamic, and *real-time* information on the patients' medication-taking behavior (N=34).
Stacy et al. 2009 [45]	IVR system: Provided three separate tailored behavioral support interactions, coupled with tailored feedback based on parents' cholesterol-related knowledge, attitudes, beliefs, and perceived barriers to medication adherence (N=253).
Bosworth et al. 2011 [46]	All intervention groups utilized wireless home blood pressure monitor (automatically transmitted) and telemedicine device—connected to a telephone line like an answering machine. (1) Behavioral management—nurse-administered encounter via software platform to provide health behavior modules focusing on hypertension self-management improvement (N=147); (2) medication management—triggers sent to physician and nurse to adjust medication dynamically with decision support, with nurse follow-up call every 3 weeks (N=149); (3) a combination of A and B (N=147).
Dusing et al. 2009 [52]	A set of medication supportive measures: Offered support to both physician and patient. Patient received 24-hour timer, reminding stickers, information brochures, and home blood measurement device. Electronic MEMS utilized (N=97).
Henriksson et al. 2015 [36]	Electronic Monitoring Drug Dispensing Device: The patients loaded the device with a week's worth of medication at a time. The device generated visual and audible signals. If the patient did not take their medication, the audible signal repeated with increasing frequency for 120 minutes. After this (or after the medication was taken), the device sent an SMS text message to the web-based software, thus registering patient compliance information (N=40).
Hosseininasab et al. 2014 [37]	Wrist self-monitoring device: A blood pressure measurement device with log-book documentation (N=97).

Author, year of publication	Description of health technology intervention (intervention group size)
Jeong et al. 2018 [38]	Patient in all groups used a Smart Care Unit (SCU), which consists of a web-enabled computer with camera (for videoconferencing and communication with caregiver), specific software, glucometer (blood glucose monitoring), and body composition organizer (for body weight measure, tracking diet, and exercise record). Other functions included automated short message feedback and access to care center education program. (1) Telemonitoring group: face-to-face outpatient hospital visit scheduled with caregiver at 8, 16, and 24 weeks. Medication was prescribed based on SCU data and caregiver received advice from clinical decision-support system (N=113); (2) Telemedicine group: in weeks 8 and 16, patients contacted physicians via the SCU, and in week 24, a face-to-face visit was scheduled (N=112).
Kooy et al. 2013 [47]	Electronic reminder device (ERD): Medication reminder device that beeped every day at the same time until the patient switched it off. Patients could adjust the beeping time. (1) Counseling with an ERD (N=130). (2) ERD with written instructions (N=123).
Liu et al. 2011 [48]	Mobile telephone-based interactive self-care system: Provided an electronic diary to record patients' daily asthma symptom score (including sleep quality, coughing severity, difficulty breathing, and daily activities affected by asthma), use of relievers, peak expiratory flow rate (PEFR), and PEFR variability (N=43).
Wakefield et al. 2012 [3]	Home telemonitoring device: Employed standard telephone line to transmit data between patient and study center. Patients in all groups manually entered blood pressure and blood glucose measures. (1) High-intensity group received health information tips and questions from the branching algorithm (N=93). (2) Low-intensity group: Did not receive the informational tips and questions from the algorithm (N=102).
Young et al. 2016 [39]	One-on-one in-hospital self-management training + telephone-based postdischarge reinforcement sessions—scheduled twice a week in the first 2 weeks, once a week in weeks 3-6, and every other week in weeks 7-12. Intervention content presented in verbal, written, visual formats with interactive ability; self-management workbooks and self-management toolkit (calendar for weight and salt daily logging), weight scale with large and bright readings, and an electronic pill organizer reminder alarm are provided. Session lasted for 45-50 minutes. Booster sessions were delivered to those struggling with self-management at home. Tailored intervention sessions were provided based on level of activation, predefined goals, and specific self-management needs (N=51).
Wald et al. 2014 [40]	SMS text messaging group: Sent daily texts for 2 weeks, and alternate-day texts for 2 weeks. Subsequently sent weekly texts for 22 weeks (6 months in all). Participants were requested to reply to each message to indicate if they had taken their medication or not and if the message reminded them to take medication. Computer sends the text message based on the schedule. Patients responses were filed and if not taking medicine, telephone follow-up was made (N=151).

Outcome Measures: Medication Adherence

Table 2 presents the definitions of different medication adherence measures and summarizes each of these measures as reported by the studies included. Among 19 studies [2,3,31,32,34-40,42-45,47,49,51,52] that evaluated and reported the impact of health technology on medication adherence, 5 [3,31,32,39,43] adopted questionnaires and scales, which are the most commonly used measures. Further, 15 [2,3,31,32,34,37-40,43-45,47,49,52] explicitly compared improvements in medication adherence as a result of the intervention with the control arms. As many as 12 [2,3,31,32,37-40,43-45,52] reported that the intervention arms had seen improvements in medication adherence at the significance level of $P<.1$. A total of 4 studies [32,38,40,43] showed significant improvement in medication adherence at the $P<.01$ level and 3 studies [32,38,43] were significant at the $P<.001$ level. These results indicate that eHealth can improve patients' medication adherence in nonhospital settings. Multimedia Appendix 2 provides details of the synthesized outcomes, risk differences, and P values between intervention and control.

Outcome Measures: Health Care Outcome Measures

Table 2 also presents the definitions of the secondary outcome measures and summarizes each of them as reported by the studies included. In all, 11 studies [32-34,38,41-43,46,48-50] reported secondary outcome measures including those that are associated with adverse events [38,43], quality of life [41,42,48,50], patient satisfaction [32,41,43,49], and health expenditure [33,34,46]. Among the 2 studies [38,43] reporting adverse events relating to the interventions (or safety outcome), no statistically meaningful difference between the intervention and control was found. Two [48,50] of the 4 studies reporting quality of life of the patients [41,42,48,50] revealed that there were significant differences between the intervention and control arms at the significance level of $P<.01$. Among the 4 studies measuring patient satisfaction [32,41,43,49], 1 [41] observed a difference when compared with the control. Three studies (all conducted in the United States) [34,42,46] reported medical spending obtained from various sources, including insurance claims [33,34] and inpatient and outpatient costs [46], and none of these showed any significant difference between intervention and control. The above evidence indicates that eHealth for self-administration of medication can improve the quality of life of patients.

Table 2. Summary of outcome measures.

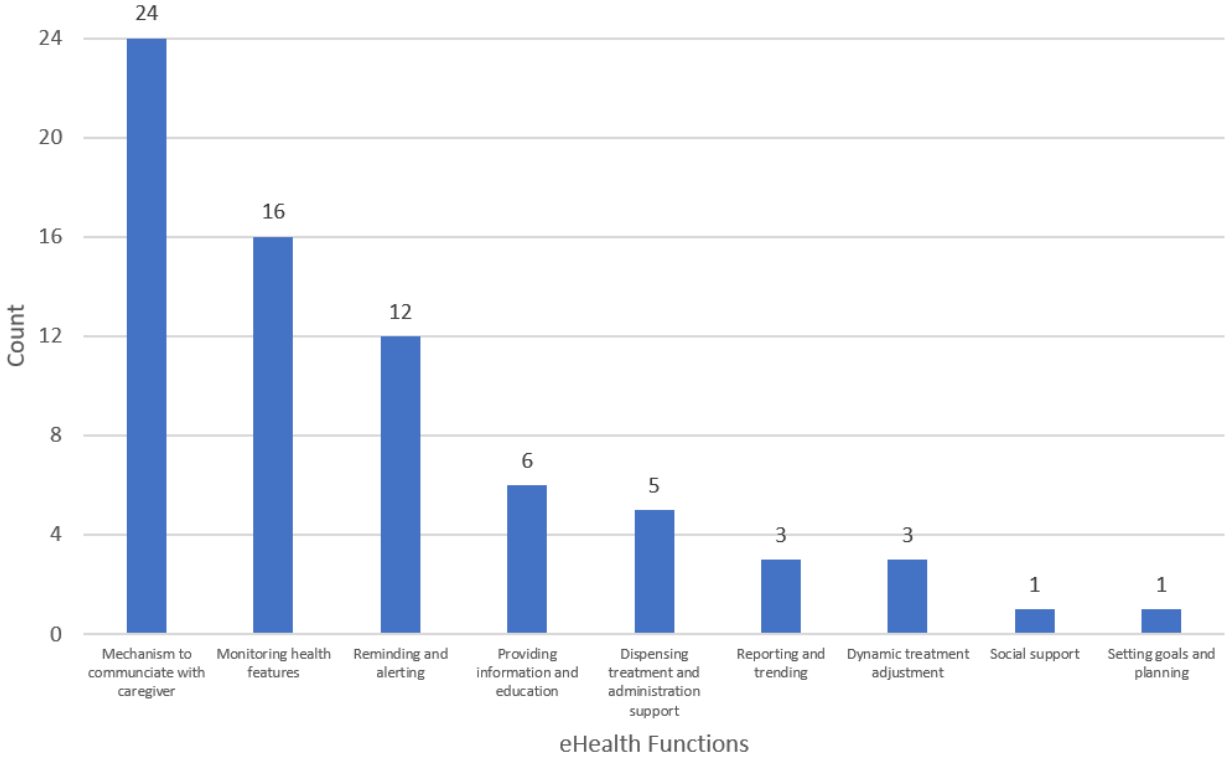
Outcome measures	Definition	Reference(s)
Medication adherence (N=19)		
Continuous, multiple-interval measures of medication acquisition (CMA)	CMA is the sum of days of medication supply obtained divided by the total number of days of study participation [27,53,54].	[44]
Proportion of days medication covered (PDC)	PDC measures the persistence to the medication therapy by calculating the total days' supply divided by the number of days of study participation. The value is capped at 100% [27]. It is a common proxy-measure of adherence.	[32,34,45,47]
Continuous measure of medication gaps	This measure refers to the total number of treatment gap days divided by the duration of the time period of interest. It indicates the proportion of time for which patients do not have drug exposure [53].	[44]
Medication possession ratio (MPR)	MPR is the proportion of days' medication supply obtained over either refill interval or fixed refill [53]. MPR is calculated for the individual patient and can create different denominators.	[45]
Eight-item Morisky Medication Adherence Scale (MMAS-8)	MMAS-8 is a validated medication adherence scale that contains 7 Yes/No responses and a 5-point Likert response [53]. The scale captures patients' medication-taking behavior and barriers to adherence.	[35,37]
Pill count	Pill count is the number of consumed pills divided by the number of total prescribed pills [53].	[37]
Measurement cutoff	This measure requires setting an arbitrary cutoff value to a continuous measure for identifying adherence and nonadherence into dichotomous outcomes [53,55]. It is typically less sensitive than the original measure [55]. For instance, taking medication with 80% cutoff of a 28-day medication cycle, or PDC with 80% or above.	[2,32,40,49]
Measures involving electronic medication devices	Electronic medication devices aim to record adherence performance for analysis. Typical features include recording dosing events, audio/visual reminders, electronic displays, and monitoring and feedback on adherence performance. However, not all features are available in all devices. In many medication adherence studies, the MEMS is commonly used [53].	[36,40,42,51,52]
Other questionnaires and scales	These questionnaires and scales are generally validated against other conditions and related measures to assess medication regime conditions for a broad range of diseased populations [53] or specific ones. Heart failure compliance scale and validated diabetes mellitus regimen adherence scale (Edwards Scale) are some examples. Self-report questionnaires are also included.	[3,31,32,39,43]
Unspecified	Medication adherence measure was not specified.	[38]
Health care outcome measures (N=11)		
Adverse event	Measures refer to the number of emergency visits or instances of hospitalization or untoward medical occurrence.	[38,43]
Quality of life	Various measures may apply, including Short Form-36/Short Form-12 Physical Component Scale and Mental Component Scale, Asthma-Related Quality of Life, and (Pediatric) Asthma Quality of Life Questionnaire.	[41,42,48,50]
Patient satisfaction	This measure refers to how well eHealth met patient expectations. Various measures may apply, including the 8-item Client Satisfaction Questionnaire and the patient satisfaction survey.	[32,41,43,49]
Health spending	Various measures may apply, including assessing monthly or yearly expenditure via claims data, and medical cost computed by direct and indirect outpatient and inpatient cost items.	[33,34,46]

Health Technology Functions

Figure 2 presents the functionalities in the studies included by inclusion frequency. It appears that all the studies included incorporated an established mechanism to communicate with caregivers. Monitoring health features [2,3,35-39,41,43,44,46,48-52], and reminding and alerting [2,31,33,35,36,39,40,42,46,47,51,52] are the second- and third-most commonly used functions, respectively. Of the 24 studies, 6 [3,32,35,38,43,52] used the function of providing information and education to users and five studies [33,34,36,39,51] incorporated dispensing treatment and administration support, such as pill organizer automatic opening devices [34]. Providing information and education and dispensing treatment and administration support are significantly different in the medication adherence (at $P<.001$ level) group (Fisher exact test, $P=.02$). This indicates that these health

technology functions are effective in improving medication adherence (Details provided in Multimedia Appendix 4).

Figure 2. Functionalities among the studies included.



Assessment of Methodological Quality

Figures 3 and 4 present a summary of the RoB assessment and judgments of each assessment domain for the studies included. Of the 24 studies, 12 [2,3,31-34,36,39-42,50] reported on the employment of random sequence generation methods that sufficiently produced comparable study arms, while 6 [2,3,36,39,43,49] utilized methods to conceal the allocation sequences before or during enrollment. These studies exhibited a low risk of selection bias. All the studies included had a high risk of performance bias, which refers to the lack of blinding of participants. It was found that some studies [2,32,34,37] managed to introduce blinding to the investigator or personnel engaged in the research. In terms of detection bias, only 2 studies [2,39] were classified as having low bias with an indication of postallocation assessors blinding and 20 studies were assessed as unclear due to unspecified information. The attrition rates were generally low among the studies and only 4 [37,44,45,48] were grouped as high risk in this domain. A total of 9 studies had high selective reporting bias owing to the limitation of the chosen outcome measures [31,39,45,49] or the design [37,42,47], or the inability to measure the control group's outcomes [36,51], etc. Apart from the above, there were a large number of studies that did not provide sufficient details on how to handle various domains of RoB in the publications and that inevitably increased the challenges involved in assessing various biases in the RCT studies included.

Figure 3. Summary of the quality of the studies included.

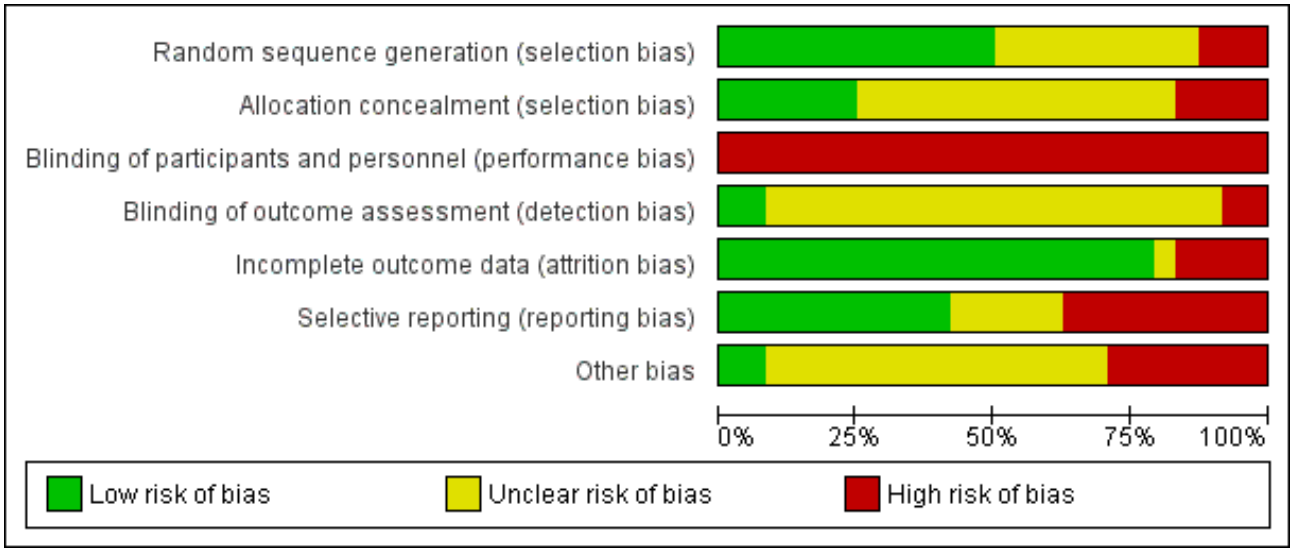
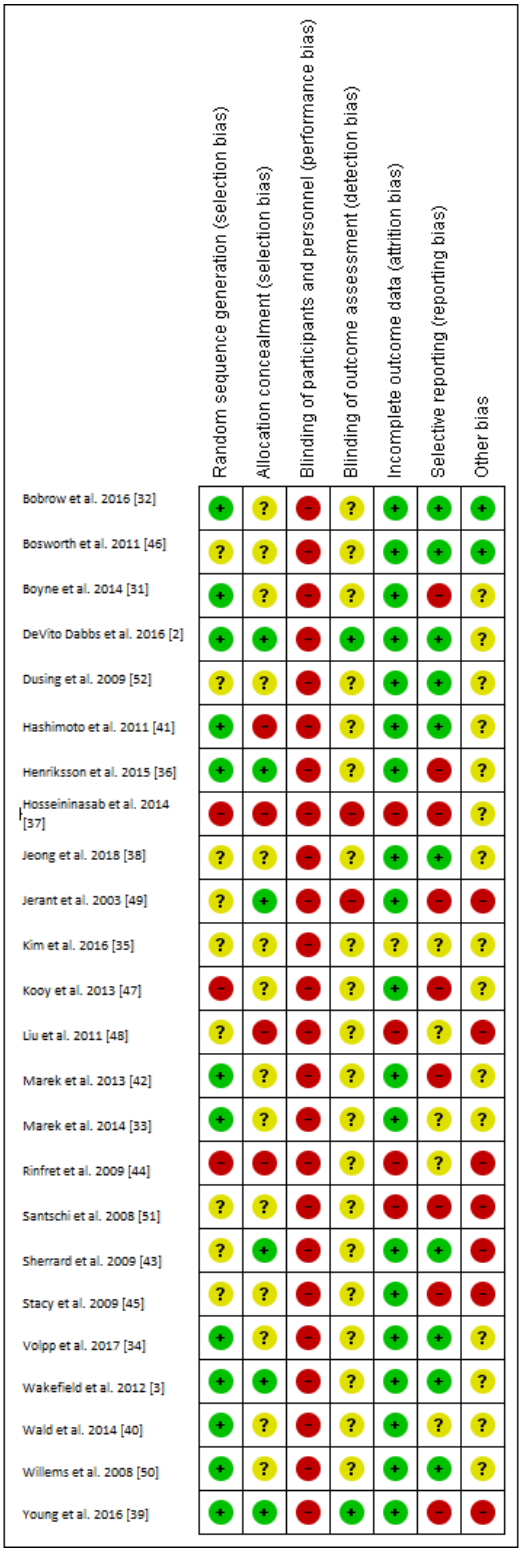


Figure 4. Risk of Bias (RoB) assessment of the studies included.



Discussion

This systematic review synthesized the existing evidence on the impact of eHealth interventions on medication adherence and selected health care outcome measures in nonhospital settings. There is evidence proving that eHealth improves patients' medication adherence and quality of life. The most frequently used functions are the mechanism of communication with caregivers, monitoring of health features, and reminding and alerting. Further, providing information and education, and dispensing treatment and administration support are most favorable when it comes to improved medication adherence outcomes.

Outcome Measures

All the studies included were designed for patient home-based care applications and chronic conditions. Chronic disease management and home health care are important health issues

in the context of a rapidly aging population. Only 3 [33,42,47] of the studies included were targeted at older patients aged 65 or above, and their findings were either insignificant [33,47] or inconclusive (no control group outcome was measured) [42]. Furthermore, the conclusions drawn from interventions in the adult population may lack generalizability in explaining the impact upon an older population. This is due to potential divergence in acceptance of technology and ability to use technology-assisted services between the older population and the general adult population. To address this issue, future studies may consider specifically designing and evaluating patient-centered eHealth interventions that cater to the special needs of older patients.

All the studies integrated more than 2 functions into the health technology intervention, indicating that multimodal intervention tends to be a commonly adopted model for eHealth design. We found that studies with significant improvements in medication adherence were all equipped with 3 or more integrated functions, except [37]. For instance, Jeong et al. [38] presented a multifunctional eHealth intervention comprising heterogeneous functions such as web-enabled videoconferencing that helped connect with the caregiver, a blood glucose monitoring device, a body composition organizer, an automated short message feedback system, and access to care center education program, which achieved significant medication adherence improvement (P<.001). This finding was in line with the non-eHealth medication nonadherence studies that showed that multiple components incorporated into the intervention tend to be successful [6]. This may be because medication nonadherence is usually multifactorial [6,53] and is a complex behavioral issue that involves socioeconomic and therapy-, patient-, condition-, and health system/health care team-related factors [56]. Single-function interventions thus tend to be insufficient when it comes to tackling such a complex problem

All the studies included provided a common eHealth function, that is, a mechanism to communicate with the caregivers. This allowed regular patient status updates, sending alerts, remote coaching, and interactive treatment plan adjustment. Considering the sociotechnical aspect of health informatics applications, future studies should carefully consider the interplay between the health system and eHealth interventions to tackle patient self-management of diseases effectively.

Our results indicated that functions of providing information and education, and dispensing treatment and administration support tend to favor improved medication adherence outcomes. This may be because dispensing treatment and administration support offer a mechanism to regulate and track medication activities [57], and to support a smooth self-medication process. It is also important to educate patients so that they understand the related disease characteristics and the benefits of following the medication regime, through the function of providing information and education.

The Quality of the RCT Studies Included

The results should be interpreted with caution. Many studies that we included did not explicitly report RoB assessment items, particularly in the domains of blinding of outcome assessment (83% [20/24] unclear risk) and method for allocation concealment (58% [14/24] unclear risk). This increased the difficulties in evaluating the quality of the RCTs in full in order to make comparisons and draw appropriate conclusions. DeVito Dabbs et al. [2] and Young et al. [39] are relatively high-quality RCTs that evince minimal risk in selection, detection, and attrition biases.

Owing to the disparities in population characteristics, intervention design, functionality, and reporting adherence measures, it is difficult to draw inferences and definitive conclusions on some potential hypotheses, such as the combination of eHealth functions that contribute to improvements in medication adherence for a particular disease. Evidence on the effectiveness of eHealth on medication adherence and health care outcomes improvement exists, but is not compelling enough. Larger-scale RCTs with greater sample sizes are necessary to draw inferences among those infrequently observed measures such as safety outcomes in adverse events, hospitalization, and emergency visits.

Subjective and objective measures have their pros and cons [53]. To increase measurement sensitivity, future studies can consider employing a combination of measures to assess medication adherence. When designing eHealth interventions, it is important to be aware of how the selection of outcome measures can contribute to the trustworthiness of a study. The choice of medication adherence measures can affect who is assessing the outcome and the objectivity of the assessment. In general, where subjective outcomes are concerned, blinding is particularly important. For instance, medication adherence via questionnaire survey can be subjectively assessed by the participants, which is considered as low-quality adherence measurement. Such a design can influence the blinding of outcome assessment (ie, increasing detection bias).

Strengths and Limitations

This study attempted to examine how eHealth for patients' medication management improves drug adherence and other health care outcome measures. Our study has a number of strengths. First, we comprehensively searched cross-sectional databases in the fields of medicine, nursing care, public health, science, engineering, and social science. We focused exclusively on studies that employed RCT, which is considered the highest standard of eHealth evaluation. Our search strategy was broad and involved a large number of studies for screening (eg, 9909). We successfully retrieved a solid body of evidence that indicated improvement in drug adherence through the application of eHealth. We also evaluated the quality of the studies included through the state-of-the-art Cochrane Collaboration tool. Our study provides practical insights into the future of eHealth design and applications for patient self-medication management.

Our study has a few limitations, as well. First, a publication bias may exist owing to the inherent tendency to publish *positive* results with significant findings. Furthermore, the studies included reported multiple outcome measures rather differently. Owing to this heterogeneity, it was not possible to carry out a meta-analysis. We synthesized review outcomes from studies that largely centered on chronic diseases. However, our search strategy was designed to consider a broad scope of studies. Based on our study framework, future studies can investigate

how eHealth impacts a specific targeted disease or patient group (such as those who suffer from mental health issues or HIV/AIDS).

Furthermore, owing to the small sample size, we were unable to use more accurate parametric methods, and thus unable to ascertain the impact of a combination of multiple functions. Large-scale evaluation studies in the future can consider examining the impact of the integration of multiple functionalities into eHealth; studying the social and behavioral differences across targeted populations in reacting to health technology; and carefully evaluating other important health outcome measures including adverse events owing to health technology introduction, and cost and economic evaluation.

Understanding the nature of the eHealth intervention, one must be aware that some limitations in the study design are inevitable. For instance, all the studies included were unable to blind the intervention group from the control group (ie, the usual care group without using technology intervention). At present, there is no way for researchers to avoid performance bias in the studies reviewed. However, this situation may change in the future when successful health technology becomes the gold standard and replaces the current usual care system which does not employ health technology. Future RCTs for evaluating eHealth interventions should follow best practice guidelines, which include observing the limitations of the study design; blinding of participants, personnel, and assessors involved; selecting the most objective outcome measures; employing fair assessment methods; and avoiding selective outcome reporting.

Conclusion

This study investigated how eHealth interventions could affect patient medication adherence and health outcomes and identified the eHealth functions that are most effective in improving medication adherence. The evidence reviewed shows that eHealth can improve patients' medication adherence and quality of life in nonhospital settings. Integrating multiple functions into health technology tends to be effective in achieving enhanced medication adherence. eHealth functions of providing information and education, and dispensing treatment and administration favored an improved medication adherence outcome. However, the literature base remains small, diffuse, and inconclusive at this time. Medication-taking behavior may vary tremendously based on the patients' medical conditions, the population studied, and the specific medications assessed. Many interesting potential medical–social–behavioral research hypotheses are yet to be posed and answered in the existing literature.

Acknowledgments

This study was supported by the Japan Society for the Promotion of Science KAKENHI (Grant No. 18H03336) and Daiwa Securities Health Foundation (2017-2018).

Authors' Contributions

The study was conceived by ZSYW. ZSYW, AG, and KDSL contributed to the study design. Research data were retrieved and interpreted by ZSYW and BS. ZSYW led the writing of the paper and all authors revised and refined the arguments. All authors approved the article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms for the systematic review.
[DOCX File , 17 KB - jmir_v22i8e17015_app1.docx]

Multimedia Appendix 2
Details of the included studies.
[XLSX File (Microsoft Excel File), 27 KB - jmir_v22i8e17015_app2.xlsx]

Multimedia Appendix 3
JMIR hand-search flow diagram.
[DOC File , 35 KB - jmir_v22i8e17015_app3.doc]

Multimedia Appendix 4
Fisher's exact test results.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v22i8e17015_app4.xlsx]

References

1. Car J, Tan WS, Huang Z, Sloot P, Franklin BD. eHealth in the future of medications management: personalisation, monitoring and adherence. BMC Med 2017 Apr 05;15(1):73 [FREE Full text] [doi: 10.1186/s12916-017-0838-0] [Medline: 28376771]
2. DeVito Dabbs A, Song MK, Myers BA, Li R, Hawkins RP, Pilewski JM, et al. A Randomized Controlled Trial of a Mobile Health Intervention to Promote Self-Management After Lung Transplantation. Am J Transplant 2016 Jul;16(7):2172-2180 [FREE Full text] [doi: 10.1111/ajt.13701] [Medline: 26729617]
3. Wakefield BJ, Holman JE, Ray A, Scherubel M, Adams MR, Hills SL, et al. Outcomes of a home telehealth intervention for patients with diabetes and hypertension. Telemed J E Health 2012 Oct;18(8):575-579. [doi: 10.1089/tmj.2011.0237] [Medline: 22873700]
4. Smahel D, Elavsky S, Machackova H. Functions of mHealth applications: A user's perspective. Health Informatics J 2017 Nov 01:1460458217740725. [doi: 10.1177/1460458217740725] [Medline: 29121831]
5. Kleinsinger F. The Unmet Challenge of Medication Nonadherence. permj 2018;22:18-033. [doi: 10.7812/tpp/18-033]
6. Kleinsinger F. The Unmet Challenge of Medication Nonadherence. Perm J 2018;22:18-033 [FREE Full text] [doi: 10.7812/TPP/18-033] [Medline: 30005722]
7. Iuga AO, McGuire MJ. Adherence and health care costs. Risk Manag Healthc Policy 2014;7:35-44 [FREE Full text] [doi: 10.2147/RMHP.S19801] [Medline: 24591853]
8. Bosworth HB, Zullig LL, Mendys P, Ho M, Trygstad T, Granger C, et al. Health Information Technology: Meaningful Use and Next Steps to Improving Electronic Facilitation of Medication Adherence. JMIR Med Inform 2016;4(1):e9 [FREE Full text] [doi: 10.2196/medinform.4326] [Medline: 26980270]
9. Reckmann MH, Westbrook JI, Koh Y, Lo C, Day RO. Does computerized provider order entry reduce prescribing errors for hospital inpatients? A systematic review. J Am Med Inform Assoc 2009;16(5):613-623 [FREE Full text] [doi: 10.1197/jamia.M3050] [Medline: 19567798]
10. Georgiou A, Prgomet M, Markewycz A, Adams E, Westbrook JI. The impact of computerized provider order entry systems on medical-imaging services: a systematic review. J Am Med Inform Assoc 2011 May 01;18(3):335-340 [FREE Full text] [doi: 10.1136/amiajnl-2010-000043] [Medline: 21385821]
11. Georgiou A, Williamson M, Westbrook JI, Ray S. The impact of computerised physician order entry systems on pathology services: a systematic review. Int J Med Inform 2007 Jul;76(7):514-529. [doi: 10.1016/j.ijmedinf.2006.02.004] [Medline: 16567121]
12. Bates DW. Using information technology to reduce rates of medication errors in hospitals. BMJ 2000 Mar 18;320(7237):788-791 [FREE Full text] [doi: 10.1136/bmj.320.7237.788] [Medline: 10720369]
13. Ammenwerth E, Schnell-Inderst P, Machan C, Siebert U. The effect of electronic prescribing on medication errors and adverse drug events: a systematic review. J Am Med Inform Assoc 2008;15(5):585-600 [FREE Full text] [doi: 10.1197/jamia.M2667] [Medline: 18579832]
14. Triberti S, Savioni L, Sebri V, Pravettoni G. eHealth for improving quality of life in breast cancer patients: A systematic review. Cancer Treat Rev 2019 Mar;74:1-14. [doi: 10.1016/j.ctrv.2019.01.003] [Medline: 30658289]
15. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
16. Holtkamp MD. Decrease Hospital Spending: There's an App for That! A Retrospective Analysis of Implementation of a Mobile Resident Handbook on Hospital Costs and Disposition. Telemed J E Health 2017 Oct;23(10):828-832. [doi: 10.1089/tmj.2016.0257] [Medline: 28488949]
17. World Health Organization. eHealth. URL: https://www.who.int/ehealth/en/ [accessed 2020-08-05]
18. World Health Organization. eHealth. World Health Organization. 2019. URL: https://www.who.int/ehealth/en/ [accessed 2020-08-05]
19. Jarvis WR. Infection control and changing health-care delivery systems. Emerg Infect Dis 2001;7(2):170-173 [FREE Full text] [doi: 10.3201/eid0702.010202] [Medline: 11294699]
20. Moro M. Infection Control in Non-Hospital Settings: The Example of Long-Term and Ambulatory Care Facilities. 2013. URL: https://www.who.int/gpsc/information_centre/moro-maria-luisa_infection-control.pdf [accessed 2020-08-05]
21. Park JYE, Li J, Howren A, Tsao NW, De Vera M. Mobile Phone Apps Targeting Medication Adherence: Quality Assessment and Content Analysis of User Reviews. JMIR Mhealth Uhealth 2019 Jan 31;7(1):e11919 [FREE Full text] [doi: 10.2196/11919] [Medline: 30702435]
22. Topol EJ. A decade of digital medicine innovation. Sci Transl Med 2019 Jun 26;11(498):eaaw7610. [doi: 10.1126/scitranslmed.aaw7610] [Medline: 31243153]
23. World Health Organization. Monitoring and Evaluating Digital Health Interventions: A Practical Guide to Conducting Research Assessment. Geneva, Switzerland: World Health Organization; 2016.
24. van den Wijngaart LS, Roukema J, Boehmer ALM, Brouwer ML, Hugen CAC, Niers LEM, et al. A virtual asthma clinic for children: fewer routine outpatient visits, same asthma control. Eur Respir J 2017 Oct;50(4):1700471 [FREE Full text] [doi: 10.1183/13993003.00471-2017] [Medline: 28982775]
25. Anonymous. Rayyan QCRI: Qatar Computing Research Institute (Data Analytics). URL: https://rayyan.qcri.org/welcome [accessed 2020-07-25]

26. Cochrane Consumers and Communication. Consumers and Communication Group Resources for Authors. 2019. URL: https://cccrg.cochrane.org/author-resources [accessed 2020-03-10]
27. Hess LM, Raebel MA, Conner DA, Malone DC. Measurement of adherence in pharmacy administrative databases: a proposal for standard definitions and preferred measures. Ann Pharmacother 2006;40(7-8):1280-1288. [doi: 10.1345/aph.1H018] [Medline: 16868217]
28. Sackett DL, Haynes RB, Gibson ES, Taylor DW, Roberts RS, Johnson AL. Patient compliance with antihypertensive regimens. Patient Couns Health Educ 1978;1(1):18-21. [doi: 10.1016/s0738-3991(78)80033-0] [Medline: 10238880]
29. Cochrane. RevMan 5 Download. URL: https://community.cochrane.org/help/tools-and-software/revman-5/revman-5-download [accessed 2020-07-25]
30. Cochrane Collaboration. RevMan 5.3 User Guide. London, UK: Cochrane Collaboration; 2014.
31. Boyne JJJ, Vrijhoef HJM, Spreeuwenberg M, De Weerd G, Kragten J, Gorgels APM, TEHAF investigators. Effects of tailored telemonitoring on heart failure patients' knowledge, self-care, self-efficacy and adherence: a randomized controlled trial. Eur J Cardiovasc Nurs 2014 Jun;13(3):243-252. [doi: 10.1177/1474515113487464] [Medline: 23630403]
32. Bobrow K, Farmer AJ, Springer D, Shanyinde M, Yu L, Brennan T, et al. Mobile Phone Text Messages to Support Treatment Adherence in Adults With High Blood Pressure (SMS-Text Adherence Support [StAR]): A Single-Blind, Randomized Trial. Circulation 2016 Feb 09;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]
33. Marek KD, Stetzer F, Adams SJ, Bub LD, Schlidt A, Colorafi KJ. Cost analysis of a home-based nurse care coordination program. J Am Geriatr Soc 2014 Dec;62(12):2369-2376 [FREE Full text] [doi: 10.1111/jgs.13162] [Medline: 25482242]
34. Volpp KG, Troxel AB, Mehta SJ, Norton L, Zhu J, Lim R, et al. Effect of Electronic Reminders, Financial Incentives, and Social Support on Outcomes After Myocardial Infarction: The HeartStrong Randomized Clinical Trial. JAMA Intern Med 2017 Aug 01;177(8):1093-1101 [FREE Full text] [doi: 10.1001/jamainternmed.2017.2449] [Medline: 28654972]
35. Kim JY, Wineinger NE, Steinhubl SR. The Influence of Wireless Self-Monitoring Program on the Relationship Between Patient Activation and Health Behaviors, Medication Adherence, and Blood Pressure Levels in Hypertensive Patients: A Substudy of a Randomized Controlled Trial. J Med Internet Res 2016 Jun 22;18(6):e116 [FREE Full text] [doi: 10.2196/jmir.5429] [Medline: 27334418]
36. Henriksson J, Tydén G, Höijer J, Wadström J. A Prospective Randomized Trial on the Effect of Using an Electronic Monitoring Drug Dispensing Device to Improve Adherence and Compliance. Transplantation 2016 Jan;100(1):203-209. [doi: 10.1097/TP.0000000000000971] [Medline: 26588006]
37. Hosseininasab M, Jahangard-Rafsanjani Z, Mohagheghi A, Sarayani A, Rashidian A, Javadi M, et al. Self-monitoring of blood pressure for improving adherence to antihypertensive medicines and blood pressure control: a randomized controlled trial. Am J Hypertens 2014 Nov;27(11):1339-1345 [FREE Full text] [doi: 10.1093/ajh/hpu062] [Medline: 24771706]
38. Jeong JY, Jeon J, Bae K, Choi Y, Park K, Kim J, et al. Smart Care Based on Telemonitoring and Telemedicine for Type 2 Diabetes Care: Multi-Center Randomized Controlled Trial. Telemed J E Health 2018 Aug;24(8):604-613. [doi: 10.1089/tmj.2017.0203] [Medline: 29341843]
39. Young L, Hertzog M, Barnason S. Effects of a home-based activation intervention on self-management adherence and readmission in rural heart failure patients: the PATCH randomized controlled trial. BMC Cardiovasc Disord 2016 Sep 08;16(1):176 [FREE Full text] [doi: 10.1186/s12872-016-0339-7] [Medline: 27608624]
40. Wald DS, Bestwick JP, Raiman L, Brendell R, Wald NJ. Randomised trial of text messaging on adherence to cardiovascular preventive treatment (INTERACT trial). PLoS One 2014;9(12):e114268 [FREE Full text] [doi: 10.1371/journal.pone.0114268] [Medline: 25479285]
41. Hashimoto S, Brinke AT, Roldaan AC, van VIH, Möller GM, Sont JK, et al. Internet-based tapering of oral corticosteroids in severe asthma: a pragmatic randomised controlled trial. Thorax 2011 Jun;66(6):514-520. [doi: 10.1136/thx.2010.153411] [Medline: 21474498]
42. Marek K, Stetzer F, Ryan P, Bub L, Adams S, Schlidt A. Nurse care coordination and technology effects on health status of frail older adults via enhanced self-management of medication: randomized clinical trial to test efficacy. Nurs Res 2013;62(4):269-278. [doi: 10.1097/nnr.0b013e318298aa55] [Medline: 23817284]
43. Sherrard H, Struthers C, Kearns SA, Wells G, Chen L, Mesana T. Using technology to create a medication safety net for cardiac surgery patients: a nurse-led randomized control trial. Can J Cardiovasc Nurs 2009;19(3):9-15. [Medline: 19694112]
44. Rinfret S, Lussier M, Peirce A, Duhamel F, Cossette S, Lalonde L, et al. The impact of a multidisciplinary information technology-supported program on blood pressure control in primary care. Circ Cardiovasc Qual Outcomes 2009 May;2(3):170-177 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.108.823765] [Medline: 20031834]
45. Stacy JN, Schwartz SM, Ershoff D, Shreve MS. Incorporating tailored interactive patient solutions using interactive voice response technology to improve statin adherence: results of a randomized clinical trial in a managed care setting. Popul Health Manag 2009 Oct;12(5):241-254. [doi: 10.1089/pop.2008.0046] [Medline: 19848566]
46. Bosworth HB, Powers BJ, Olsen MK, McCant F, Grubber J, Smith V, et al. Home blood pressure management and improved blood pressure control: results from a randomized controlled trial. Arch Intern Med 2011 Jul 11;171(13):1173-1180. [doi: 10.1001/archinternmed.2011.276] [Medline: 21747013]

47. Kooy MJ, van Wijk BLG, Heerdink ER, de Boer A, Bouvy ML. Does the use of an electronic reminder device with or without counseling improve adherence to lipid-lowering treatment? The results of a randomized controlled trial. Front Pharmacol 2013;4:69 [FREE Full text] [doi: 10.3389/fphar.2013.00069] [Medline: 23755014]
48. Liu W, Huang C, Wang C, Lee K, Lin S, Kuo H. A mobile telephone-based interactive self-care system improves asthma control. Eur Respir J 2011 Feb;37(2):310-317 [FREE Full text] [doi: 10.1183/09031936.00000810] [Medline: 20562122]
49. Jerant AF, Azari R, Martinez C, Nesbitt TS. A randomized trial of telenursing to reduce hospitalization for heart failure: patient-centered outcomes and nursing indicators. Home Health Care Serv Q 2003;22(1):1-20. [doi: 10.1300/J027v22n01_01] [Medline: 12749524]
50. Willems DCM, Joore MA, Hendriks JJE, Nieman FHM, Severens JL, Wouters EFM. The effectiveness of nurse-led telemonitoring of asthma: results of a randomized controlled trial. J Eval Clin Pract 2008 Aug;14(4):600-609. [doi: 10.1111/j.1365-2753.2007.00936.x] [Medline: 19126178]
51. Santschi V, Rodondi N, Bugnon O, Burnier M. Impact of electronic monitoring of drug adherence on blood pressure control in primary care: a cluster 12-month randomised controlled study. Eur J Intern Med 2008 Oct;19(6):427-434. [doi: 10.1016/j.ejim.2007.12.007] [Medline: 18848176]
52. Düsing R, Handrock R, Klebs S, Tousset E, Vrijens B. Impact of supportive measures on drug adherence in patients with essential hypertension treated with valsartan: the randomized, open-label, parallel group study VALIDATE. J Hypertens 2009 Apr;27(4):894-901. [doi: 10.1097/HJH.0b013e328323f9be] [Medline: 19300114]
53. Lam WY, Fresco P. Medication Adherence Measures: An Overview. Biomed Res Int 2015;2015:217047 [FREE Full text] [doi: 10.1155/2015/217047] [Medline: 26539470]
54. Steiner JF, Prochazka AV. The assessment of refill compliance using pharmacy records: methods, validity, and applications. J Clin Epidemiol 1997 Jan;50(1):105-116. [Medline: 9048695]
55. Farmer KC. Methods for measuring and monitoring medication regimen adherence in clinical trials and clinical practice. Clin Ther 1999 Jun;21(6):1074-90; discussion 1073. [doi: 10.1016/S0149-2918(99)80026-5] [Medline: 10440628]
56. Sabaté E. Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003.
57. Quisel T, Foschini L, Zbikowski SM, Juusola JL. The Association Between Medication Adherence for Chronic Conditions and Digital Health Activity Tracking: Retrospective Analysis. J Med Internet Res 2019 Mar 20;21(3):e11486 [FREE Full text] [doi: 10.2196/11486] [Medline: 30892271]

Abbreviations

CMA: continuous, multiple-interval measures of medication acquisition
MeSH: Medical Subject Headings
PDC: proportion of days medication covered
RCT: randomized controlled trial
RoB: risk of bias
SMS: short messaging service

Edited by G Eysenbach; submitted 12.11.19; peer-reviewed by J Lacroix, K Gashu, C Kruse; comments to author 30.12.19; revised version received 24.02.20; accepted 10.04.20; published 20.08.20.

Please cite as:
Wong ZSY, Siy B, Da Silva Lopes K, Georgiou A
Improving Patients' Medication Adherence and Outcomes in Nonhospital Settings Through eHealth: Systematic Review of Randomized Controlled Trials
J Med Internet Res 2020;22(8):e17015
URL: http://www.jmir.org/2020/8/e17015/
doi:10.2196/17015
PMID:32663145

Review

Online Health Information Seeking by Parents for Their Children: Systematic Review and Agenda for Further Research

Christian Kubb[1], MSc; Heather M Foran[1], PhD

Health Psychology Unit, Institute of Psychology, Universität Klagenfurt, Klagenfurt, Austria

Corresponding Author:
Christian Kubb, MSc
Health Psychology Unit
Institute of Psychology
Universität Klagenfurt
Universitätsstraße 65-67
Klagenfurt, 9020
Austria
Phone: 43 463 2700 1631
Email: christian.kubb@aau.at

Abstract

Background: Parents commonly use the internet to search for information about their child's health-related symptoms and guide parental health-related decisions. Despite the impact of parental online health seeking on offline health behaviors, this area of research remains understudied. Previous literature has not adequately distinguished searched behaviors when searching for oneself or one`s child.

Objective: The purpose of this review is to examine prevalences and associated variables of parent-child online health information seeking; investigate parents' health-related online behavior regarding how they find, use, and evaluate information; and identify barriers and concerns that they experience during the search. Based on this analysis, we develop a conceptual model of potentially important variables of proxy online health information seeking, with a focus on building an agenda for further research.

Methods: We conducted a comprehensive systematic literature review of the PsycINFO, JMIR, and PubMed electronic databases. Studies between January 1994 and June 2018 were considered. The conceptual model was developed using an inductive mixed methods approach based on the investigated variables in the study sample.

Results: A total of 33 studies met the inclusion criteria. Findings suggest that parents worldwide are heavy online users of health-related information for their children across highly diverse circumstances. A total of 6 studies found high parental health anxiety, with prevalences ranging from 14% to 52%. Although parents reported wishing for more guidance from their pediatrician on how to find reliable information, they rarely discussed retrieved information from the web. The conceptual model of proxy online health information seeking includes 49 variables.

Conclusions: This systematic review identifies important gaps regarding the influence of health-related information on parents' health behavior and outcomes. Follow-up studies are required to offer parents guidance on how to use the web for health purposes in an effective way, as well as solutions to the multifaceted problems during or after online health information seeking for their child. The conceptual model with the number of studies in each model category listed highlights how previous studies have hardly considered relational variables between the parent and child. An agenda for future research is presented.

(J Med Internet Res 2020;22(8):e19985) doi:10.2196/19985

KEYWORDS

information seeking behavior; parents; child; internet; health behavior; digital health

Introduction

The proportion of individuals looking for health-related topics online has increased significantly in recent years [1]. Every second internet user in Europe has searched for health-related topics, such as diseases, injuries, or health promotion activities, at least once in the past 3 months [1]. Online health information–seeking (OHIS) behavior has been shown to affect the patient-doctor relationship [2], health care utilization [3,4], and multiple health outcomes [5].

However, large-scale studies suggest that about half of online health-related search sessions are not for one's own health, but rather for someone else's health situation [6,7]. These online health seekers are described in the literature under various terms [8-10] and there is no consensus about the definition of OHIS on behalf of others. The term "surrogate seekers" is sometimes used but has potential for confusion because "surrogate" is associated with surrogate pregnancies, surrogate motherhood, or sexual surrogates. The term is also misleading from an etymological point of view because it suggests that the search is a replacement or substitute for an action that would normally be done by the individual (Latin *surrogatus* means replace). This is particularly not the case in the parent-child search relationship.

We expand the definition on interactive health communication introduced by Robinson et al [11] by adding the term "proxy" seeking. The term "proxy OHIS" refers to any behavior of interactive health communication to obtain information in order to receive support or guidance on a health-related topic for someone else (eg, child, parent, grandparent, friend, neighbor, or any other relative or nonrelative).

Proxy searches are likely when there is a strong emotional tie between two people, which applies especially to intrafamily relations like parent-child relationships, intimate partner relationships, or other family relationships [7,10,12]. Parents consult the web widely for information about their child´s health symptoms and to assist in determining whether they need to seek medical aid for their child [13-15]. Therefore, information from the web can have a crucial impact on a child's health status, as parents use it to make health-related decisions by proxy.

The literature offers numerous studies on parental online information seeking related to their child's health but so far, to our knowledge, there is only 1 literature review that attempts to summarize the findings. This integrative review by Park et al [13] included studies that do not differentiate between self-seeking and proxy seeking. Research has shown significant differences in characteristics of self-seekers and proxy seekers [7-9,16]. OHIS for oneself is typically based on different motives, needs, and circumstances than searching for someone else [8,16,17]. In addition, a recent study by Reifegerste and Bachl [18] suggests that it is not merely the individual factors in the seeker that have an influence on proxy seeking, but also relational factors between seeker and search subject and the relationship's individual characteristics. For these reasons it is unclear whether reviews on the connection between OHIS and other outcomes, like patient-physician relationship [2], health anxiety [19], health literacy, or evaluation of online information [20] can explain the behavior of proxy seekers specifically. Other reviews have focused on general internet behavior of parents [21,22], maternal information-seeking behavior [23], or OHIS during pregnancy [24]; however, the parent-child search relationship was not examined exclusively.

Further, commonly used theoretical models only partially apply to understanding proxy-seeking behaviors. The comprehensive model of information seeking (CMIS) [25] is an established model to predict information-seeking behavior for individuals in different health contexts [26-28]. The influence of demographic variables, such as age, gender, or education, has been inconsistent in the literature on predicting proxy seeking [7-9,16]. Reifegerste and Bachl [18] concluded that further relational variables between searcher and search subject must also be considered in theoretical models to explain these differences. As another consequence, study results on prevalences and associated factors of proxy seeking are not readily transferable to parent-child proxy seeking, since the studies either did not specifically target parents but instead the general public [7-9], or they had a special search relation (eg, family caregivers to cancer survivors [16]).

For these reasons, this review specifically targets research on OHIS by parents. Online health seeking by parents for their children represents an understudied yet important area in the field of health internet research. The aims of this systematic review are (1) to examine prevalences and associated demographic variables of parent-child OHIS, (2) to investigate how parents find, evaluate, and use online health information, (3) to identify which barriers or concerns parents experience online, (4) to document important research gaps and formulate a research agenda, and (5) to develop a conceptual model on proxy OHIS.

Methods

Overview

This systematic review has been performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [29]. For a detailed description, see the PRISMA checklist in Multimedia Appendix 1.

Data Sources and Search Strategy

A comprehensive analysis of the databases of PsycINFO and PubMed was performed. JMIR was also systematically searched. Starting from the word "surrogate seeking," relevant core terms were identified and used for database analysis by applying the pearl finding and growing strategy [30]. These results were combined by using Boolean operators with family-related terms (mother, father, family, caregiver, parent, child). To consider the linguistic variations, these terms were truncated accordingly: (Mother* OR Father* OR Famil* OR Caregiver* OR Parent* OR Child*) AND (Internet OR Web OR Online OR Cyber* OR eHealth OR e-Health OR Health Information OR Information Seeking).

Studies from January 1994 to June 2018 were considered. The year 1994 was chosen because in this year, the first International World Wide Web Conference took place [31]. The existing web did not have essential health services at that time, and internet use was not common.

Records were summarized in a text-based database. After elimination of the duplicates, titles and relevant abstracts were reviewed. The full texts of the remaining records were reviewed to determine whether they met all inclusion criteria. A protocol of the process for selecting studies is available in Multimedia Appendix 2.

Inclusion and Exclusion Criteria

To study OHIS by parents for their children, we included papers that met the following criteria: (1) the participants were parents; (2) the focus of the investigated behavior was OHIS on publicly available websites; (3) the online health-seeking behavior was for their own child; and (4) the study was written in English, presented quantitative data, and was published in a journal between 1994 and 2018.

First, the participants were parents. We defined parents as the primary caregivers who substantially support the child over a stable period in daily routines like feeding, hygiene, play, sleep, or health. Studies including other caregivers (eg, grandparents, other family members) besides parents were excluded if the percentage of other caregivers was greater than 5% of the total sample.

Second, the focus of the investigated behavior was OHIS on publicly accessible websites. Excluded papers included those about special online behaviors (evaluation of one specific website) or areas that are only accessible with registration (support groups, discussion boards, chats), papers that focused only on offline information-seeking behavior (books, television, physicians), and papers with a focus on non–health relevant search behavior.

Third, the online health-seeking behavior was for their own child. Studies with self-seeking behavior only (searching for own medical issues) and studies in which a self-seeking and proxy-seeking distinction was not made or was not possible (eg, pregnancy) were excluded.

Fourth, only papers written in English, presenting quantitative data, and published in a journal between 1994 and 2018 were included.

Based on studies that met the inclusion criteria, we manually reviewed their references to identify further studies that may not have been found through the literature review. Further, we used Google Scholar in June 2018 to identify cited papers that met the inclusion criteria.

Selection of Studies

One author (CK) manually merged the studies from the different databases, removed duplicates, screened titles and abstracts for relevance, and hand-searched additional citations. The remaining records after screening by title and abstract were independently checked for eligibility by an author (CK) and a psychology master's student (PS) (Cohen κ=0.84). In cases of nonagreement (7 out of 136), studies were discussed and a consensus for inclusion or exclusion was reached.

Data Extraction and Analytical Strategy

The formal study characteristics were extracted from all 33 papers by 1 main author (CK) and can be found in their entirety in Multimedia Appendix 3 with a description of the studies (author, year of publication, location, survey period), study design (survey methodology, prospective vs nonprospective, cross-sectional vs longitudinal, hypothesis generating vs hypothesis testing, sampling technique), and sample characteristics (sample size, amount of parents in the sample, clinical vs community sample, parental gender, parental age, race, education, income, occupation, health insurance, child's age). Subsequently, the content focus of each paper was coded in 3 category clusters: (1) studies with OHIS related to a child's specific disease, (2) studies with OHIS related to a treatment or circumstance, and (3) studies that investigated parent-child OHIS in general.

We extracted the quantitative surveyed prevalences on parental OHIS as well as the related item that was used to assess prevalence because the study-specific prevalences are based on varying defined timespans. Significant and nonsignificant associated factors related to these items were extracted as well. To develop a research agenda, further information on theoretical frameworks, study limitations, and mentioned research gaps were extracted from the reviewed studies.

The heterogeneity of the sample composition of studies and the lack of a sufficient sample size of studies with similar outcome variables made the use of meta-analytical methods inappropriate for this review. Therefore, data were summarized by conducting a descriptive analysis and narrative synthesis. Frequency counts of key variables were coded and summarized.

Coding for Conceptual Model

The conceptual model was developed with an inductive approach by the 2 authors (CK, HMF) based on the investigated variables in the studies. The CMIS by Johnson and Meischke [25] provides a theoretical framework and served as a basic structure to categorize the extracted variables. The underlying assumption of the CMIS is that characteristics of the individual and characteristics of the medium jointly influence health information–seeking behavior. Specifically, the model considers antecedents in the seeker (demographics, personal experience, beliefs, and salience), the characteristics and perceived utility of the information carrier (eg, health information on a website), and the final health information–seeking action (eg, decision to see a doctor). Based on the Johnson and Meischke [25] classification, we renamed the category names to make it more suitable for an internet search and distributed the constructs according to this distribution. Demographics, personal experience, beliefs, and salience of the CMIS are subsumed under "personal factors" and "environmental factors." Characteristics and utilities of the CMIS are classified under "online search factors." Information-seeking actions of the CMIS correspond to the outcome category labeled "health decision making and behaviors." Finally, we have added the relational categories "relational factors" and "search subject" to our model. These are unique to health information seeking by proxy.

One author (CK) scanned the papers for quantified variables and created a binary coding system (1=variable is investigated; 0=variable is not investigated) with definitions for 49 variable categories to examine the frequencies of considered variables in the whole study sample (Multimedia Appendix 4). A psychology bachelor's student (AS) and 1 author (CK) coded the variables in the studies independently (Cohen κ=0.69).

Results

Description of Studies

A total of 33 studies met the inclusion criteria (Figure 1 [32]). All studies were cross-sectional. Studies were conducted using in-person questionnaires (n=23), online surveys (n=5), interviews (n=2), telephone surveys (n=2), and mailed questionnaires (n=1).

The papers were sorted into 3 groups based on the focus of the paper (Table 1). A total of 13 of the studies focused on OHIS related to a specific disease or disease cluster, including asthma [34], attention-deficit/hyperactivity disorder [35], brachial plexus birth palsies [36], congenital heart disease [33,37,38], diabetes [39], hearing loss [40], hydrocephalus [41], scoliosis [42], skin disorders [43], and rare diseases [44,45]. An additional 13 studies addressed specific circumstances, mainly prior to a surgical procedure [46-51], after childbirth [52], 24 hours before an emergency department visit [53], during stay in a neonatal intensive care unit [54], in a pediatric outpatient clinic [55,56], in palliative care [57], and regarding attitudes towards human papillomavirus vaccination [58]. In addition, 7 studies dealt with general OHIS without specified diseases or particular circumstances [14,59-64].

Figure 1. Flowchart of the systematic review search process. Adapted from Moher et al [32].

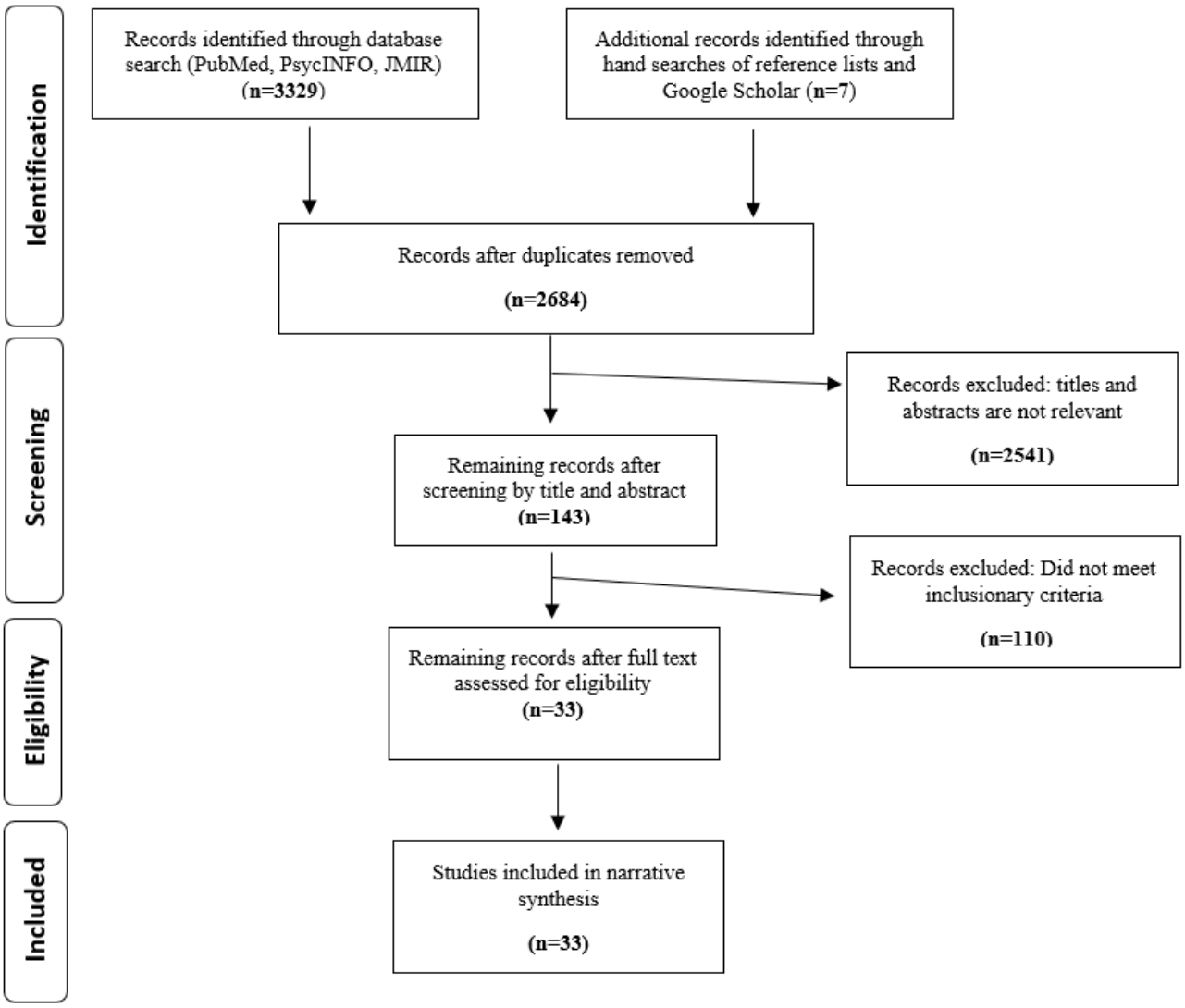


Table 1. Study focus cluster.

Author	Year	Group	Specification
AlSaadi [34]	2012	Disease[a]	Asthma
Baker et al [42]	2012	Disease[a]	Scoliosis
Balkhi et al [39]	2015	Disease[a]	Diabetes
Ikemba et al [37]	2002	Disease[a]	Congenital heart disease
Kasparian et al [33]	2017	Disease[a]	Cogenital heart malformation
Lai and Mallory [43]	2000	Disease[a]	Skin disorders
Massin et al [38]	2006	Disease[a]	Congenital heart disease
Naftel et al [41]	2013	Disease[a]	Hydrocephalus
Nicholl et al [44]	2017	Disease[a]	Rare diseases
Porter and Edirippulige [40]	2007	Disease[a]	Deafness
Sage et al [35]	2017	Disease[a]	Attention-deficit/hyperactivity disorder
Shah et al [36]	2006	Disease[a]	Brachial plexus birth palsies
Tozzi et al [45]	2013	Disease[a]	Rare diseases
Boston et al [46]	2005	Circumstance[b]	Otolaryngology procedures
Dhillon et al [54]	2003	Circumstance[b]	Neonatal intensive care
Glynn et al [56]	2013	Circumstance[b]	Otolaryngology services
Hand et al [47]	2013	Circumstance[b]	Surgical procedure
Knapp et al [57]	2010	Circumstance[b]	Palliative care program
McRee et al [58]	2012	Circumstance[b]	Human papillomavirus vaccination
Nogueira et al [48]	2013	Circumstance[b]	Otolaryngology surgery
Semere et al [49]	2003	Circumstance[b]	Surgical procedure
Shroff et al [53]	2017	Circumstance[b]	24 hours before emergency department
Sim et al [50]	2007	Circumstance[b]	Surgical procedure
Slomian et al [52]	2017	Circumstance[b]	After childbirth
Tuffrey and Finlay [55]	2002	Circumstance[b]	Pediatric outpatients
Wong et al [51]	2017	Circumstance[b]	Surgical procedure
Harvey et al [59]	2017	General[c]	N/A[d]
Opeoluwa et al [60]	2017	General[c]	N/A
Pehora et al [61]	2015	General[c]	N/A
Sebelefsky et al [62]	2015	General[c]	N/A
Skranes et al [63]	2014	General[c]	N/A
Whyte and Hunter [64]	2008	General[c]	N/A
Yardi et al [14]	2018	General[c]	N/A

[a]"Disease" group indicates studies investigating parental online health information seeking related to specific illnesses, diseases, or disorders.
[b]"Circumstance" group indicates studies investigating parental online health information seeking related to a treatment or circumstance/situation.
[c]"General" group indicates studies investigating parental online health information seeking in general without a specified disease or circumstance.

[d]N/A: not applicable.

Sample Characteristics

The samples from all studies included a total of 8665 participants and varied from a sample size of 70 [35] to 848 [58] participants, with a median of 209 participants (Table 2). A total of 26 out of 33 studies specified the proportion of mothers and fathers in the samples; with 4758 mothers and 1353 fathers, 77.86% were mothers (4758/6111). Of the 33 studies, 29 (88%) were conducted in the Western world, of which one-third of all studies (n=11) were conducted in the United States. Samples from other parts of the world included Nigeria [60], Singapore [51], Saudi Arabia [34], and Brazil [48].

Reported parental mean ages were all between 30 and 42 years, but only half of studies (n=16) reported ages of parent and child. The child's age varied from neonates [37,54] to adults [40,42,44,45,55], but studies with reported mean ages or distributions consisted mainly of toddlers, preschoolers, and school-aged children aged 1 to 12 years. Adolescents were targeted in only one study explicitly [58]. Only 2 studies [34,62] differentiated between mothers' and fathers' demographic data and listed their information separately.

The samples consisted of highly educated parents, with more than 50% [33,34,36,41,44,48,54] and up to more than 75% of parents holding academic degrees [52,58,63], but 14 studies did not report any educational levels. The proportion of persons with only primary education varied between 0% [52] and 21.9% [57] among the studies that reported on education levels.

Table 2. Sample characteristics.

Author	Location	Sample size, n	Proportion of mothers, %	Sample	Child's age	Parent's age (years)
AlSaadi [34]	Saudi Arabia	500	__[a]	Clinical	<5 y (63.3%) >5 y (36.7%)	—
Baker et al [42]	Ireland	167	81	Clinical	Mean 11.9 y (SD 4)	<20 (12%) 20-35 (7%) 35-50 (75%) 50-65 (7%)
Balkhi et al [39]	United States	209	72	Clinical	Mean 12.26 y (SD 4.7)	Mean 42.15 (SD 8.94)
Boston et al [46]	United States	204	64	Clinical	—	Mean 34, range 16-65
Dhillon et al [54]	Canada	90	67	Clinical	2-148 days	Median 32
Glynn et al [56]	Ireland	501	75	Clinical	—	<18 (2%) 18-40 (68%) 41-65 (30%) >65 (<1%)
Hand et al [47]	Ireland	214	79	Clinical	—	<18 (1%) 18-40 (77%) 41-65 (21%)
Harvey et al [59]	Ireland	100	81	Clinical	<3 y (35%) 4-6 y (15%) 7-9 y (13%) 10-12 y (16%) >12 y (21%)	—
Ikemba et al [37]	United States	275	45	Clinical	Mean 4.3 y, range 7 d-24 y	—
Kasparian et al [33]	Australia	132	63	Clinical	Mean 21.8 months (SD 5.6)	Mean 35.2 (SD 7)
Knapp et al [57]	United States	129	90	Clinical	Mean 9.9 y (SD 6.1)	Mean 42.9 (SD 11.7)
Lai and Mallory [43]	United States	467	—	Clinical	—	—
Massin et al [38]	Belgium	389	47	Clinical	Mean 6 y (SD 4.9)	—
McRee et al [58]	United States	848	92	Nonclinical	First sample: Mean 14.7 y (SD 3.5); Second sample: Mean 13.9 y (SD 2.2)	First sample: <45 (28.2%) >45 (71.8%); Second sample: <45 (63.5%) <45 (36.5%)
Naftel et al [41]	United States	300	—	Clinical	Mean 8.2 y (SD 5.8)	Mean 36.7 (SD 10.4)
Nicholl et al [44]	Ireland	93	87	Clinical	<1 y (4%) 1-3 y (20.5%) 4-7 y (28.2%) 8-12 y (23.9%) 13-19 y (12.8%) 20-29 y (7.7%) 30-39 y (2.6%)	18-34 (24%) 35-49 (67%) 50-64 (10%)

Author	Location	Sample size, n	Proportion of mothers, %	Sample	Child's age	Parent's age (years)
Nogueira et al [48]	Brazil	132	83	Clinical	range 2-14 y	Mean 42, range 18-66
Opeoluwa et al [60]	Nigeria	142	100	Clinical	—	<20 (31.2%) 21-30 (42.2%) 31-40 (22%) >40 (4.6%)
Pehora et al [61]	Canada	146	—	Clinical	—	—
Porter and Edirippulige [40]	Australia	166	89	Clinical	<1 y (6%) 1-2 y (11%) 2-5 y (26%) 5-10 y (23%) 10-15 y (20%) 15-18 y (9%) 18-21 y (5%)	18-34 (29%) 34-49 (67%) 50-64 (4%)
Sage et al [35]	United States	70	81	Clinical	Mean 12 y (SD 2.6)	Mean 42.9 (SD 7.1)
Sebelefsky et al [62]	Austria	500	82	Clinical	Mean 2.4 y (SD 2.6)	Mean 34 (SD 6.4)
Semere et al [49]	United States	150	83	Clinical	—	Mean 35 (SD 11)
Shah et al [36]	United States	122	77	Clinical	—	—
Shroff et al [53]	United States	262	84	Clinical	Median 4 y (IQR 1.3-11)	Median 31 (IQR 25-37)
Sim et al [50]	United Kingdom	271	70	Clinical	—	—
Skranes et al [63]	Norway	99	100	Nonclinical	Mean 1.6 y, range 0.3-11 y	Mean 33.1, range 21-58
Slomian et al [52]	Belgium	349	100	Nonclinical	Mean 12.7 months (SD 14.5)	Mean 30.6 (SD 4.05)
Tozzi et al [45]	Italy	516	68	Clinical	Mean 10.3 y (SD 9)	Mean 42.7 (SD 9)
Tuffrey and Finlay [55]	United Kingdom	485	—	Clinical	Mean 6.3 y, range 4 weeks-23 y	—
Whyte and Hunter [64]	United Kingdom	245	—	Clinical	—	—
Wong et al [51]	Singapore	84	63	Clinical	Sample 1: median 5.1 y (range 0.2-15.7); Sample 2: median 9.8 y (range 0.6-15.9)	—
Yardi et al [14]	Australia	308	—	Clinical	—	<25 (7%) 25-44 (76%) 45-55 (15%) >55 (2%)

Author	Location	Sample size, n	Proportion of mothers, %	Sample	Child's age	Parent's age (years)
Total	N/A[b]	N=8665	77.86[c]	N/A	N/A	N/A

[a]Not available (exact numbers are not given by the author).

[b]N/A: not applicable.

[c]4758/6111. Calculated from the studies (n=26) that provided information on parental gender.

Prevalence of Parental Online Health Information Seeking

Table 3 presents the prevalences of OHIS by proxy and associated factors, separated into general OHIS and OHIS for specific conditions in the child. In studies that reported prevalence by parental OHIS in general (n=9), prevalence ranged from 52% to 98%. Only 3 studies explicitly distinguished between general OHIS and specific OHIS [36,53,62]. Recent studies from 2017 or later showed the highest prevalences, with around three-fourths [51,59,60] to roughly 9 out of 10 parents who searched for health information related to their child [14,33,35,44]. Likewise, the health-related internet use among parents of children with rare diseases seems to be relatively high [44,45], even for older studies that deal with rare conditions [36,37]. Most of the data are related to OHIS before or because of surgical intervention [46-51]. In those cases, the prevalence varied between 38% and 90%.

In Table 3, all variables investigated in relation to parental OHIS are reported. Only a small proportion of studies provided bivariate or multivariate analysis of associated factors with OHIS, often presenting only selective data with significant outcomes. Education was shown to be the most common associated factor with parental OHIS [34,36,40,41,47,53,54,56,57], although some studies found no significant association with education [35,46,62].

The gender of the parent was not related to whether a parent searched the internet for their child in most studies [35,42,53,54,57,62], but it was related in one study [33]. The influence of the age of the parents on OHIS was inconsistent; some studies found younger age to be associated [56,62], one found older age to be associated [53], and other studies found no association with age and search behavior [14,35,40,42,47,57].

Table 3. Prevalence of online health information seeking and related factors.

Author	General OHIS[a] for child[b,c]	Specific OHIS for child[b,c]	Associated factors	Nonassociated factors
AlSaadi [34][d]	—[e]	79% (—/505) "Using the Internet to gain information on their children's [asthmatic] condition"	Father's education, mother's education, occupation of mother, nationality of father (Saudi vs non-Saudi), nationality of mother	Father's nationality, occupation of father, history of allergy
Baker et al [42][d]	—	58% (97/165) "Have you searched the internet for information on scoliosis?"	Corrective surgery, private health insurance	Postoperative complications, parent gender, education, child age, parent age group, visit type, home internet access
Balkhi et al [39][d]	—	64% (133/209) "Using the Internet for diabetes information"	Child's age	HbA_{1C}[f] level
Ikemba et al [37][d]	—	58% (93/160) "Used the Internet to obtain information related to their child's cardiac diagnosis"	—	Type of congenital heart defect
Kasparian et al [33][d]	—	91% (—) "Identified the internet as a source of congenital heart disease information"	Parents' gender (mothers)	—
Lai and Mallory [43][d]	—	13% (62/467) "Used the Internet to search for information related to their child's skin disorders"	—	—
Massin et al [38][d]	—	35% (84/238) "Used the Internet to obtain information related to their child's cardiac diagnosis"	Expected treatment modalities	Type of congenital heart defect, internet access at home
Naftel et al [41][d]	—	82% (225/275) "Searching for hydrocephalus-related information online"	Caucasian, income, education	Geographic location (urban vs rural), parents' age, etiology of hydrocephalus
Nicholl et al [44][d]	—	92% (105/114) "Use the Internet to find information about your child's condition [at least every few month]"	—	—
Porter and Edirippulige [40][d]	—	82% (131/159) "Use the Internet to find information about deafness and related topics [at least every few months]"	Education	Parents' age, child's age, geographic area, employment status, type of hearing loss
Sage et al [35][d]	—	87% (61/70) "Searching the Internet for ADHD[g] information"	—	Parents' age, parents' gender, years of education
Shah et al [36][d]	90% (—/122) "Searched the Internet for health-related information at least once a month"	88% (108/122) "Used the Internet to search for information on Brachial Plexus Birth Palsies"	Education, income	—
Tozzi et al [45][d]	—	99% (462/468) "Information searched on the web [related to disease characteristics]"	—	—

Author	General OHIS[a] for child[b,c]	Specific OHIS for child[b,c]	Associated factors	Nonassociated factors
Boston et al [46][h]	—	49% (83/170) "Used the Internet to look for information about their child's diagnosis and/or surgical [otolaryngology] procedure"	—	Education, frequency of internet use
Dhillon et al [54][h]	—	44% (40/90) "Having searched the Internet for information related to the medical condition of their baby [in the neonatal intensive care unit]"	Education, parents' age	Parents' gender, employment status, comfort in English
Glynn et al [56][h]	—	30% (149/497) "Had searched online for information regarding their child's ENT[[i]] problem"	Education, parents' age, private health insurance, daily internet use, smartphone	—
Hand et al [47][h]	—	38% (82/214) "Searched the internet regarding their child's surgical issue"	Education, private health insurance, daily internet use, smartphone	Parents' age
Knapp et al [57][h]	—	81% (92/114) "Used Internet information about their children's health [who have life-threatening illnesses]"	Education, parents' race, language spoken at home (English)	Parents' gender, parents' age, marital status, type of household, sibling in household, children's age, children's health
McRee et al [58][h]	—	21% (154/773) "[Mothers] having heard about HPV[[j]] vaccine through the Internet" 17% (19/115) "[Fathers] having heard about HPV vaccine through the Internet"	Greater knowledge about HPV	—
Nogueira et al [48][h]	—	90% (117/130) "Look[ed] for information on the Web on the condition of your child/guardian [with undergoing otolaryngology surgical procedure]"	—	—
Semere et al [49][h]	—	69% (88/128) "Searched for information relating to their child's surgery procedure or treatment"	—	—
Shroff et al [53][h]	52% (117/224) "At least one episode of Internet use for general pediatric health information in the preceding 3 months"	12% (31/262) "Used Internet in 24 hours prior to emergency department visit"	Education, income, older children, older parents	Parents' gender, race of parent, race of child, insurance, triage classification, time of enrollment, disposition
Sim et al [50][h]	—	53% (144/271) "Had accessed the Internet to seek more information regarding their children's condition [surgical outpatient]"	—	—

Author	General OHIS[a] for child[b,c]	Specific OHIS for child[b,c]	Associated factors	Nonassociated factors
Slomian et al [52][h]	—	12% (43/349) "Seeking information for the baby only [after childbirth]" 75% (262/349) "Seeking information about themselves or about their baby [after childbirth]"	—	—
Tuffrey and Finlay [55][h]	—	22% (107/485) "Used the internet to find information about the problem for which they were being seen in clinic that day"	Internet access at home	—
Wong et al [51][h]	—	74% (62/84) "Use the Internet to access child's current condition [surgical procedure]"	—	—
Harvey et al [59][k]	72% (72/100) "Frequency of use of the Internet to access healthcare information at least yearly or less"	—	—	Children with chronic diseases
Opeoluwa et al [60][k]	77% (109/142) "Had ever consulted the Internet to find answers to their babies' medical problems or health-related issues"	—	Self-medication, health-seeking behaviors	—
Pehora et al [61][k]	98% (143/146) "Using the Internet to search for health information regarding their child [at least few times a year]"	—	—	—
Sebelefsky et al [62][k]	94% (471/500) "General internet use to obtain child health information [at least occasionally]"	21% (105/499) "Internet use to be informed about the reason for consultation [pediatric outpatient clinic]"	Parents' age (younger parents), younger children	Parents' gender, nationality, education, children's sex, children's diet
Skranes et al [63][k]	98% (97/99) "Used the Internet regularly to search for child health information"	—	—	—
Whyte and Hunter [64][k]	64% (121/190) "Used Internet to search for information regarding child's health"	—	—	Scottish Index of Multiple Deprivation

Author	General OHIS[a] for child[b,c]	Specific OHIS for child[b,c]	Associated factors	Nonassociated factors
Yardi et al [14][k]	90% (276/308) "Searching for medical information about their child's health"	—	—	Parents' age, number of children, in-patient/outpatient, parent-perceived seriousness of child's condition

[a]OHIS: online health information seeking.

[b]Percentages are rounded.

[c]Textual information in brackets has been added for better understanding.

[d]Group 1: study investigated parental OHIS related to specific illnesses, diseases, or disorders.

[e]Not available (exact numbers or information not given by the author).

[f]HbA_{1C}: glycated hemoglobin.

[g]ADHD: attention-deficit/hyperactivity disorder.

[h]Group 2: study investigated parental OHIS related to a specific treatment or circumstance/situation.

[i]ENT: ear, nose, and throat.

[j]HPV: human papillomavirus.

[k]Group 3: study investigated parental OHIS in general.

How Parents Find, Evaluate, and Use Health Information

Google was reported to be the most common starting point for gathering health information [33,42,44,48,50,51,53,62,63]. The most recent studies found that 9 out of 10 parents use Google [33,44,51] and many of the daily internet users go online via their mobile phones [45]. The rising trend in mobile phone use over desktop computer use was already evident in the studies since 2013 [41,44,45,47,52,53,56]. There is some evidence that smartphone owners are more likely to look for health-related information relating to their child [47,56] than people without a smartphone. Yardi et al [14] reported first that smartphones have overtaken the desktop computer as the most used device for proxy health information seeking.

Parents described information from the web mostly as helpful and useful, with a fundamentally positive attitude towards OHIS [33,34,37,38,41-52,56,57,63,64]. The most frequent underlying search motive was the need for a better understanding of the child's condition, which gave parents the opportunity to play a more active role in the management of their child's health [14,33,34,42,43,45,49,51,52,55].

Parents used the internet to decide if their child needed a doctor [14,60,63] and in some cases also in emergency situations [53]. Likewise, they searched the web before a doctor's visit to prepare for the appointment and after the doctor's visit to address unanswered questions [14,33,42]. The web was also reported to be used as a second medical opinion, but the amount varied from 1% to 57% across studies covering different medical circumstances [44,45,51,52].

Unfortunately, only a few studies examined search content in detail. Information about characteristics of specific diseases, current treatments, and diagnoses were the most common search topics [36,44,45,50,51,55,59], while looking for alternative treatments [40,44,50,51,55] was comparatively less common. However, the choice of treatment could be influenced by the information from the internet [46-48,56]. Studies that did not restrict their items on search content to a specific disease showed a greater variety of search content, including searches for health purposes like children's nutrition [44,45,52,61] or development [44,52,61].

A consistent finding across studies over time was the search for or use of support groups [14,33,39-41,44,45,49-51,55,57]. In particular, parents of children with chronic, acute, or rare diseases showed a high need for support groups [40,41,44,45,57].

Barriers and Concerns That Parents Experience Online

Parents perceived the information on the web as easy to understand [14,43,46-48], but studies found that parents sometimes had problems distinguishing between trusted and untrusted websites [14,33] or finding reliable information [54,60]. Some studies showed only a small proportion of parents who considered the reliability and trustworthiness of the information [14,49,51], while other studies showed greater skepticism of the participants towards the internet as a reliable source [41,54]. Further, parents did not necessarily navigate to the pages that they trusted or that provided trustworthy information [61]. The web as a trusted source was ranked lowest, but it is used frequently as a source of health information [46,54], and even unreliable information was reported as helpful [52,54].

Although parents wished for more guidance regarding good websites from their physicians [14,41,48,52,53], parents rarely or never discussed their findings with them [14,34,40,46-48,50,51]. Reasons for not discussing findings included a lack of time and a fear of doctors' disapproval [33,51,52,59]. Other problems mentioned included nonnative language information [34,38], technical language [34,51], and information overload [33,38,51].

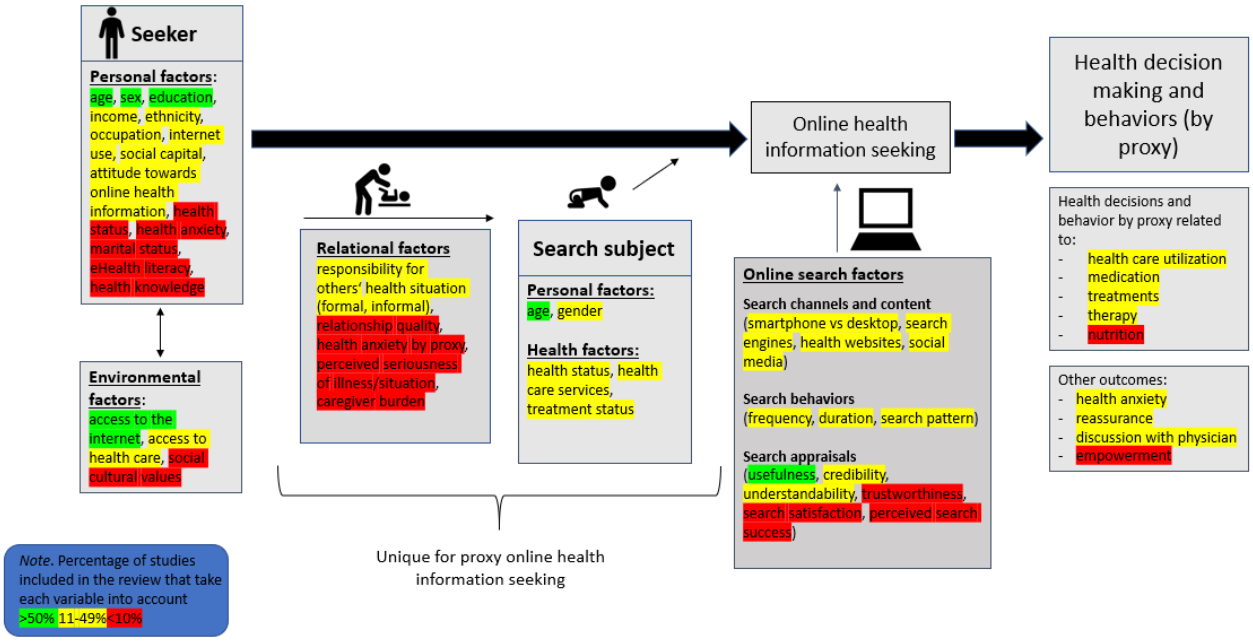
There were 6 studies that reported anxiety, distress, or worries caused by information from the internet [33,42,44,45,50,51]. The proportion of affected parents was between 14% [51] and 52% [45]. Nicholl et al [44] reported that online searching lowered anxiety in 16% of their participants, but the number of people with increased fear was twice as large. Likewise,

attention should also be paid to parents who spend a lot of time searching the web for health purposes or who visit many different sites, like in Shah et al [36] or Porter and Edirippulige [40], which indicates that some parents may not be able to find the health information they need.

Model of Online Health Information Seeking by Proxy

The conceptual model (Figure 2) consists of 6 categories: personal factors within the seeker, environmental factors, relational factors between seeker and search subject, factors within the search subject (ie, the child in regard to personal and health variables), online search factors (search channels, content, behaviors, and appraisals), and outcomes. Factors of the search subject and the relational factors are unique for OHIS by proxy.

Figure 2. Model of proxy online health information seeking and decision making.



The most commonly studied variables were the age of the parent (21/33), their gender (23/33), their education (21/33), and their access to the internet (25/33), as well as the age of the child (21/33). Attitude on usefulness of online health information was the only variable from the other categories that was frequently included in studies (22/33). The most neglected study variables were relational variables. However, variables from the category of online search factors were also rarely assessed, especially trustworthiness of information (2/33), search satisfaction during the search (2/33), and the perceived search success after the search (1/33). In addition, influencing factors such as trait health anxiety (0/33), health knowledge (2/33), and eHealth literacy (2/33) were also considered in less than 10% of the studies.

The surveyed variables of the studies that met the inclusion criteria for narrative synthesis (N=33) were counted. On average, studies considered 12.12 variables (SD 5.36) out of 49 coded variables (Multimedia Appendix 4). The study by Whyte and Hunter [64] presented the fewest number of variables (n=2) and Kasparian et al [33] considered the most variables (n=28).

Discussion

Principal Findings

Parents are heavy users of health-related information on the internet for their children across highly diverse circumstances. Across studies, results showed that the majority of parents have searched the web at least once for general health information for their child. This indicates that the percentage of parents who search for their child is well above the national averages for self-seekers [1]. Education was the most consistent predictor for proxy OHIS across all studies. Well-educated parents used the internet for their children more than parents with little education. The most recent studies show that Google search engine was used by almost all parents as a starting point for OHIS.

Further, information retrieved from the web was reported to be used by caregivers for decision making about children's health. Physicians should be aware that parents reported using information found on the web under certain circumstances for treatment choice or to make health care utilization decisions. Although parents rarely or never discussed information from the web with their doctor, studies showed that they would like more support from doctors on how to find reliable sources. However, there is a tremendous lack of understanding about which criteria parents use to make decisions and about individual and external factors that contribute to parental empowerment. More research is needed on offering parents tangible knowledge and appropriate guidance, using the web for health purposes in an effective way, and problem solving approaches to the multifaceted problems that come up during or after OHIS for their child (eg, unmet information needs, obstacles in

parent-doctor communication, false proxy lay diagnoses by parents with wrong conclusions for child's needed treatment, unnecessary or missed doctor visits, and parental health anxiety due to online health information). All 6 studies that surveyed anxiety and distress by proxy showed a significant proportion of affected parents.

Implications for Future Research

Overall, this review identified the need for more developed research in the area of OHIS. As can be seen from the results of the review, most studies provided descriptive information, but process- and theory-driven advances in this research area are still in infancy. To facilitate more systematic research in the area of OHIS, we present a summary of research gaps in the context of the conceptual framework provided. A total of 17 studies included in this review named research gaps explicitly or gave suggestions for further research. We synthesized these into the results of the developed model on proxy OHIS and the current research on proxy health searches. This is a framework that can be used for future studies.

Differentiation of Self-Seekers and Proxy Seekers

First, we recommend a clear separation between parental self-seeking and parental proxy seeking. These health behaviors represent two independent processes with different motivations, circumstances, and predictors [7-9,16,17]. If both are considered in one study, authors must state explicitly which they are referring to. Numerous excluded studies mixed them or formulated the research items in a vague or undifferentiated way. Furthermore, it is largely unknown whether findings from parent-child OHIS also apply to other types of proxy seeking, such as searching for a spouse or parents. For instance, existing research indicated that proxy seekers tend to be women [7,16], but this review showed that the gender of the parent had no influence on whether they searched the internet for their child. As suggested by Reifegerste et al [17], relational factors are relevant variables for proxy seeking and therefore a fundamental part of our conceptual model. They could explain contradictory results from past research. Relational variables like relationship closeness and quality should be considered in future studies.

Representative Samples and Generalizability

Second, studies with generalizable samples are urgently needed to provide an accurate estimate of the actual prevalence and influencing factors of parental proxy seeking. The lack of generalizability of the results is the most frequently mentioned limitation, which leads to the recommendation for larger and more diverse samples in further studies [33-35,51,54,57,61,63]. None of the studies in this systematic review had a representative sample. Existing literature mainly offered convenience samples in clinical environments with specific populations of ill children. However, it is unclear whether the results are also generalizable to parents of children who are not seriously ill and whether general patterns across proxy seekers can be established. In addition, systematic studies from non-Western countries have so far been lacking, for example in Asian and African regions, where smartphone and internet use has increased substantially in recent years [65]. There is virtually nothing known about how parents search and behave in low-income countries, where they have access to the web but may have limited access to some health care options.

Theoretical Frameworks

Third, the theoretical approaches are still lacking after 20 years of research in the field of parental OHIS. Only 3 studies [57,58,60] referred to existing theories at all, and none of these studies used them to interpret their results. Existing theory-based literature on models of OHIS did not consider proxy seekers [66]. Nonetheless, health characteristics of the supported search participants are associated with the search behavior of the proxy seeker [17,67]. To address this gap, we present a conceptual model on OHIS by proxy. This framework can be used for future studies in order to consider important influencing variables on parent-child OHIS.

Advanced Modeling Techniques

Fourth, data analyses in previous studies have been limited in scope. New studies should analyze the collected data with advanced statistical methods and go beyond the solely descriptive approach that has commonly been used so far. Structural equation modeling could be beneficial for testing the conceptual model proposed. Further, longitudinal analyses would help explore search behaviors and their connection with health care decisions and health care utilization behaviors over time.

Dyadic data analysis could be used to test both parents' search behaviors in the context of the conceptual model. Results of this review found that both mothers and fathers searched for health-related information regarding their children [35,42,53,54,62]. To what extent they differ in search behaviors and whether interpersonal interactions influence search behaviors could be examined in future dyadic studies. Literature suggests that fathers' involvement can impact a child's social, behavioral, and psychological outcomes [68], and the results of this review reveal that social capital is an important variable that has been included in approximately half of the studies. Dyadic modeling could help address the question of how co-occurring proxy seeking by mother and father affects their health decisions and their child's health outcomes.

Social Media

Fifth, upcoming studies need to focus more on the new possibilities on the internet. The landscape for consuming health-related information is completely different than it was ten years ago but it is hardly studied for parental proxy seeking. Facebook, Twitter, and YouTube are heavily frequented to find and share health information, but parental social media behaviors are not well understood. In particular, trustworthiness of information online was found in our review to be understudied, with only 2 studies examining this construct. This may be more important than ever due to the challenges related to health misinformation and fake news on social media [69,70]. In this context, the impact of far-reaching influencer personalities on platforms such as Instagram on the health behavior of young parents has barely been considered in the literature on OHIS. So far, it is also unknown whether the use of smartphones instead of desktop computers has fundamentally changed the

search for health-related information, since these devices are now accessible immediately and everywhere.

Factors of a Successful Online Health Search and Interventions to Improve Search Skills

Sixth, evidence on factors that result in search success among parents searching for health information is lacking. Mixed method approaches with eye tracking, desktop tracking, or think-aloud protocols with evaluation immediately after a health-related search could contribute to better understanding of which parental factors (eg, eHealth literacy) and search process factors (eg, number and choice of sources, search duration) might be associated with a positive search outcome that empowers parents. Based on those types of studies, evidence-based recommendations for parents could be formulated for use on health-related sites on the internet. The online search factors category of the conceptual model (Figure 2) presents nonpersonal related variables that may play a role in the search process only. In addition, to move forward in the research area of search success, new psychometrically tested measures that operationalize search success in a valid and reliable way will need to be developed.

The question of how to improve parental searching skills with interventions is also in need of further research [14,58,63]. It is unclear how parents can be empowered effectively for OHIS [14] and if educational interventions are able to improve parents' health information–seeking skills on a long-term basis [33]. Research on approaches and skills to teach parents appropriate and effective methods of proxy OHIS are still needed [14,33,40].

Suffering From Online Health Information Seeking

Seventh, the negative accompaniments of OHIS, such as uncertainty, anxiety, or triggered health care utilization, are well described among self-seekers [4,71-73] but rarely investigated for proxy seekers [74]. Some studies have documented that parents are also negatively impacted from information seeking [33,42,44,45,50,51]. However, there is a lack of research that applies approaches to improve the outcomes for parents who currently do not benefit from proxy OHIS. Moreover, taking into consideration the relational aspect between seeker and subject may lead to a better understanding of the prevention of negative outcomes for parents searching for health information [75].

Effects on the Doctor-Parent Relationship

Eighth, the role of health professionals and their reciprocal communication with online health–seeking parents needs more investigation. Research gaps concern doctors' perceptions of eHealth resources [33], their responses to parents' retrieved online information [44], and the effects of doctor engagement in the doctor-parent relationship [33]. Searching for health information on the internet can have a positive effect on the doctor-patient relationship among self-seekers [76]. Future studies will need to examine if and under what circumstances this applies to proxy seekers. Subsequently, more research is needed on how pediatricians can support parents in their OHIS behavior (eg, with a proactive conversational approach during appointments or evidence-based leaflets with instructions and links to reputable websites). Unfortunately, studies on doctor-parent communication improvements related to OHIS by proxy or on standardized information leaflets are lacking.

Limitations

This systematic review has several limitations. We included studies from a period of 17 years, while the manner of OHIS has undoubtedly changed much faster. The circumstances in which the studies were conducted may be difficult to compare due to differences at the point of data collection, geographical location, characteristics of the parents, and the underlying diseases of the children. Further, almost all studies were conducted in clinical settings, and the findings in this review may not generalize to other populations. There is a strong need for research on representative samples of parents. Estimates of the prevalence of proxy OHIS should be treated with caution, as it was often not consistently defined in the previous studies, with different time periods being queried and the health status of the child varying.

Conclusions

Our systematic review has important implications for future research. The results suggest that more studies on parental OHIS are needed to understand parental online search behaviors and support parents in their medical decision making by proxy. There is evidence that parental proxy OHIS is a very common but understudied behavior. Our presented agenda has highlighted research gaps that will hopefully lead to more systematic, theoretically informed research in this field.

Acknowledgments

There is no funding to declare. We thank Paulina Slonka for performing the second eligibility coding, Alexander Stoppek for performing the model coding, and Amina Halilović for the final proofread.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 26 KB - jmir_v22i8e19985_app1.docx]

Multimedia Appendix 2

Protocol of the process for selecting studies.
[XLSX File (Microsoft Excel File), 550 KB - jmir_v22i8e19985_app2.xlsx]

Multimedia Appendix 3
Study and sample characteristics.
[XLSX File (Microsoft Excel File), 19 KB - jmir_v22i8e19985_app3.xlsx]

Multimedia Appendix 4
Coding scheme for model of proxy online health information seeking and decision making.
[XLSX File (Microsoft Excel File), 19 KB - jmir_v22i8e19985_app4.xlsx]

References

1. Individuals using the internet for seeking health-related information.: Eurostat URL: http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tin00101&plugin=1 [accessed 2020-08-12]
2. Tan SS, Goonawardene N. Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
3. Eastin MS, Guinsler NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. Cyberpsychol Behav 2006 Aug;9(4):494-498. [doi: 10.1089/cpb.2006.9.494] [Medline: 16901253]
4. White RW, Horvitz E. From health search to healthcare: explorations of intention and utilization via query logs and user surveys. J Am Med Inform Assoc 2014 Jan;21(1):49-55 [FREE Full text] [doi: 10.1136/amiajnl-2012-001473] [Medline: 23666794]
5. Jiang S, Street RL. Pathway Linking Internet Health Information Seeking to Better Health: A Moderated Mediation Study. Health Commun 2017 Dec;32(8):1024-1031. [doi: 10.1080/10410236.2016.1196514] [Medline: 27464036]
6. Fox S, Duggan M. Health Online 2013.: Pew Research Center; 2013 Jan 15. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2020-08-12]
7. Reifegerste D, Bachl M, Baumann E. Surrogate health information seeking in Europe: Influence of source type and social network variables. Int J Med Inform 2017 Jul;103:7-14. [doi: 10.1016/j.ijmedinf.2017.04.006] [Medline: 28551004]
8. Sadasivam RS, Kinney RL, Lemon SC, Shimada SL, Allison JJ, Houston TK. Internet health information seeking is a team sport: analysis of the Pew Internet Survey. Int J Med Inform 2013 Mar;82(3):193-200. [doi: 10.1016/j.ijmedinf.2012.09.008] [Medline: 23149121]
9. Cutrona SL, Mazor KM, Vieux SN, Luger TM, Volkman JE, Finney Rutten LJ. Health information-seeking on behalf of others: characteristics of "surrogate seekers". J Cancer Educ 2015 Mar;30(1):12-19 [FREE Full text] [doi: 10.1007/s13187-014-0701-3] [Medline: 24989816]
10. Selwyn N, Johnson N, Nemorin S, Knight E. Going online on behalf of others: An investigation of 'proxy' internet consumers. Sydney, Australia: Australian Communications Consumer Action Network; 2016. URL: https://accan.org.au/files/Grants/ACCAN_Monash_2016_Going%20online%20on%20behalf%20of%20others_WEB.pdf [accessed 2020-08-12]
11. Robinson TN, Patrick K, Eng TR, Gustafson D. An evidence-based approach to interactive health communication: a challenge to medicine in the information age. Science Panel on Interactive Communication and Health. JAMA 1998 Oct 14;280(14):1264-1269. [doi: 10.1001/jama.280.14.1264] [Medline: 9786378]
12. Abrahamson JA, Fisher KE, Turner AG, Durrance JC, Turner TC. Lay information mediary behavior uncovered: exploring how nonprofessionals seek health information for themselves and others online. J Med Libr Assoc 2008 Oct;96(4):310-323 [FREE Full text] [doi: 10.3163/1536-5050.96.4.006] [Medline: 18974809]
13. Park E, Kim H, Steinhoff A. Health-Related Internet Use by Informal Caregivers of Children and Adolescents: An Integrative Literature Review. J Med Internet Res 2016 Mar 03;18(3):e57 [FREE Full text] [doi: 10.2196/jmir.4124] [Medline: 26940750]
14. Yardi S, Caldwell PH, Barnes EH, Scott KM. Determining parents' patterns of behaviour when searching for online information on their child's health. J Paediatr Child Health 2018 Nov;54(11):1246-1254. [doi: 10.1111/jpc.14068] [Medline: 29864197]
15. Jaks R, Baumann I, Juvalta S, Dratva J. Parental digital health information seeking behavior in Switzerland: a cross-sectional study. BMC Public Health 2019 Feb 21;19(1):225 [FREE Full text] [doi: 10.1186/s12889-019-6524-8] [Medline: 30791927]
16. Oh YS. Predictors of Self and Surrogate Online Health Information Seeking in Family Caregivers to Cancer Survivors. Soc Work Health Care 2015 Dec;54(10):939-953. [doi: 10.1080/00981389.2015.1070780] [Medline: 26671245]
17. Reifegerste D, Czerwinski F, Rosset M, Baumann E, Kludt E, Weg-Remers S. Demographic and cancer-related differences between self-seeking patients and supported patients: Analysis of cancer information-service data. Psychooncology 2019 Apr;28(4):759-766. [doi: 10.1002/pon.5016] [Medline: 30707476]
18. Reifegerste D, Bachl M. Informationssuche als Beziehungstat. Der Zusammenhang zwischen relationalen Faktoren und Motiven der stellvertretenden Suche nach Gesundheitsinformationen. SCM 2019;8(3):378-412. [doi: 10.5771/2192-4007-2019-3-378]

19. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: Systematic review and meta-analysis. J Affect Disord 2019 Dec 15;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
20. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
21. Plantin L, Daneback K. Parenthood, information and support on the internet. A literature review of research on parents and professionals online. BMC Fam Pract 2009 May 18;10:34 [FREE Full text] [doi: 10.1186/1471-2296-10-34] [Medline: 19450251]
22. Dworkin J, Connell J, Doty J. A literature review of parents' online behavior. Cyberpsychology 2013 Jul 01;7(2):1. [doi: 10.5817/CP2013-2-2]
23. Lee HS. Research trends in mothers' health information seeking behaviors: A review of the literature. Proc. Assoc. Info. Sci. Tech 2016 Dec 27;53(1):1-6. [doi: 10.1002/pra2.2016.14505301130]
24. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
25. Johnson J, Meischke H. A Comprehensive Model of Cancer-Related Information Seeking Applied to Magazines. Human Comm Res 1993 Mar;19(3):343-367. [doi: 10.1111/j.1468-2958.1993.tb00305.x]
26. Ruppel EK. Scanning Health Information Sources: Applying and Extending the Comprehensive Model of Information Seeking. J Health Commun 2016;21(2):208-216. [doi: 10.1080/10810730.2015.1058438] [Medline: 26716985]
27. Van Stee SK, Yang Q. Online Cancer Information Seeking: Applying and Extending the Comprehensive Model of Information Seeking. Health Commun 2018 Dec;33(12):1583-1592. [doi: 10.1080/10410236.2017.1384350] [Medline: 29083231]
28. Xiao Z, Lee J, Zeng L, Ni L. Information seeking in the context of cigarette smoking: predictors from the Comprehensive Model of Information Seeking (CMIS). Psychol Health Med 2020 Feb 20:1-19. [doi: 10.1080/13548506.2020.1728348] [Medline: 32077752]
29. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
30. Schlosser RW, Wendt O, Bhavnani S, Nail-Chiwetalu B. Use of information-seeking strategies for developing systematic reviews and engaging in evidence-based practice: the application of traditional and comprehensive Pearl Growing. A review. Int J Lang Commun Disord 2006;41(5):567-582. [doi: 10.1080/13682820600742190] [Medline: 17050471]
31. Cailliau R. A little history of the World Wide Web. URL: https://www.w3.org/History.html [accessed 2020-08-12]
32. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
33. Kasparian NA, Lieu N, Winlaw DS, Cole A, Kirk E, Sholler GF. eHealth literacy and preferences for eHealth resources in parents of children with complex CHD. Cardiol Young 2016 Sep 19;27(4):722-730. [doi: 10.1017/s1047951116001177]
34. AlSaadi MM. Evaluation of internet use for health information by parents of asthmatic children attending pediatric clinics in Riyadh, Saudi Arabia. Ann Saudi Med 2012;32(6):630-636 [FREE Full text] [doi: 10.5144/0256-4947.2012.630] [Medline: 23396028]
35. Sage A, Carpenter D, Sayner R, Thomas K, Mann L, Sulzer S, et al. Online Information-Seeking Behaviors of Parents of Children With ADHD. Clin Pediatr (Phila) 2018 Jan;57(1):52-56. [doi: 10.1177/0009922817691821] [Medline: 28183190]
36. Shah A, Kuo A, Zurakowski D, Waters PM. Use and satisfaction of the internet in obtaining information on brachial plexus birth palsies and its influence on decision-making. J Pediatr Orthop 2006;26(6):781-784. [doi: 10.1097/01.bpo.0000229971.93812.37] [Medline: 17065946]
37. Ikemba CM, Kozinetz CA, Feltes TF, Fraser CD, McKenzie ED, Shah N, et al. Internet use in families with children requiring cardiac surgery for congenital heart disease. Pediatrics 2002 Mar;109(3):419-422. [doi: 10.1542/peds.109.3.419] [Medline: 11875135]
38. Massin MM, Montesanti J, Gerard P. Use of the Internet by Parents of Children with Congenital Heart Disease. Acta Cardiologica 2006 Aug 01;61(4):406-410. [doi: 10.2143/ac.61.4.2017300]
39. Balkhi AM, Olsen B, Lazaroe L, Silverstein J, Geffken GR. Paging Dr. Google: parents' report of internet use for type 1 diabetes management. Diabetes Care 2015 Feb;38(2):e18-e19. [doi: 10.2337/dc14-2461] [Medline: 25614694]
40. Porter A, Edirippulige S. Parents of deaf children seeking hearing loss-related information on the internet: the Australian experience. J Deaf Stud Deaf Educ 2007;12(4):518-529. [doi: 10.1093/deafed/enm009] [Medline: 17456507]
41. Naftel RP, Safiano NA, Falola MI, Shannon CN, Wellons JC, Johnston JM. Technology preferences among caregivers of children with hydrocephalus. PED 2013 Jan;11(1):26-36. [doi: 10.3171/2012.9.peds12208]
42. Baker JF, Devitt BM, Lynch S, Green CJ, Byrne DP, Kiely PJ. Internet use by parents of children attending a dedicated scoliosis outpatient clinic. Eur Spine J 2012 Oct;21(10):1972-1977 [FREE Full text] [doi: 10.1007/s00586-012-2429-2] [Medline: 22820914]

43. Lai C, Mallory SB. Internet use among parents of patients of a pediatric dermatology clinic. Pediatr Dermatol 2000;17(6):493-494. [doi: 10.1046/j.1525-1470.2000.01837-5.x] [Medline: 11123793]
44. Nicholl H, Tracey C, Begley T, King C, Lynch AM. Internet Use by Parents of Children With Rare Conditions: Findings From a Study on Parents' Web Information Needs. J Med Internet Res 2017 Feb 28;19(2):e51 [FREE Full text] [doi: 10.2196/jmir.5834] [Medline: 28246072]
45. Tozzi AE, Mingarelli R, Agricola E, Gonfiantini M, Pandolfi E, Carloni E, et al. The internet user profile of Italian families of patients with rare diseases: a web survey. Orphanet J Rare Dis 2013 May 16;8:76 [FREE Full text] [doi: 10.1186/1750-1172-8-76] [Medline: 23680013]
46. Boston M, Ruwe E, Duggins A, Willging JP. Internet use by parents of children undergoing outpatient otolaryngology procedures. Arch Otolaryngol Head Neck Surg 2005 Aug;131(8):719-722. [doi: 10.1001/archotol.131.8.719] [Medline: 16103305]
47. Hand F, McDowell DT, Mc Dowell DT, Glynn RW, Rowley H, Mortell A. Patterns of internet use by parents of children attending a pediatric surgical service. Pediatr Surg Int 2013 Jul;29(7):729-733. [doi: 10.1007/s00383-013-3317-5] [Medline: 23615872]
48. Nogueira Júnior JF, Hermann DR, Silva MLS, Santos FP, Pignatari SSN, Stamm AC. Is the information available on the Web influencing the way parents see ENT surgical procedures? Braz J Otorhinolaryngol (Impr) 2009 Aug;75(4):517-523. [doi: 10.1590/s1808-86942009000400009]
49. Semere W, Karamanoukian HL, Levitt M, Edwards T, Murero M, D'Ancona G, et al. A pediatric surgery study: parent usage of the Internet for medical information. J Pediatr Surg 2003 Apr;38(4):560-564. [doi: 10.1053/jpsu.2003.50122] [Medline: 12677566]
50. Sim NZ, Kitteringham L, Spitz L, Pierro A, Kiely E, Drake D, et al. Information on the World Wide Web--how useful is it for parents? J Pediatr Surg 2007 Feb;42(2):305-312. [doi: 10.1016/j.jpedsurg.2006.10.003] [Medline: 17270540]
51. Wong M, Sivasegaran D, Choo C, Nah S. Parental Internet Use and Health Information Seeking Behavior Comparing Elective and Emergency Pediatric Surgical Situations. Eur J Pediatr Surg 2018 Feb;28(1):89-95. [doi: 10.1055/s-0037-1604021] [Medline: 28662533]
52. Slomian J, Bruyère O, Reginster J, Emonts P. The internet as a source of information used by women after childbirth to meet their need for information: A web-based survey. Midwifery 2017 May;48:46-52. [doi: 10.1016/j.midw.2017.03.005] [Medline: 28324809]
53. Shroff PL, Hayes RW, Padmanabhan P, Stevenson MD. Internet Usage by Parents Prior to Seeking Care at a Pediatric Emergency Department: Observational Study. Interact J Med Res 2017 Sep 28;6(2):e17 [FREE Full text] [doi: 10.2196/ijmr.5075] [Medline: 28958988]
54. Dhillon AS, Albersheim SG, Alsaad S, Pargass NS, Zupancic JAF. Internet use and perceptions of information reliability by parents in a neonatal intensive care unit. J Perinatol 2003;23(5):420-424 [FREE Full text] [doi: 10.1038/sj.jp.7210945] [Medline: 12847540]
55. Tuffrey C, Finlay F. Use of the internet by parents of paediatric outpatients. Arch Dis Child 2002 Dec;87(6):534-536 [FREE Full text] [doi: 10.1136/adc.87.6.534] [Medline: 12456558]
56. Glynn RW, O'Duffy F, O'Dwyer TP, Colreavy MP, Rowley HM. Patterns of Internet and smartphone use by parents of children attending a pediatric otolaryngology service. Int J Pediatr Otorhinolaryngol 2013 May;77(5):699-702. [doi: 10.1016/j.ijporl.2013.01.021] [Medline: 23434200]
57. Knapp C, Madden V, Marcu M, Wang H, Curtis C, Sloyer P, et al. Information seeking behaviors of parents whose children have life-threatening illnesses. Pediatr Blood Cancer 2011 May;56(5):805-811. [doi: 10.1002/pbc.22674] [Medline: 21370415]
58. McRee A, Reiter PL, Brewer NT. Parents' Internet use for information about HPV vaccine. Vaccine 2012 May 28;30(25):3757-3762 [FREE Full text] [doi: 10.1016/j.vaccine.2011.11.113] [Medline: 22172505]
59. Harvey S, Memon A, Khan R, Yasin F. Parent's use of the Internet in the search for healthcare information and subsequent impact on the doctor-patient relationship. Ir J Med Sci 2017 Nov;186(4):821-826. [doi: 10.1007/s11845-017-1555-6] [Medline: 28130665]
60. Opeoluwa F, Foluke A, Oluwatosin A, Bethel A. The Influence Of The Internet On Health Seeking Behaviour Of Nursing Mothers In Ekiti State, Nigeria. International Journal of Advanced Research and Publications 2017;1(5):1-6 [FREE Full text]
61. Pehora C, Gajaria N, Stoute M, Fracassa S, Serebale-O'Sullivan R, Matava CT. Are Parents Getting it Right? A Survey of Parents' Internet Use for Children's Health Care Information. Interact J Med Res 2015 Jun 22;4(2):e12 [FREE Full text] [doi: 10.2196/ijmr.3790] [Medline: 26099207]
62. Sebelefsky C, Karner D, Voitl J, Klein F, Voitl P, Böck A. Internet health seeking behaviour of parents attending a general paediatric outpatient clinic: A cross-sectional observational study. J Telemed Telecare 2015 May 29;21(7):400-407. [doi: 10.1177/1357633x15583431]
63. Skranes LP, Løhaugen GCC, Botngård A, Skranes J. Internet use among mothers of young children in Norway—a survey of Internet habits and perceived parental competence when caring for a sick child. J Public Health 2014 Jul 2;22(5):423-431. [doi: 10.1007/s10389-014-0631-x]

64. Whyte KL, Hunter I. Internet access, utilisation and perception by parents. Arch Dis Child 2008 May;93(5):448-449. [doi: 10.1136/adc.2008.137851] [Medline: 18426941]
65. Poushter J. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. Pew Research Center. URL: https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2020-08-14]
66. Marton C, Wei Choo C. A review of theoretical models of health information seeking on the web. Journal of Documentation 2012 Apr 20;68(3):330-352. [doi: 10.1108/00220411211225575]
67. Bogg T, Vo PT. Openness, neuroticism, conscientiousness, and family health and aging concerns interact in the prediction of health-related Internet searches in a representative U.S. sample. Front Psychol 2014;5:370 [FREE Full text] [doi: 10.3389/fpsyg.2014.00370] [Medline: 24808880]
68. Sarkadi A, Kristiansson R, Oberklaid F, Bremberg S. Fathers' involvement and children's developmental outcomes: a systematic review of longitudinal studies. Acta Paediatr 2008 Feb;97(2):153-158. [doi: 10.1111/j.1651-2227.2007.00572.x] [Medline: 18052995]
69. Chou WS, Oh A, Klein WMP. Addressing Health-Related Misinformation on Social Media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
70. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
71. Mathes BM, Norr AM, Allan NP, Albanese BJ, Schmidt NB. Cyberchondria: Overlap with health anxiety and unique relations with impairment, quality of life, and service utilization. Psychiatry Res 2018 Mar;261:204-211. [doi: 10.1016/j.psychres.2018.01.002] [Medline: 29324396]
72. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: Systematic review and meta-analysis. J Affect Disord 2019 Dec 15;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
73. Norr AM, Albanese BJ, Oglesby ME, Allan NP, Schmidt NB. Anxiety sensitivity and intolerance of uncertainty as potential risk factors for cyberchondria. J Affect Disord 2015 Mar 15;174:64-69. [doi: 10.1016/j.jad.2014.11.023] [Medline: 25486275]
74. Aiken M, Kirwan G. The psychology of cyberchondria and cyberchondria by proxy. In: Power A, Kirwan G, editors. Cyberpsychology and New Media: A Thematic Reader. London, UK: Psychology Press; 2013:158-169.
75. Thorgaard MV, Frostholm L, Walker L, Jensen JS, Morina B, Lindegaard H, et al. Health anxiety by proxy in women with severe health anxiety: A case control study. J Anxiety Disord 2017 Dec;52:8-14. [doi: 10.1016/j.janxdis.2017.09.001] [Medline: 28950218]
76. Van Riel N, Auwerx K, Debbaut P, Van Hees S, Schoenmakers B. The effect of Dr Google on doctor–patient encounters in primary care: a quantitative, observational, cross-sectional study. BJGP Open 2017 May 16;1(2):bjgpopen17X100833. [doi: 10.3399/bjgpopen17x100833]

Abbreviations

CMIS: comprehensive model of information seeking
OHIS: online health information seeking
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 08.05.20; peer-reviewed by K Scott, J Volkman; comments to author 15.06.20; revised version received 02.07.20; accepted 07.07.20; published 25.08.20.

Please cite as:
Kubb C, Foran HM
Online Health Information Seeking by Parents for Their Children: Systematic Review and Agenda for Further Research
J Med Internet Res 2020;22(8):e19985
URL: http://www.jmir.org/2020/8/e19985/
doi:10.2196/19985
PMID:32840484

Review

Relationship Between Depression and the Use of Mobile Technologies and Social Media Among Adolescents: Umbrella Review

Jorge Arias-de la Torre[1,2,3,4*], MSc, PhD; Elisa Puigdomenech[3,5*], MSc; Xavier García[3], MSc; Jose M Valderas[6], MD, PhD, Prof Dr; Francisco Jose Eiroa-Orosa[7], MSc, PhD; Tania Fernández-Villa[4], MSc, PhD; Antonio J Molina[4], MSc, PhD; Vicente Martín[2,4], MD, PhD, Prof Dr; Antoni Serrano-Blanco[2,8], MD, PhD; Jordi Alonso[2,9,10], MD, PhD, Prof Dr; Mireia Espallargues[3,5], MD, PhD

[1]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[2]CIBER Epidemiología y Salud Pública, Barcelona, Spain

[3]Agency for Health Quality and Assessment of Catalonia, Barcelona, Spain

[4]Institute of Biomedicine, University of Leon, Leon, Spain

[5]Health Services and Chronic Diseases Research Network, Barcelona, Spain

[6]Health Services and Policy Research Group, University of Exeter Medical School, Exeter, United Kingdom

[7]Section of Personality, Assessment and Psychological Treatment, Department of Clinical Psychology and Psychobiology, University of Barcelona, Barcelona, Spain

[8]Parc Sanitari Sant Joan de Déu, Barcelona, Spain

[9]Health Services Research Group, Hospital del Mar Medical Research Institute, Barcelona, Spain

[10]Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona, Spain

[*]these authors contributed equally

Corresponding Author:
Jorge Arias-de la Torre, MSc, PhD
Institute of Psychiatry, Psychology and Neuroscience
King's College London
16 De Crespigny Park
Denmark Hill campus
London, SE5 8AF
United Kingdom
Phone: 44 634722677
Email: jorgeariasdelatorre@gmail.com

Abstract

Background: Despite the relevance of mobile technologies and social media (MTSM) for adolescents, their association with depressive disorders in this population remains unclear. While there are previous reviews that have identified the use of MTSM as a risk factor for developing depression, other reviews have indicated their possible preventive effect.

Objective: The aim of this review was to synthesize the current evidence on the association between MTSM use and the development or prevention of depressive disorders in adolescents.

Methods: An umbrella review was conducted using information published up to June 2019 from PubMed/MEDLINE, PsycINFO, Web of Science, and The Cochrane Library. Systematic reviews focusing on the adolescent population (up to 20 years old) and depression and its potential relationship with MTSM use were included. Screening of titles, abstracts, and full texts was performed. After selecting the reviews and given the heterogeneity of the outcome variables and exposures, a narrative synthesis of the results was carried out.

Results: The search retrieved 338 documents, from which 7 systematic reviews (3 meta-analyses) were selected for data extraction. There were 11-70 studies and 5582-46,015 participants included in the 7 reviews. All reviews included quantitative research, and 2 reviews also included qualitative studies. A statistically significant association between social media and developing depressive symptoms was reported in 2 reviews, while 5 reviews reported mixed results.

Conclusions: Excessive social comparison and personal involvement when using MTSM could be associated with the development of depressive symptomatology. Nevertheless, MTSM might promote social support and even become a point of assistance for people with depression. Due to the mixed results, prospective research could be valuable for providing stronger evidence.

(J Med Internet Res 2020;22(8):e16388) doi:10.2196/16388

KEYWORDS

mobile technologies and social media; depression; adolescents; review

Introduction

Depression is one of the most frequently occurring mental diseases worldwide, generating significant disability, dependence, and expenditure for health systems [1-4]. As shown in previous literature [5-10], adolescence is a particularly relevant period for developing depressive disorders. It should be noted that during adolescence, depressive symptomatology may be broader than in adulthood, manifesting itself through irritability, aggression, avoidance, or other behaviors in addition to the typical depressive behaviors [11]. Furthermore, during this period, young people can be especially influenced by sociocontextual factors, such as the use of mobile technologies and social media (MTSM). However, the effect of the exposure to these technologies on the development of depressive disorders in this age group remains unclear.

The use of MTSM has greatly increased over recent years, particularly since the 1990s, and adolescents can now be considered "digital natives," meaning they have been exposed to mobile devices and technologies like cellphones or tablets since birth [12-14]. This generalized exposure to social media implies a change in the way adolescents interact and communicate, naturally integrating the use of these technologies within their schemes of social perception [15,16]. Therefore, the use of MTSM could be particularly relevant, given the potential influence on adolescents' health, specifically their mental health and the development or prevention of depression.

One of the main uses of MTSM among adolescents is communication and social interaction with their peer groups through various means, including instant messaging apps (eg, WhatsApp and social networks). A few that stand out for their use in this population are Instagram, Snapchat, Twitter, and Facebook [15,17]. Using MTSM could prove beneficial in the sense that they may promote creativity, increase presence and social participation, and provide adolescents with quick access to different types of information, including that related to promoting healthy behaviors and habits [12,13,18]. However, the use of MTSM could also be related to problems like addictive internet behavior, absenteeism and failure in school, deterioration of family relationships and friendships, and different physical and mental health problems (including self-inflicted bodily impairment, eating disorders, and depression) [12,13,19]. Furthermore, MTSM use may also promote behavior that is damaging to health including, among other things, autolytic behavior, suicide, violence, and specific harmful behaviors such as cyberbullying, grooming, or sexting that are derived from the use of these technologies. Despite the abundance of literature, including systematic reviews and meta-analyses, most of the existing evidence is based on cross-sectional studies or surveys. Pooling or synthesizing data and using the broadest possible approach (eg, an umbrella review) could be valuable in determining the current knowledge on whether the use of MTSM is the cause or consequence of depressive symptomatology.

Although there is a wide variety of advantages and disadvantages that the use of new technologies can present for young people, the influence that their use could have on developing depression is unclear. Therefore, the aim of this review was to synthesize the evidence available on the association (intensity and direction) between depression and the use of MTSM in adolescents.

Methods

Study Design and Information Sources

An umbrella review on the association between the use of MTSM and depression was conducted, reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses criteria (PRISMA), and registered in PROSPERO. The following databases were used as sources of information: PubMed/MEDLINE, PsycINFO, Web of Science, and Cochrane Reviews. All documents included in these databases published up to June 2019 were considered.

A search filter (Multimedia Appendix 1) was specifically designed to achieve the study objectives, taking into account pathology, target population, exposure (social network OR social media OR mobile phone OR *phone), and the languages in which the search was performed. After carrying out a preliminary search and observing the number of systematic reviews and meta-analyses found as well as the differences between the studies, an additional filter for study design was included. The filter was designed for PubMed/MEDLINE and adapted for other databases. The search strategy was based on previous studies in other areas with the intention of maximizing the number of identified documents [20,21]. In addition, the references in the final selected studies were used to identify other systematic reviews and meta-analyses, and key authors were contacted.

Inclusion and Exclusion Criteria

The PICO (Population, Intervention, Comparison, and Outcome) criteria were used to identify and include reviews in English that focused on the adolescent population (up to 20 years old), depression (in a broad sense, not specific diagnoses like major depressive disorder or dysthymia), and the possible relationship between depression and the use of MTSM.

Reviews that included studies with participants older than 20 years and studies that did not differentiate the effect by age

group, if they included people older than 20 years, were excluded. Due to difficulties in extrapolating the results for the general adolescent population, studies on genetic or environmental factors and studies carried out in specific population groups, like those with specific characteristics or pathologies (eg, attention deficit hyperactivity disorder), were excluded. Finally, studies focusing on treatments administered through an electronic device or the internet as well as opinion articles and proposals with theoretical or conceptual frameworks that were not based on a systematic literature review or meta-analysis were also excluded.

Review Process

A review of titles, abstracts, and full texts was carried out independently by two expert reviewers (JAT and XG), and discrepancies were resolved by a third researcher (EP) with expertise in conducting systematic reviews. After study selection, a synthesis of the evidence obtained from the 7 selected reviews was carried out. The quality of each review was considered by taking into account the quality of the studies evaluated and the tools used to assess the studies. Owing to heterogeneity in the characteristics of the studies and in the presentation of outcome variables and exposures, a meta-analysis of the results was not possible; therefore, a narrative synthesis of the results was carried out. Information from the included reviews was extracted and summarized in 2 tables of evidence [22].

Results

The search retrieved 338 articles (154 from PubMed, 80 from the Cochrane Library, 41 from PsycINFO, 55 from Web of Science, and 8 from a manual search). After removing 34 duplicates, a total of 304 studies were deemed potentially eligible. The full text of 20 documents was reviewed, and 13 articles were excluded (7 non-systematic or narrative reviews, 5 documents based on other pathologies, and 1 for the inability to differentiate between results reported for adults versus adolescents). Finally, 7 systematic reviews were selected for data extraction (Figure 1) [21,23-30].

Figure 1. Flow diagram of the review process.

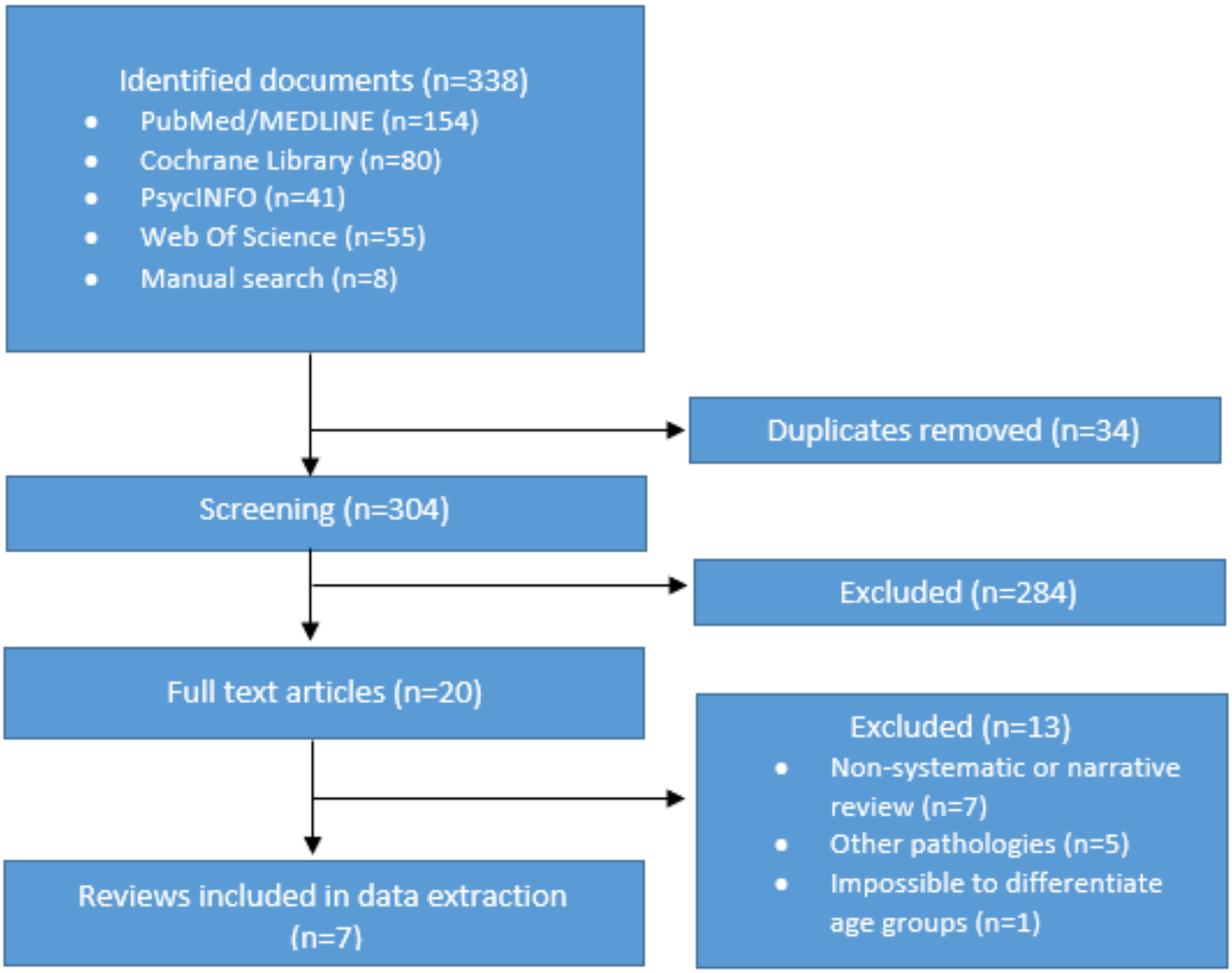


Table 1 shows the characteristics of the included systematic reviews, all of which were published between 2014 and 2019. In these reviews, PsycINFO, Medline, and CINAHL databases were searched most frequently. Two reviews explored dissertations and thesis databases [25,31]. Most reviews assessed the relationship between depression and use of social networks in general [23,25,26,28] or problematic Facebook use in particular [31]. One study by Wu et al [29] reviewed the association between internet use in general and depression. Wellbeing, anxiety, and loneliness were also assessed in 2

reviews [26,29]. There were 11-70 studies and 5582-46,015 participants included in the reviews. Most studies included in the reviews were quantitative and used cross-sectional and survey-based data. While 2 reviews used specific criteria developed by the authors to assess the quality of studies [29,31], 4 used validated assessment tools [21,23,26,28], and 1 did not specify the tool [25]. In addition, 2 meta-analyses were included [25,31].

Table 2 shows the results of the included reviews. Four studies were undertaken specifically with adolescents (age range 10-21 years) [21,23,28,29]. Seabrook et al [26] also included adults in their review (2 studies with adults and 18 studies with the general population), and Marino et al [31] reported a mean age range of 16.5-32.4 years. While 2 reviews reported a positive association between depressive symptoms and social media use (overall random effects pooled estimate: r=0.13, 95% CI 0.05-0.2) [23] and problematic Facebook use (r=0.34, 95% CI 0.28-0.39) [31], the other 5 reviews reported mixed associations between social media use and depression. Keles et al [28] reported a positive association for the relationships between time spent on social media and depression and between social media addiction and depression. Two reviews reported a gender influence with mixed effects [23]. McCrae et al [23] found that 4 studies reported girls having more depressive symptoms related to social media use and 2 studies showed that boys were more likely to show depressive symptoms. The rest of the studies included in their review did not show a gender effect. In the review by Keles et al [28], one study found that social media might have negative effects in girls but could be considered a positive leisure activity for boys, and 2 studies did not show gender effects. In addition to mixed results for the associations between social media use and wellbeing, associations with anxiety and loneliness were also found [21,26,29].

Table 1. Characteristics of the included reviews.

Author (year)	Objective of the review	Databases searched	Number of studies included	Number of participants	Quality assessment of studies included	Methodology
Best et al (2014) [21]	To assess the impact of social media use on mental wellbeing in young people	ASSIA[a], Communication abstracts, CINAHL, ERIC[b], Medline (Ovid), PsycINFO (Ovid), SCOPUS, SSCI[c]	43	NS[d]	Specific criteria developed by the authors of the review	32 quantitative, 9 qualitative, 2 mixed methods or others
Wu et al (2016) [29]	To examine the association between internet use, social connection, and levels of depression, anxiety, and loneliness	CINAHL, ERIC, Psychology and Behavioral Series Collection, Science and Technology Collection, EBSCO social sciences database	12	5582	Specific criteria developed by the authors of the review	9 quantitative (all cross-sectional), 1 mixed methods, and 2 qualitative
Seabrook et al (2016) [26]	To examine the relationship between the use of social networks and depression and anxiety as well as links with wellbeing and potential mediators and moderators of these relationships	PsycINFO, MEDLINE (Ovid), Scopus, IEEE Xplore, CINAHL, Education Resources Information Center, SSCI, Communication and Mass Media Complete	70	46,015	Adaptation of the Cochrane bias tool	NS
McCrae et al (2017) [23]	To examine the association between social media (websites used primarily for social interaction) and depression or depressive symptoms	Medline, PsycINFO, EMBASE	11	12,646	Robins-I[e], Cochrane Collaboration Methods Group Tool to assess risk of bias in cohort studies	Quantitative (7 cross-sectional, 4 longitudinal)
Marino et al (2018) [31]	To examine the association between Facebook use (problematic, abusive, overuse, compulsive) and psychological disorders in adolescents and young adults	PsycINFO, PubMed, Scopus, ResearchGate, Google Scholar, Dissertation Abstracts International, Pro-Quest Dissertations and Theses Open, Open Access Theses and Dissertations	23	13,929	Specific criteria developed by the authors of the review	Quantitative
Keles et al (2019) [28]	To examine the influence of using social networks on depression in adolescents	PsycINFO, Medline, EMBASE, CINAHL, SSCI	13	21,231	NIH[f]	Quantitative (12 cross-sectional, 1 longitudinal)
Yoon et al (2019) [25]	To examine the relationship between the use of social networking sites and depression	PsycINFO, PubMed, ProQuest Dissertations & Theses Global	55	22,099	NS	Quantitative

[a]ASSIA: Applied Social Sciences Index and Abstracts.

[b]ERIC: Education Resources Information Center.

[c]SSCI: Social Sciences Citation Index.

[d]NS: not specified.

[e]Risk of bias tool to assess nonrandomized studies of interventions.

[f]NIH: National Institutes of Health Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies.

Table 2. Results of the included reviews.

Author (year)	Sample (number of studies) or age (years)	Use of MTSM[a]	Association(s)	Gender effect	Other associations
Best et al (2014) [21]	Adolescents (age range not specified)	Communication and social interaction	Mixed results in the association of social media technologies and depression	Does not distinguish nor consider this factor	Mixed results on self-esteem, social support, loneliness, and cyberbullying
Wu et al (2016) [29]	10-21	Use of internet and related technologies	1 of 5 studies found that social media technology use can lead to depressive feelings; 4 of 5 studies did not find an association.	Takes into account the population of the studies (10 mixed gender, 2 only boys), but not in terms of the results	Mixed results on social connectivity, anxiety, and loneliness
Seabrook et al (2016) [26]	Adolescents (8), young adults (40), general population (18), adults (2), clinical depression (1), others (1)	Use of social networks	Mixed results: positive interactions, social support, and connectivity in social networks related with lower levels of depression; negative interactions and social comparison related with higher levels of depression	Not considered as a variable in the included studies but considered in the discussion of the results	Mixed results for anxiety and wellbeing
McCrae et al (2017) [23]	10-17 (one study included "high school students" but did not specify age range)	Use of social media	Small but statistically significant overall correlation between social media use and depressive symptoms	4 studies found that girls had more depressive symptoms related to social media use; 2 studies showed that boys were more likely to show depressive symptoms; the rest showed no gender differences	NS[b]
Marino et al (2018) [31]	Mean 21.9 (SD 3.97); 16.5-32.4 (mean age range)	Problematic Facebook use	Association between problematic Facebook use and depression	Proportion of girls (60.7%) did not moderate the effect	Correlation between problematic Facebook use and psychological distress was greater in samples with a higher mean age.
Keles et al (2019) [28]	13-18	Time spent, activity (quality and quantity of user's engagement and interaction with social media sets and other users), investment (time spent on social media), addiction (state of being dependent on social media)	Time spent: 1 study showed association, 1 did not, 2 did not find association; activity: 2 studies showed positive association, and 1 did not; investment: 3 studies showed association; addiction: 3 studies showed positive association	4 studies measured the effect of gender between social media–related variables and mental health outcomes. 2 studies did not find effects on gender, while 1 found that social media might have negative effects in girls and can be considered a positive leisure activity for boys. Facebook had a negative impact on both genders.	There was a relationship between age, heavy social media use, and negatively internalizing symptoms. Younger adolescents were more likely to experience internalizing symptoms (being anxious, depressed, withdrawn). Most studies highlighted the fact that the relationships observed were too complex for straightforward statements and mediating and moderating factors should be taken into account.
Yoon et al (2019) [25]	17.83-24.76 (mean age range)	Use of SNS[c]: time spent and SNS checking; social comparison and "upward" social comparison	Positive statistically significant difference between depression and time spent on its use, frequency of use, social comparison, and "upwards" comparison	No difference	NS

[a]MTSM: mobile technologies and social media.

[b]NS: not specified.

[c]SNS: social networking sites.

Discussion

The results from the included reviews suggest that social comparison and excessive personal involvement by adolescents when using MTSM could be related to the development of depressive symptoms. However, the use of MTSM when properly adapted could also promote healthy behaviors, improve social support, and even become a point of access of information and help for adolescents at risk of depression.

Both mobile technologies and social media are important aspects of how we interact today and have transformed the way in which the generations adopting MTSM and digital natives communicate [18,32]. The use of MTSM presents great opportunities in terms of creativity and ways of learning but can also entail certain risks such as isolation and restricted social interaction. Despite this, studying the possible effects on health, specifically on depression, of adolescents using MTSM is a relatively recent phenomenon. As such, it should be noted that all reviews included in this study were published in the last 5 years.

The evidence from different studies published until now, and particularly since 2017, suggests a positive and significant association between some aspects of social media use and the presence of depressive symptoms among adolescents [23-25]. Two relevant factors that increased the magnitude of this association were the problematic use of social networks and excessive social comparison [23-25]. There is less relevant evidence pointing to other factors related to the undesirable effects of social networks, like a higher level of personal involvement on the networks, defined as the degree of exposure and personal information that adolescents publish on networks or the exposure to content that promotes depressive-like behaviors [27,28]. Finally, it is worth mentioning that a high volume of studies indicating associations between the use of social networks and other undesirable effects like anxiety, harassment, or internet or smartphone addiction was identified [21,26,28-30]. Regarding internet addiction, the total usage time, frequency of consultation, and other variables related to excess use, both in frequency and time, may be more relevant than the variables found in this study, which focus specifically on depressive symptomatology.

It should be noted that the impact of the identified factors, particularly of social comparison, on the development of depression might be affected by the level of the welfare and wealth of the family [7-12]. Accordingly, those who are from families with lower socioeconomic status might have a high risk of developing depression when exposed to more wealthy people. In addition, these factors might be particularly related to the development of some specific depressive symptoms (eg, sleep problems or diminished ability to think or concentrate). Further longitudinal research focused on specific factors, like family environment, and accounting for specific depressive symptoms might be valuable in preventing the potential development of depression in MTSM users.

Emphasizing the fact that social networks do not necessarily imply a negative impact on young people's moods, other studies have described the desirable effects that social media use might have [12,33,34]. In this sense and in line with the results of these studies, the evidence found in this study suggests that social networks can promote social support and even become points of access to information and help for people with depressive disorders [26,27]. As suggested, the use of MTSM under adult supervision might be related to promoting healthy use of MTSM, as well as preventing possible negative consequences that arise like depressive symptomatology [35]. In addition, the use of new technologies could facilitate young people's connection with multiple social circles, reducing their perception of loneliness or isolation [29].

Some studies identified differences between boys and girls in the impact that social networks have on developing depressive symptoms. Previous research proposed [30] that the prevalence of intensive use of mobile technologies might be greater in women than in men. Furthermore, the use of mobile technologies could be mainly for relational purposes among teenage women and instrumental or for leisure among teenage men, making women more likely to be exposed to the effects of social networks [23,24,28]. Although the meta-analysis by McCrae et al [23] did not determine a theoretical basis for the potential differences, there are some studies included within the analysis and 1 study included in the systematic review by Keles et al [28] that show a greater correlation between social comparison and depression in women. This might allow us to hypothesize that focusing preventive measures on social comparison in adolescent women and on leisure platforms, like gaming platforms, in adolescent men could be effective in preventing the undesirable effects of social networks and mobile technology use among adolescents. Further research aimed at proving this hypothesis could be valuable.

Several limitations of the current study deserve discussion. First is the lack of longitudinal or experimental evidence in relation to the use of social networks and mobile technologies and their impact on depressive symptomatology. In this sense, most of the studies included in the literature reviewed were cross-sectional and survey-based, precluding the establishment of causal relationships between variables. As such, it is difficult to determine whether the use of social networks and mobile technologies is the cause or consequence of depressive symptomatology, and further longitudinal studies to test these hypotheses could be valuable. We should also mention the possible heterogeneity of health problems and of the patterns made or activities observed in the studies when using MTSM. While some were focused on clinical depression diagnosed by a professional, others were focused on less valid depression criteria, which could limit the comparability of the reviews included. Furthermore, some of the reviews included internet addicts. Despite this, the broad aim of this review was to determine the relationship between depressive disorders and the use of MTSM, which we consider completed through the studies included in this article, independent of the depression metrics and specific populations used in the selected reviews.

Another limitation is the lack of solid evidence or a conceptual framework on the specific behaviors, like online gaming or uploading photos to social networks, that could be related to depressive symptomatology. This lack of evidence may be due to the relative novelty of the social network phenomenon and

the shortage of valid, reliable health information pertaining to it. However, certain behaviors that could be related to the development of depression as a protective factor were identified, like searching for help or preventive information. Another limitation is that only reviews in English were included, possibly omitting scientific literature written in other languages. Finally, we should mention the limitation of having actively excluded studies on cyberbullying, addiction to new technologies, or other symptoms and harmful behaviors that could be part of or related to a depressive disease. Given their importance and the abundance of evidence on these phenomena, these behaviors deserve to be treated as separate entities, and, as previous research suggests, specific reviews should be performed on these behaviors [21,26].

In conclusion, our study shows that, during adolescence, the use of MTSM and particularly excessive social comparison and personal involvement when using it could be associated with developing depressive symptomatology. Nevertheless, the adaptive use of MTSM could also help prevent the development of depression, promote social support, and even become a point of information access and help for people with depressive disorders or symptoms. Other variables, like time spent on the internet and social networks, the frequency of consultation, and factors related to excess use, both in frequency and in time, may be more relevant in developing other problems like internet addiction. Due to the heterogeneity in methodology and the contradictory findings from the reviews included in this umbrella review, prospective research, especially longitudinal cohort studies and randomized controlled trials, could be valuable in providing stronger evidence on these relationships.

Acknowledgments

We acknowledge Antoni Parada and Kayla Smith for their support and CIBER Epidemiology and Public Health (CIBERESP) for its funding.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PubMed/MEDLINE filter.
[DOCX File , 12 KB - jmir_v22i8e16388_app1.docx]

References

1. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet 2017 Sep;390(10100):1211-1259. [doi: 10.1016/S0140-6736(17)32154-2]
2. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. The Lancet 2007 Sep;370(9590):851-858. [doi: 10.1016/S0140-6736(07)61415-9]
3. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. The Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/s2215-0366(15)00505-2]
4. Arias-de la Torre J, Vilagut G, Martín V, Molina AJ, Alonso J. Prevalence of major depressive disorder and association with personal and socio-economic factors. Results for Spain of the European Health Interview Survey 2014-2015. J Affect Disord 2018 Oct 15;239:203-207. [doi: 10.1016/j.jad.2018.06.051] [Medline: 30014961]
5. Thapar A, Collishaw S, Pine DS, Thapar AK. Depression in adolescence. The Lancet 2012 Mar;379(9820):1056-1067. [doi: 10.1016/S0140-6736(11)60871-4]
6. Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A. Prevalence and development of psychiatric disorders in childhood and adolescence. Arch Gen Psychiatry 2003 Aug 01;60(8):837-844. [doi: 10.1001/archpsyc.60.8.837] [Medline: 12912767]
7. Arias-de la Torre J, Fernández-Villa T, Molina A, Amezcua-Prieto C, Mateos R, Cancela J, et al. Psychological Distress, Family Support and Employment Status in First-Year University Students in Spain. Int J Environ Res Public Health 2019 Apr 04;16(7):1209 [FREE Full text] [doi: 10.3390/ijerph16071209] [Medline: 30987309]
8. Arnett JJ. Emerging adulthood. A theory of development from the late teens through the twenties. Am Psychol 2000 May;55(5):469-480. [Medline: 10842426]
9. Harris A. Mental health and moving from school to further and higher education. London, England: Center for Mental Health; 2019.
10. Johnson D, Dupuis G, Piche J, Clayborne Z, Colman I. Adult mental health outcomes of adolescent depression: A systematic review. Depress Anxiety 2018 Aug 07;35(8):700-716. [doi: 10.1002/da.22777] [Medline: 29878410]
11. Lewis M, Rudolph K, editors. Handbook of Developmental Psychopathology, Third Edition. New York, NY: Springer; 2014.
12. Hur J, Gupta M. Growing up in the Web of Social Networking: Adolescent Development and Social Media. APS 2013 Aug 01;3(3):233-244. [doi: 10.2174/2210676611303030004]

13. Pantic I. Online social networking and mental health. Cyberpsychol Behav Soc Netw 2014 Oct;17(10):652-657 [FREE Full text] [doi: 10.1089/cyber.2014.0070] [Medline: 25192305]
14. Sánchez-Martínez M, Otero A. Factors associated with cell phone use in adolescents in the community of Madrid (Spain). Cyberpsychol Behav 2009 Apr;12(2):131-137. [doi: 10.1089/cpb.2008.0164] [Medline: 19072078]
15. Smith A, Anderson M. Social Media Use 2018: Demographics and Statistics | Pew Research Center. Pew Res Cent Internet Technol 2018. [doi: 10.1007/978-1-349-70154-4_163]
16. Rich M. Moving from child advocacy to evidence-based care for digital natives. JAMA Pediatr 2014 May;168(5):404-406. [doi: 10.1001/jamapediatrics.2014.55] [Medline: 24687218]
17. Richards D, Caldwell PH, Go H. Impact of social media on the health of children and young people. J Paediatr Child Health 2015 Nov 26;51(12):1152-1157. [doi: 10.1111/jpc.13023]
18. Reid Chassiakos YL, Radesky J, Christakis D, Moreno M, Cross C, Council on Communications and Media. Children and Adolescents and Digital Media. Pediatrics 2016 Nov;138(5):e1-e18 [FREE Full text] [doi: 10.1542/peds.2016-2593] [Medline: 27940795]
19. Hussain Z, Griffiths MD. Problematic Social Networking Site Use and Comorbid Psychiatric Disorders: A Systematic Review of Recent Large-Scale Studies. Front Psychiatry 2018 Dec;9:686 [FREE Full text] [doi: 10.3389/fpsyt.2018.00686] [Medline: 30618866]
20. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009 Aug 27;18(8):1115-1123. [doi: 10.1007/s11136-009-9528-5]
21. Best P, Manktelow R, Taylor B. Online communication, social media and adolescent wellbeing: A systematic narrative review. Children and Youth Services Review 2014 Jun;41:27-36. [doi: 10.1016/j.childyouth.2014.03.001]
22. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol 2011 Feb 03;11(1):15 [FREE Full text] [doi: 10.1186/1471-2288-11-15] [Medline: 21291558]
23. McCrae N, Gettings S, Purssell E. Social Media and Depressive Symptoms in Childhood and Adolescence: A Systematic Review. Adolescent Res Rev 2017 Mar 2;2(4):315-330. [doi: 10.1007/s40894-017-0053-4]
24. Camporesi E, Vezzani G, Zanon V, Manelli D, Enten G, Quartesan S, et al. Review on hyperbaric oxygen treatment in femoral head necrosis. UHM 2017 Nov 1;44(6):497-508. [doi: 10.22462/11.12.2017.1]
25. Yoon S, Kleinman M, Mertz J, Brannick M. Is social network site usage related to depression? A meta-analysis of Facebook–depression relations. Journal of Affective Disorders 2019 Apr;248:65-72. [doi: 10.1016/j.jad.2019.01.026]
26. Seabrook EM, Kern ML, Rickard NS. Social Networking Sites, Depression, and Anxiety: A Systematic Review. JMIR Ment Health 2016 Nov 23;3(4):e50. [doi: 10.2196/mental.5842]
27. Memon AM, Sharma SG, Mohite SS, Jain S. The role of online social networking on deliberate self-harm and suicidality in adolescents: A systematized review of literature. Indian J Psychiatry 2018;60(4):384-392 [FREE Full text] [doi: 10.4103/psychiatry.IndianJPsychiatry_414_17] [Medline: 30581202]
28. Keles B, McCrae N, Grealish A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. International Journal of Adolescence and Youth 2019 Mar 21;25(1):79-93. [doi: 10.1080/02673843.2019.1590851]
29. Wu YJ, Outley C, Matarrita-Cascante D, Murphrey TP. A Systematic Review of Recent Research on Adolescent Social Connectedness and Mental Health with Internet Technology Use. Adolesc Res Rev 2016;1(2):153-162. [doi: 10.1007/s40894-015-0013-9]
30. Marchant A, Hawton K, Stewart A, Montgomery P, Singaravelu V, Lloyd K, et al. A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown. PLoS ONE 2017 Aug 16;12(8):e0181722. [doi: 10.1371/journal.pone.0181722]
31. Marino C, Gini G, Vieno A, Spada MM. The associations between problematic Facebook use, psychological distress and well-being among adolescents and young adults: A systematic review and meta-analysis. Journal of Affective Disorders 2018 Jan;226:274-281. [doi: 10.1016/j.jad.2017.10.007]
32. Moreno MA, Goniu N, Moreno PS, Diekema D. Ethics of social media research: common concerns and practical considerations. Cyberpsychol Behav Soc Netw 2013 Sep;16(9):708-713 [FREE Full text] [doi: 10.1089/cyber.2012.0334] [Medline: 23679571]
33. Van Rode V, Rotsaert M, Delhaye M. [Loneliness and adolescence: clinical implications and outlook. Literature review]. Rev Med Brux 2015;36(5):415-420. [Medline: 26749631]
34. Grist R, Porter J, Stallard P. Mental Health Mobile Apps for Preadolescents and Adolescents: A Systematic Review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
35. Moreno M, Pumper M. Adolescent Health Displays. In: Rosen LD, Cheever NA, Carrier LM, editors. Wiley Handbook of Psychology, Technology, and Society. Hoboken, NJ: John Wiley & Sons; 2015:287-300.

Abbreviations

ASSIA: Applied Social Sciences Index and Abstracts
ERIC: Education Resources Information Center
MTSM: mobile technologies and social media
NIH: National Institutes of Health Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies
NS: not specified
PICO: Population, Intervention, Comparison and Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses criteria
SSCI: Social Sciences Citation Index

Edited by G Eysenbach; submitted 25.09.19; peer-reviewed by KB Lankarani, D Saleh, C Summers, A Benetoli; comments to author 14.05.20; revised version received 21.05.20; accepted 03.06.20; published 26.08.20.

Please cite as:
Arias-de la Torre J, Puigdomenech E, García X, Valderas JM, Eiroa-Orosa FJ, Fernández-Villa T, Molina AJ, Martín V, Serrano-Blanco A, Alonso J, Espallargues M
Relationship Between Depression and the Use of Mobile Technologies and Social Media Among Adolescents: Umbrella Review
J Med Internet Res 2020;22(8):e16388
URL: http://www.jmir.org/2020/8/e16388/
doi:10.2196/16388
PMID:32663157

Review

Perceptions and Experiences of Internet-Based Testing for Sexually Transmitted Infections: Systematic Review and Synthesis of Qualitative Research

Tommer Spence[1], BSc; Inès Kander[1], BSc, MSc; Julia Walsh[1], BA, MSc; Frances Griffiths[1], BA, MBBS, PhD; Jonathan Ross[2], MBChB, MD, FRCP

[1]Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, United Kingdom

[2]Whittall Street Clinic, University Hospitals Birmingham National Health Service Foundation Trust, Birmingham, United Kingdom

Corresponding Author:
Tommer Spence, BSc
Division of Health Sciences
Warwick Medical School
University of Warwick
Gibbet Hill Road
Coventry, CV4 7AL
United Kingdom
Phone: 44 (0)2476 528009
Email: T.Spence@warwick.ac.uk

Abstract

Background: Internet-based testing for sexually transmitted infections (STIs) allows asymptomatic individuals to order a self-sampling kit online and receive their results electronically, reducing the need to attend a clinic unless for treatment. This approach has become increasingly common; however, there is evidence that barriers exist to accessing it, particularly among some high-risk populations. We review the qualitative evidence on this topic, as qualitative research is well-placed to identify the complex influences that relate to accessing testing.

Objective: This paper aims to explore perceptions and experiences of internet-based testing for STIs among users and potential users.

Methods: Searches were run through 5 electronic databases (CINAHL, EMBASE, MEDLINE, PsycINFO, and Web of Science) to identify peer-reviewed studies published between 2005 and 2018. Search terms were drawn from 4 categories: STIs, testing or screening, digital health, and qualitative methods. Included studies were conducted in high-income countries and explored patient perceptions or experiences of internet-based testing, and data underwent thematic synthesis.

Results: A total of 11 studies from the 1735 studies identified in the initial search were included in the review. The synthesis identified that internet-based testing is viewed widely as being acceptable and is preferred over clinic testing by many individuals due to perceived convenience and anonymity. However, a number of studies identified concerns relating to test accuracy and lack of communication with practitioners, particularly when receiving results. There was a lack of consensus on preferred media for results delivery, although convenience and confidentiality were again strong influencing factors. The majority of included studies were limited by the fact that they researched hypothetical services.

Conclusions: Internet-based testing providers may benefit from emphasizing this testing's comparative convenience and privacy compared with face-to-face testing in order to improve uptake, as well as alleviating concerns about the self-sampling process. There is a clear need for further research exploring in depth the perceptions and experiences of people who have accessed internet-based testing and for research on internet-based testing that explicitly gathers the views of populations that are at high risk of STIs.

Trial Registration: PROSPERO CRD42019146938; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=146938

(J Med Internet Res 2020;22(8):e17667) doi:10.2196/17667

KEYWORDS

sexually transmitted infections; self-sampling; screening; testing; internet; digital health; eHealth; qualitative research; thematic synthesis

Introduction

Background

Sexually transmitted infections (STIs) are a serious public health problem, with the incidence of many infections rising rapidly [1-3]. In England, syphilis diagnoses have risen 126% in the past 5 years, and gonorrhea diagnoses rose 26% in a year from 2017 to 2018 [3,4]. This statistic is of particular concern, given the increasing risk of antibiotic-resistant gonorrhea [3,5].

One of the challenges in preventing the spread of STIs is that they frequently remain asymptomatic [6,7]. This allows them to be spread unknowingly and increases the likelihood of developing complications such as pelvic inflammatory disease and infertility from chlamydia and gonorrhea and damage to the heart, bones, and central nervous system from syphilis [5,8,9].

Screening for STIs is therefore crucial in tackling their impact, ensuring that people are treated soon after infection, and reducing the risk of passing the infection onto others. It is well established, however, that numerous barriers exist to accessing testing, including stigma, aversion to the sampling process, or the time and travel required to access clinics [10-12]. This contributes to low uptake of testing, identified as an obstacle to reducing STI prevalence in a number of countries, including Australia, England, France, and the United States [1,3,13,14].

One new method to improve access to and uptake of STI screening is internet-based testing. Its use has grown rapidly in recent years, and it now accounts for over 17% of chlamydia tests undertaken by young people in England [3]. This figure is likely to continue rising, in part due to increased provision, as the cost effectiveness of internet-based testing mitigates considerable cuts to the budgets of sexual health services seen since 2013 [15]. Although variations exist between internet-based testing services, they almost all involve users ordering a self-sampling kit online, which they then return to a laboratory for testing before receiving their results remotely [16,17]. Common media for results delivery include SMS text messaging, email, phone, mail, and websites [16].

Although internet-based testing appears to address many of the barriers users face in accessing traditional face-to-face testing, there is a lack of conclusive data on how it is perceived or experienced. The existing systematic reviews focusing exclusively on self-sampling for STIs have found it to be acceptable, but these were not limited to internet-based testing and only included the views of people who had already accessed self-sampling [18-20]. Other reviews of attitudes towards STI testing have found that participants identified waiting times and clinic opening hours as examples of barriers to accessing face-to-face testing, but again these studies did not focus exclusively on internet-based testing and reported only limited data on self-sampling [21-23]. This review seeks to fill this gap and develop the understanding of how internet-based testing specifically is perceived and experienced. It focuses on qualitative research, as this approach is uniquely well placed to aid nuanced analysis of people's engagement with sexual health services [24].

Review Question

This review aims to answer the following question: What are the perceptions and experiences of internet-based testing for STIs among users and potential users?

Methods

The review protocol was registered on PROSPERO during the review process (identification number CRD42019146938) [25].

Search Strategy

The search used 5 electronic databases that specialize in health research: MEDLINE, EMBASE, CINAHL, Web of Science, and PsycINFO. The search terms were developed through experimentation with the support of a specialist librarian, using a population, intervention, context, and outcome model adapted for qualitative research [26,27]. This resulted in the following 4 search term categories: (1) Population: individuals with or at risk of STIs (eg, chlamydia); (2) Intervention: testing or screening for STIs (eg, test); (3) Context: online (eg, internet); and (4) Outcome: qualitative perception or experience (eg, interview).

An example list of search terms is included in Multimedia Appendix 1.

The search period spanned from January 1, 2005, to December 31, 2018. We chose 2005 because this was when internet-based testing emerged, and fewer than half of UK households had access to the internet prior to this period [28,29]. The search was limited to studies published in English, as there were insufficient resources available to arrange translation.

Eligibility

Studies were eligible for inclusion if they (1) reported user (or prospective user) perceptions or experiences of any aspect of internet-based testing for STIs, either hypothetically or in practice, and how this affected whether users might access it; (2) collected the relevant data using qualitative methods, including the qualitative component of mixed methods research and free text responses to questionnaires; (3) were published in English in peer-reviewed academic journals between 2005 and 2018; and (4) collected data in countries defined as high-income countries by the World Bank.

Inclusion was limited to high-income countries due to their similar STI profiles, health care infrastructure, and rates of internet access [30,31]. Self-sampling for HPV was not included within the scope of this review, as it is normally conducted in the context of cancer screening and is not usually a component of STI screening [32].

Screening

The search results from each database were combined and duplicates were removed. Initially, all studies were screened via their titles and abstracts to determine eligibility, with potentially eligible studies subsequently being read in full. These stages of the screening process were both undertaken in full by the lead reviewer (TS), with a second reviewer (IK) screening a random 20% (267/1332 for title and abstract screening; 16/79 for full-text screening) of studies at each stage to determine interrater reliability. Any disagreements were resolved through discussion, and a third reviewer was brought in if consensus could not be reached. Reviewers always erred on the side of inclusion to ensure all relevant data were identified.

Quality Assessment

The quality of the studies was critically appraised using the Critical Appraisal Skills Programme checklist for qualitative research [33]. The checklist provides a holistic overview of the rigor of research, and any studies found to be methodologically weak were planned to be included in a sensitivity analysis once the synthesis was complete.

Data Extraction and Synthesis

The included studies underwent thematic synthesis following an adapted version of a framework developed by Thomas and Harden [34] for qualitative systematic reviews. The results section of each study was uploaded to NVivo (QSR International), with each line of text that addressed the research question being coded according to its meaning and content. All relevant data were coded at least once, and once this process was complete, codes were grouped into descriptive themes that captured the meaning of multiple initial codes. These were in turn developed into broader analytical themes, which go beyond the findings of the included studies. The lead reviewer (TS) undertook all analysis, with support and interrater reliability being provided on 20% of included studies by a second reviewer (JW) and feedback being provided from the lead reviewer's supervisors (FG and JR).

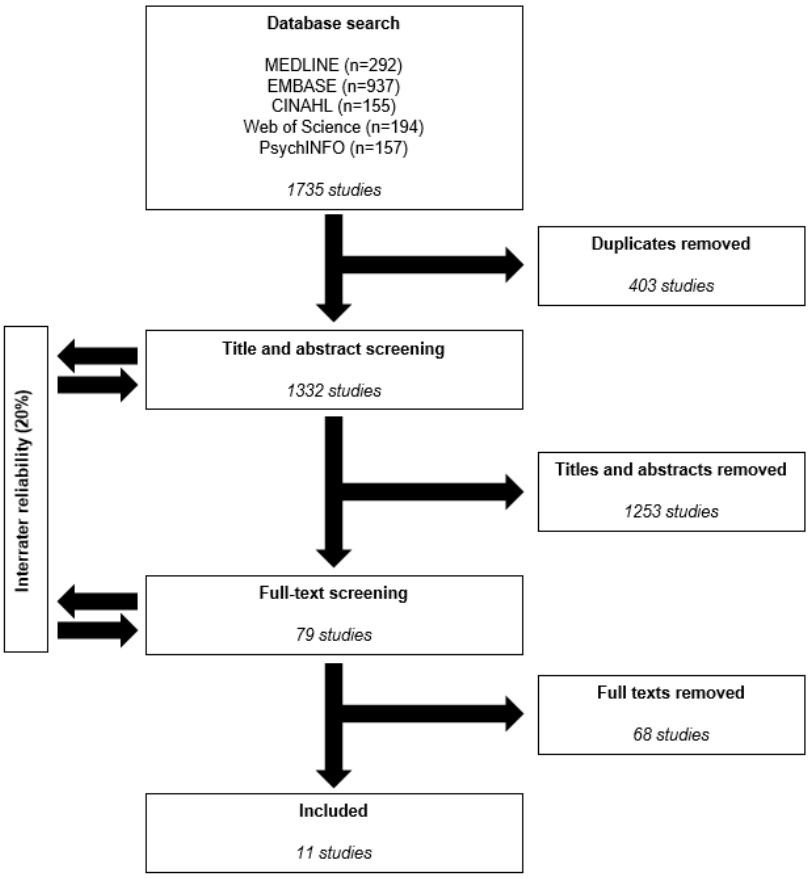
Results

Search

A total of 11 studies met the inclusion criteria for the review. The initial search identified 1735 studies for screening, which reduced to 1332 studies once duplicates were removed. The search and screening process is outlined in Figure 1, along with the number of studies removed at each stage in the process.

The initial rate of agreement between reviewers was 96.2% for title and abstract screening and 93.3% for full-text screening.

Figure 1. Summary of literature search process.



Study Characteristics

An overview of the included studies is provided in Table 1. All were published between 2006 and 2018. One was conducted in Australia, with the remainder split equally between the United Kingdom and United States.

Table 1. Characteristics of included studies.

Lead author	Year	Country	Aim	Participants	Study size	Average age (age range)	Actual or potential users	Methods
Ahmed-Little [35]	2016	United Kingdom	To explore attitudes of participants towards a pilot internet-based HIV-testing program	People who provided a free-text response within a questionnaire	756	Mean 22.6 (16-72)	Actual	Free text space in a feedback questionnaire, thematically analyzed
Baraitser [17]	2015	United Kingdom	To obtain stakeholder input on a theory of change for online sexual health services	Potential service users sampled to include men who have sex with men, young people, and ethnic minorities	4	—[a]	Potential	Interviews, analyzed using a framework approach
Friedman [36]	2013	United States	To explore young women's technology use, preferences for STD[b]-testing venues, attitudes toward nontraditional venues, and acceptability of different test results delivery methods	Women (35% Black, 34% Hispanic, 31% White), recruited through market research firms	80	— (15-25)	Potential	Ethnographic semistructured interviews, thematically analyzed
Gaydos [37]	2006	United States	To inform the design of an effective educational website that facilitates self-sampling and is appealing to women	Women (57% Black, 2% Hispanic, 43% White), recruited via educational institutions	42	— (14-49)	Potential	Focus groups, reported
Gibbs [38]	2018	United Kingdom	To evaluate the results component of a pilot online sexual health service	People purposively sampled from users of the pilot service	36	— (18-35)	Actual	In-depth interviews, analyzed using a framework approach
Lorimer [39]	2013	United Kingdom	To inform the design of an internet-based approach to chlamydia screening targeting young men	Men recruited from the community	60	— (16-24)	Potential	Focus groups, analyzed using a framework approach
Roth [40]	2011	United States	To explore preferences for accessing STI[c]-screening services among men	Men (55% Black, 14% Hispanic, 31% White) recruited in a sexual health clinic	29	Median 34 (19-60)	Potential	Interviews and focus groups, thematically analyzed
Stahlman [41]	2015	United States	To explore attitudes towards potential interventions to increase testing and reduce transmission among MSM[d] with repeat syphilis infection	MSM (16% Black, 32% Hispanic, 53% White; 68% HIV+), recruited via local government public health database	19	Mean 38 (21-54)	Potential	Semistructured interviews, analyzed using grounded theory and thematically via axial coding
Tobin [42]	2018	United States	To assess the acceptability and feasibility of a program to train young Black MSM to use and promote HIV and STI home testing to their social network	Young Black MSM (2 self-reported HIV+) recruited from the community	15	Mean 26.2 (—)	Potential	2 in-depth structured interviews, 1 week apart, reviewed for range, consensus, and divergence of responses

Lead author	Year	Country	Aim	Participants	Study size	Average age (age range)	Actual or potential users	Methods
Tomnay [43]	2014	Australia	To examine rural young people's perceptions of barriers and facilitators to using face-to-face and online sexual health testing and treatment	Young people recruited from the community	50	— (16-25)	Potential	Focus groups, thematically analyzed
Wayal [44]	2011	United Kingdom	To inform the development of a service offering home sampling kits for STI/HIV	MSM (4% Black, 92% White, 4% Asian; 17% HIV+), recruited from a sexual health clinic	24	Median 39 (22-68)	Potential	Focus groups, analyzed using a framework approach

[a]Not available.

[b]STD: sexually transmitted disease.

[c]STI: sexually transmitted infection.

[d]MSM: men who have sex with men.

Of the 11 studies included, 2 studies reported the experience of users who had accessed internet-based testing, with the remaining 9 exploring perceptions among potential users of hypothetical services. Of these, 2 studies explored the views of women, 2 explored the views of men who have sex with men (MSM), and 2 explored the views of men whose sexual orientation was unspecified. The remaining 3 studies explored the views of both men and women.

Young people aged younger than 30 years were exclusively recruited to 4 of the studies, including one of the studies that explored the views of women, one that explored the views of men, and one that explored the views of MSM. This latter study also had exclusively Black participants; another 4 studies reported over 45% of participants as people of color.

Critical Appraisal

The results of the critical appraisal are outlined in Figure 2. All but 3 [35,42,43] of the studies met at least 9 of the 10 appraisal criteria, with the remainder [17,36-41,44] meeting at least 6. All of the studies were deemed valuable, had clear aims and statements of findings, and had appropriate methodologies, study designs, and data collection methods. However, most did not provide enough information to determine whether they met all of the criteria, with only 3 [36-38] reporting sufficient consideration of the researcher-participant relationship. Separately, 3 studies [35,38,42] described their analysis with limited detail, which meant it was not clear how themes were derived from the data and the methods were not replicable.

Synthesis

A total of 12 themes were identified, which were organized into 4 categories. These are outlined in Table 2.

Figure 2. Critical appraisal of included studies according to the Critical Appraisal Skills Programme.

Table 2. Summary of themes.

Categories	Themes
1. Positive aspects of internet-based testing	1.1 Internet-based testing is acceptable
	1.2 Attractive due to convenience
	1.3 Attractive due to the stigma associated with face-to-face testing
	1.4 Avoids undesirable aspects of face-to-face testing
	1.5 Improves accessibility of STI[a] testing
2. Negative aspects of internet-based testing	2.1 Loss of positive aspects of face-to-face testing
	2.2 Concerns about self-sampling processes
	2.3 Privacy concerns with internet-based testing
3. Positive aspects of remote delivery of results	3.1 Remote delivery of results is acceptable
	3.2 Convenience drives preference of results medium
4. Negative aspects of remote delivery of results	4.1 Concern about interception
	4.2 Concern over well-being

[a]STI: sexually transmitted infection.

Category 1: Positive Aspects of Internet-Based Testing

Internet-Based Testing Is Acceptable

There was a broad yet incomplete consensus that internet-based testing was acceptable, with around half of studies explicitly reporting that participants were open or positive towards using it [35,37,39,41,42]. Ahmed-Little et al [35], for example, reported:

> *There was overwhelming support that this method of testing offered ease and was considered acceptable.*

A minority of studies did report uncertainty or negativity among some potential users of hypothetical internet-based testing services, however [36,37,39]. Friedman and Bloodgood [36] reported that overall:

> *Participants were slightly more negative than positive about the option of ordering [a sexually transmitted disease] test from a website.*

Attractive Due to Convenience

The perceived convenience of internet-based testing was a prominent theme, appearing in almost all of the included studies [35-37,39-44]. Approximately half of them identified that internet-based testing appealed to participants because it meant they would not have to take time out of their day to get tested or go to the effort of travelling to a clinic [35,37,39,40,43]:

> *You just go in there, find the information that you need and don't have to worry about travelling, getting gas, whatever, so...it's quick and easy.* [40]

The convenience of not having to make or wait for an appointment also enhanced the appeal of internet-based testing among participants, especially those who had accessed it:

> *I think this is a great service as I have tried to do this through my doctor and will have to wait 3 weeks for an appointment.* [35]

Attractive Due to the Stigma Associated With Face-to-Face Testing

Internet-based testing appealed to many participants due to the perceived anonymity, confidentiality, or privacy it offers compared with face-to-face testing. This theme appeared in most of the studies included in the review and applied to women and those who had accessed internet-based testing in particular [17,35-37,39-42,44]. Many participants felt embarrassed, anxious, or ashamed about the prospect of others seeing them at a clinic or finding out they had attended one:

> *If you can do it all remotely and without anybody knowing or seeing you waiting outside a sexual health clinic and going, "Oh, what are you doing here?" then I think it's going to be absolutely brilliant.* [17]

Avoids Undesirable Aspects of Face-to-Face Testing

Around a third of the included studies reported that participants were attracted to internet-based testing because it allowed them to avoid specific aspects of face-to-face testing that they disliked, a finding that was prominent among those who had accessed internet-based services [17,35-37]. The most prevalent of these aspects was interacting with clinic staff, a negative prospect for many participants, particularly women. Friedman and Bloodgood [36] reported one participant stating that they liked the idea of internet-based testing, as it meant:

> *I don't have to…have this long talk with a professional about sexual education.*

Ahmed-Little et al [35] and Gaydos et al [37] both reported participants stating that internet-based testing would be appealing for users who experience anxiety about interacting with a health care professional. These same two studies also identified that some participants preferred internet-based testing to face-to-face testing, as it was more comfortable and allowed them to avoid sampling methods they were averse to, such as venipuncture:

> *I am terrified of needles so this small lancet is much easier I would rather prick my finger than have a needle in my arm.* [35]

Improves Accessibility of STI Testing

Almost half of the included studies reported that participants felt internet-based testing would improve access to testing for STIs, as it would allow them to overcome barriers such as not being able to afford face-to-face testing, not having easy access to it, or feeling averse to using it. This finding emerged strongly among participants who had accessed internet-based testing and was seen as particularly advantageous for young people, with Gaydos et al [37] reporting one participant stating:

> *It is always good to have several ways to get tested at no cost or low cost especially for teens.*

There were a small number of studies, however, that recorded concerns about potential barriers to internet-based testing, such as cost or a lack of internet access. Tomnay et al [43], for example, found that:

> *An important consideration for all groups was that online STI testing is a free service…[R]esearchers were asked by participants about the cost of using a website for online testing with questions such as 'so this is all free?' and including specific questions such as 'would they [referring to the website] pay the postage to send it back?' in reference to returning the testing kit.*

Category 2: Negative Aspects of Internet-Based Testing

Loss of Positive Aspects of Face-to-Face Testing

Over half of the included studies found that some participants were dissuaded from using internet-based testing, as they felt it was lacking important aspects of face-to-face testing [36,37,39-41,44]. One common example was the opportunity to speak with a health care provider about health holistically. This attitude appeared to be more widely held by men, as evidenced by Roth et al [40], who reported a participant saying:

> *[I want to talk to a professional] when I'm thinking if I need to take a test. You know, what was the probability I was infected? What do I need to know about how transmission occurs? Are there any studies or statistical data that correspond with my particular case that would give me a clue in on how worried about it do I really need to be?*

There was an implication in some of the studies reporting this finding that this view was predominantly held by participants who were familiar and comfortable with the provider they would use to get tested for STIs and that it was not found among participants who had accessed internet-based testing [35,39,41].

Concerns About Self-Sampling Processes

Over half of the included studies identified concerns among participants about the self-sampling process [35-37,41,42,44]. One aspect of this was the prospect of challenges or discomfort caused by self-sampling, as reported by one participant in Tobin et al [42]:

> *I showed them the packet. And the first question they asked was 'Is it painful?' ... Once I told them it was painless they were a little more interested.*

Tobin et al [42] also identified the other prevalent concern about self-sampling: it may be inaccurate or unreliable. They report one participant stating:

> *If it is not 100% accurate they [peers] would probably prefer to go to a clinic. Even if I am telling them, they might not feel they know enough about the at-home test and might think it is better to go to a clinic.* [42]

Another aspect of the process that a number of studies identified as a concern was the return of samples, which a number of participants felt may compromise the test in some way [36,40,44]. Roth et al [40] reported:

> *The top ranked method for sample return was in person, even among individuals who preferred self-sampling … Recurrent themes for personally returning the sample to the clinic included the possibility of sample misidentification, the possibility of loss of confidentiality, mistrust of the postal system and immediate access to treatment.*

Uncertainty around the self-sampling appeared to be more prevalent among men and in studies with ethnically diverse participants, although such concerns were not ubiquitous in these populations.

Privacy Concerns With Internet-Based Testing

A number of studies researching perceptions of hypothetical services identified privacy concerns among participants relating to internet-based testing [37,39-41,43]. These largely centered around obtaining or returning self-sampling kits, with young people in particular expressing concern about their parents finding a kit in the mail. Tomnay et al [43] reported a teenaged participant stating:

> *You don't want your parents to know about it, every day you're going to be the first one to that mailbox checking to see whether it's there.*

Gaydos et al [37] and Roth et al [40] both reported concerns among participants about being seen collecting or returning a kit in a public location, a theme that appeared to be stronger in studies with ethnically diverse populations, while Stahlman et al [41] noted concern about data being collected by providers:

> *One participant noted that he would not be comfortable submitting identifying information, such as his name, online.*

Category 3: Positive Aspects of Remote Delivery of Results

Remote Delivery of Results Is Acceptable

The included studies that sought participant views on results delivery covered a wide range of media, including SMS, email, websites, and mail. A variety of options for phone delivery were also covered, including a phone call from a health practitioner, an answering machine message, and having to proactively call in to obtain results. The diversity of the media investigated means it is difficult to draw firm conclusions on any particular option, as opposing views were identified on almost every medium. Nevertheless, almost half of the included studies reported positive attitudes towards the delivery of results electronically. This included Gibbs et al [38], who explored users' experiences of receiving results via a website and found that:

> *They welcomed the online results service, for the ability it gave them to log on when they felt ready.*

Ahmed-Little et al [35], the other study that explored participants' experiences of using internet-based testing, also found participants to be positive about electronic results.

Results by phone were also found to be viewed positively in almost a third of studies [36,37,40,44]. The most mixed response was found towards the delivery of results by mail [36,37,40].

Convenience Drives Preference of Results Medium

Almost a third of studies identified that participants' preferences for a results medium were frequently influenced by its perceived convenience [36,38,40]. Roth [40], for example, quoted a participant discussing SMS results:

> *I can just flip my phone up real quick, even at work, and like okay, cool. And then I can just know what the results were and it'd just be nice. It'd be easy.*

Friedman and Bloodgood [36], meanwhile, found the same motivation among participants who preferred email notification and an answering machine message on their phone.

Category 4: Negative Aspects of Remote Delivery of Results

Concern About Interception

Almost all of the studies that had data on attitudes towards results notification reported that one of users' main concerns was interception, although this manifested itself in differing preferences [36-40]. Gaydos et al [37] explored attitudes towards a number of media and found concerns relating to all of them from some participants. One said of results via a website, for example:

> *Typing in a passcode for results on the internet is a good idea but most families use the same computer so you will have to be careful not to leave your passcode lying around.* [37]

Other participants shared concerns about mail being read by family members and calls being overheard. This theme appeared to be prominent among women and those who had accessed internet-based testing.

Concern Over Well-Being

Almost half of the included studies reported that attitudes towards results delivery were motivated in part by a concern for the well-being of users [36-38,40,42]. This was most frequently expressed through the concern that communication of results via electronic media would mean that users would not receive sufficient support or advice, particularly if the results were positive. This also appeared to be a commonly held attitude by women, with Friedman and Bloodgood [36] quoting one participant who had received results face-to-face:

> *It was nice to speak to somebody and for them to tell me 'okay, this is what we did, this is the tests we ran and thank God, you're negative.' Again, just somebody to talk to so that if you had questions, you had somebody face to face to talk to.*

A number of participants were also concerned that the use of certain media for results delivery may result in users being left unaware of a positive diagnosis, for example, if they had to proactively call in to be informed or needed to remember a password to access results online. One participant was quoted by Roth et al [40] as saying:

> *I don't think email would be good because everybody gets junk mail they might just delete the email without even knowing.*

In contrast, Gibbs et al [38] reported that participants appreciated being able to log on to a website and access their results whenever they felt mentally prepared to do so.

Discussion

Principal Results

This review and synthesis identified a wide range of perceptions and experiences of internet-based testing held by users and potential users. There was a clear finding that internet-based testing is attractive due to its convenience and the fact that it alleviates concerns around stigma associated with being tested in a clinic, and many participants were drawn to it because it allowed them to avoid elements of face-to-face testing that they disliked. However, there was also a concern among some participants relating to the privacy of internet-based testing, the self-sampling process, and the fact that internet-based testing would be missing positive aspects of face-to-face testing. There was no universally accepted medium for results delivery, but preference was largely motivated by perceptions of convenience and concerns over privacy. Overall, internet-based testing appears to be acceptable despite some reservations expressed about it.

Strengths and Limitations

This review is the first attempt we are aware of to bring together qualitative data that relate to this growing medium of STI testing. The inclusion of data from 1115 participants—a substantial number for qualitative research—is a notable strength, as is the fact that the data were synthesized using a well-established and transparent method. All of the studies were assessed to be of satisfactory quality, imbuing confidence in the results, and the fact that all studies were undertaken in 3 countries with similar socioeconomic profiles enhances the generalizability of the results.

The synthesis was limited by the small number of studies eligible for inclusion, and the quantity of data varied widely between these studies. This meant that although there was a relatively strong consensus among included studies on most themes, it was difficult to draw definitive conclusions on subpopulations. Only 2 of the studies collected data from people who had actually used internet-based testing, for example, and one of these collected data only on their experience receiving results, which limited the distinctions that could be drawn between their findings and the findings of the 9 studies that collected data from potential users of hypothetical services. This similarly affected the findings on perceptions held by different sociodemographic populations, and the review may also have benefited from studies undertaken in a wider number of geographical settings and published in languages other than English. The included studies are also unlikely to have recruited people from vulnerable populations, such as homeless people or those with serious mental health issues, meaning these populations' views may not have been captured in the review. The screening of studies for inclusion was limited, as only a proportion of results were screened by 2 authors, and although including studies published between 2005 and 2018 ensured that all relevant data were included, this relatively long time frame means that findings from some of the earlier studies may now be less relevant due to considerable changes in internet usage over this period.

Comparison With Prior Work

Internet-based STI testing differs from many other digital health interventions (such as telehealth, patient portals, and remote monitoring), as it does not involve two-way communication with clinical staff and is designed as a one-off engagement with the health care system rather than part of the management of a long-term condition. This review offers the opportunity to explore whether users of STI-testing services interact with the service in the same way as those using other digital health interventions, and one notable difference is the fact that internet-based testing is highly associated with convenience. This contrasts with the findings of a review by O'Connor et al [45] about the factors affecting more general engagement with digital health interventions, which found that many people could be deterred from accessing interventions if they felt they lacked the time or energy to do so, which may be because such interventions are frequently targeted towards individuals who are expected to have more sustained engagement [46].

However, other findings were similar to those found for alternative digital interventions, including the finding that some participants had privacy concerns over electronically providing the personal data required for access and that other participants appreciated the anonymity offered by digital interactions over face-to-face ones [45,47]. This concern over the provision of personal data parallels evidence on engagement with digital media unrelated to health care, such as social media, which regularly forces users to weigh privacy concerns against perceived benefits of use [48,49]. The normalization of social media can frequently lead users to overlook their privacy concerns, however, and this is unlikely to apply in the same way when people access internet-based testing, given that it is less likely to feature as often in their day-to-day lives [48-50].

The stigma associated with sexual health also undoubtedly heightens privacy concerns held by users of internet-based testing [23]. The concerns over privacy and anonymity identified in this review simultaneously highlight the potential of digital interventions to overcome stigma as a barrier to accessing health care—in the context of both sexual health and other conditions—and the underresearched phenomenon of the role that stigma plays when using internet-based interventions

[23,45,51,52]. Our finding that many people are concerned about aspects of the internet-based testing process that could allow others to know they had used sexual health services emphasizes the role providers have in mitigating that risk and suggests a need for further exploration of the role that stigma plays when individuals access, or consider accessing, internet-based health care. It is noteworthy that this finding was prominent in studies with ethnically diverse populations and that women were found to be particularly concerned about the stigma associated with clinics and interacting with clinic staff, as users of internet-based testing are disproportionately women and White [53-55].

The concern identified that internet-based testing may deprive users of important aspects of face-to-face testing is significant, particularly as internet-based testing is one of the few health interventions that allows users to have no direct contact with clinical staff. Although it did not appear to deter the majority of participants from the prospect of using internet-based testing, it aligns with evidence from other contexts that digital health care is seen as supplementary and that service users are willing to use it provided it does not replace face-to-face care [51]. This may also be the case for the delivery of test results remotely, a topic on which limited evidence exists and to which this review therefore makes a notable contribution [56]. Our finding that people who had used internet-based testing were satisfied to receive their results by SMS is in agreement with other research, which has found high levels of satisfaction among patients who have received results electronically [57]. That we found no consensus on preferred media for results delivery among potential users corresponds with conflicting data from other sexual health studies, suggesting that service users' preferences when conceptualizing hypothetical services may not be an accurate predictor for what they find acceptable when they start using them [58-60].

Recommendations for Practice

Providers of internet-based testing may wish to emphasize the approach's comparative convenience and privacy compared with face-to-face testing in order to improve uptake, as these appear to be the most appealing aspects of the service. Uptake may also be improved through attempts to alleviate concerns around self-sampling, for example, by providing reassurance about discomfort and emphasizing that the sensitivity and specificity of self-sampling is comparable to samples obtained in a clinic. It may also be worthwhile for providers to consider patient confidentiality in results delivery, for example, by ensuring that text messages are worded so they could not inadvertently reveal a diagnosis in a phone notification, and to ensure that patients receive adequate signposting to support if results are delivered electronically.

Conclusions

This study has identified a wide range of perceptions and experiences of internet-based testing by actual and potential users, including both positive and negative comparisons with clinic-based testing. There is a clear need for further qualitative research exploring in depth the experiences of people who have accessed internet-based testing, given the paucity of data on this, and for qualitative research on internet-based testing that explicitly gathers the views of populations that are at high risk of STIs or that have testing behaviors that require more in-depth understanding. There would also be value in further research on attitudes towards communicating results, given that no consensus could be found on a preferred medium.

Acknowledgments

The research was funded by the Warwick Economic and Social Research Council Doctoral Training Centre and the University Hospitals Birmingham National Health Service Foundation Trust.

Conflicts of Interest

JR reports personal fees from GSK Pharma, Mycovia, and Nabriva Therapeutics, as well as ownership of shares in GSK Pharma and AstraZeneca Pharma; is author of the UK and European Guidelines on Pelvic Inflammatory Disease; and is a member of the European Sexually Transmitted Infections Guidelines Editorial Board and the National Institute for Health Research Funding Committee (Health Technology Assessment Programme). He is an NIHR Journals editor and associate editor of Sexually Transmitted Infections journal. He is an officer of the International Union against Sexually Transmitted Infections (treasurer) and a charity trustee of the Sexually Transmitted Infections Research Foundation.

Multimedia Appendix 1
Example electronic database search, MEDLINE.
[DOCX File , 14 KB - jmir_v22i8e17667_app1.docx]

References

1. Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2017. Atlanta, GA: US Department of Health and Human Services; 2018. URL: https://www.cdc.gov/std/stats17/2017-STD-Surveillance-Report_CDC-clearance-9.10.18.pdf [accessed 2020-08-17]
2. Australian Institute of Health and Welfare. Australia's health 2018. Canberra, Australia: Australian Institute of Health and Welfare; 2018. URL: https://www.aihw.gov.au/getmedia/7c42913d-295f-4bc9-9c24-4e44eff4a04a/aihw-aus-221.pdf [accessed 2020-08-17]

3. Public Health England. Sexually transmitted infections and screening for chlamydia in England, 2018. 2019. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/806118/hpr1919_stis-ncsp_ann18.pdf [accessed 2020-08-17]
4. Public Health England. Addressing the increase in syphilis in England: PHE Action Plan. 2019. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/806076/Addressing_the_increase_in_syphilis_in_England_Action_Plan_June_2019.pdf [accessed 2020-08-17]
5. Unemo M, Shafer WM. Antimicrobial resistance in Neisseria gonorrhoeae in the 21st century: past, evolution, and future. Clin Microbiol Rev 2014 Jul;27(3):587-613 [FREE Full text] [doi: 10.1128/CMR.00010-14] [Medline: 24982323]
6. Farley TA, Cohen DA, Elkins W. Asymptomatic sexually transmitted diseases: the case for screening. Prev Med 2003 Apr;36(4):502-509. [doi: 10.1016/s0091-7435(02)00058-0] [Medline: 12649059]
7. Newman L, Rowley J, Vander HS, Wijesooriya NS, Unemo M, Low N, et al. Global Estimates of the Prevalence and Incidence of Four Curable Sexually Transmitted Infections in 2012 Based on Systematic Review and Global Reporting. PLoS One 2015;10(12):e0143304 [FREE Full text] [doi: 10.1371/journal.pone.0143304] [Medline: 26646541]
8. Jackson LJ, Auguste P, Low N, Roberts TE. Valuing the health states associated with Chlamydia trachomatis infections and their sequelae: a systematic review of economic evaluations and primary studies. Value Health 2014;17(1):116-130 [FREE Full text] [doi: 10.1016/j.jval.2013.10.005] [Medline: 24438725]
9. Global strategy for the prevention and control of sexually transmitted infections: 2006-2015: breaking the chain of transmission. Geneva, Switzerland: World Health Organization; 2007. URL: https://apps.who.int/iris/bitstream/handle/10665/43853/9789241563475_eng.pdf [accessed 2020-08-04]
10. Balfe M, Brugha R. What prompts young adults in Ireland to attend health services for STI testing? BMC Public Health 2009 Aug 26;9:311 [FREE Full text] [doi: 10.1186/1471-2458-9-311] [Medline: 19709430]
11. Gilbert M, Thomson K, Salway T, Haag D, Grennan T, Fairley CK, et al. Differences in experiences of barriers to STI testing between clients of the internet-based diagnostic testing service GetCheckedOnline.com and an STI clinic in Vancouver, Canada. Sex Transm Infect 2019 Mar;95(2):151-156 [FREE Full text] [doi: 10.1136/sextrans-2017-053325] [Medline: 29437984]
12. Tilson EC, Sanchez V, Ford CL, Smurzynski M, Leone PA, Fox KK, et al. Barriers to asymptomatic screening and other STD services for adolescents and young adults: focus group discussions. BMC Public Health 2004 Jun 09;4:21 [FREE Full text] [doi: 10.1186/1471-2458-4-21] [Medline: 15189565]
13. Kirby Intitute. HIV, viral hepatitis and sexually transmissible infections in Australia: Annual surveillance report 2018. Sydney, Australia: Kirby Institute, UNSW Sydney; 2018. URL: https://kirby.unsw.edu.au/sites/default/files/kirby/report/KI_Annual-Surveillance-Report-2018.pdf [accessed 2020-08-04]
14. La Ruche G, Le Strat Y, Fromage M, Berçot B, Goubard A, de Barbeyrac B, et al. Incidence of gonococcal and chlamydial infections and coverage of two laboratory surveillance networks, France, 2012. Euro Surveill 2015 Aug 13;20(32):6-15 [FREE Full text] [Medline: 26290487]
15. Robertson R. Sexual health services and the importance of prevention. London, UK: The King's Fund; 2018 Dec 30. URL: https://www.kingsfund.org.uk/blog/2018/12/sexual-health-services-and-importance-prevention [accessed 2020-08-04]
16. Wilson E, Free C, Morris TP, Kenward MG, Syred J, Baraitser P. Can Internet-Based Sexual Health Services Increase Diagnoses of Sexually Transmitted Infections (STI)? Protocol for a Randomized Evaluation of an Internet-Based STI Testing and Results Service. JMIR Res Protoc 2016 Jan 15;5(1):e9 [FREE Full text] [doi: 10.2196/resprot.4094] [Medline: 26772143]
17. Baraitser P, Syred J, Spencer-Hughes V, Howroyd C, Free C, Holdsworth G. How online sexual health services could work; generating theory to support development. BMC Health Serv Res 2015 Dec 05;15:540 [FREE Full text] [doi: 10.1186/s12913-015-1200-x] [Medline: 26637295]
18. Paudyal P, Llewellyn C, Lau J, Mahmud M, Smith H. Obtaining self-samples to diagnose curable sexually transmitted infections: a systematic review of patients' experiences. PLoS One 2015;10(4):e0124310 [FREE Full text] [doi: 10.1371/journal.pone.0124310] [Medline: 25909508]
19. Odesanmi TY, Wasti SP, Odesanmi OS, Adegbola O, Oguntuase OO, Mahmood S. Comparative effectiveness and acceptability of home-based and clinic-based sampling methods for sexually transmissible infections screening in females aged 14-50 years: a systematic review and meta-analysis. Sex Health 2013 Dec;10(6):559-569. [doi: 10.1071/SH13029] [Medline: 24160747]
20. Ogale Y, Yeh PT, Kennedy CE, Toskin I, Narasimhan M. Self-collection of samples as an additional approach to deliver testing services for sexually transmitted infections: a systematic review and meta-analysis. BMJ Glob Health 2019;4(2):e001349 [FREE Full text] [doi: 10.1136/bmjgh-2018-001349] [Medline: 31139454]
21. Balfe M, Brugha R, O'Connell E, Vaughan D, O'Donovan D. Men's attitudes towards chlamydia screening: a narrative review. Sex Health 2012 May;9(2):120-130. [doi: 10.1071/SH10094] [Medline: 22498155]
22. Jackson LJ, Roberts TE. Conceptualising quality of life outcomes for women participating in testing for sexually transmitted infections: A systematic review and meta-synthesis of qualitative research. Soc Sci Med 2015 Oct;143:162-170. [doi: 10.1016/j.socscimed.2015.08.048] [Medline: 26360418]

23. Hood JE, Friedman AL. Unveiling the hidden epidemic: a review of stigma associated with sexually transmissible infections. Sex Health 2011 Jun;8(2):159-170. [doi: 10.1071/SH10070] [Medline: 21592429]
24. Evans D, Farquhar C. An interview based approach to seeking user views in genitourinary medicine. Genitourin Med 1996 Jun;72(3):223-226 [FREE Full text] [doi: 10.1136/sti.72.3.223] [Medline: 8707330]
25. Spence T, Kander I, Walsh J, Griffiths F, Ross J. Attitudes towards internet-based testing for sexually transmitted infections: a systematic review and thematic synthesis of qualitative research. PROSPERO - University of York Internet.: National Institute for Health Research URL: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=146938 [accessed 2019-11-28]
26. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res 2014 Nov 21;14:579 [FREE Full text] [doi: 10.1186/s12913-014-0579-0] [Medline: 25413154]
27. Ring N, Ritchie K, Mandava L, Jepson R. A guide to synthesising qualitative research for researchers undertaking health technology assessments and systematic reviews. Edinburgh, Scotland: NHS Quality Improvement Scotland; 2011. URL: https://dspace.stir.ac.uk/bitstream/1893/3205/1/HTA_MethodsofSynthesisingQualitativeLiterature_DEC10%5b1%5d.pdf [accessed 2020-08-17]
28. Internet access – households and individuals, Great Britain: 2017. Office for National Statistics. 2017. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017 [accessed 2020-08-17]
29. Gaydos CA, Barnes M, Aumakhan B, Quinn N, Agreda P, Whittle P, et al. Can e-technology through the Internet be used as a new tool to address the Chlamydia trachomatis epidemic by home sampling and vaginal swabs? Sex Transm Dis 2009 Sep;36(9):577-580 [FREE Full text] [doi: 10.1097/OLQ.0b013e3181a7482f] [Medline: 19543145]
30. Internet Usage Data.: The World Bank; 2019. URL: https://data.worldbank.org/indicator/IT.NET.USER.ZS?end=2018&most_recent_value_desc=false&start=2005 [accessed 2020-08-04]
31. UNDP/UNFPA/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction, World Health Organization. Sexually transmitted infections (STIs) : the importance of a renewed commitment to STI prevention and control in achieving global sexual and reproductive health. Geneva, Switzerland: World Health Organization; 2012. URL: https://apps.who.int/iris/handle/10665/75838 [accessed 2020-08-04]
32. Yeh PT, Kennedy CE, de Vuyst H, Narasimhan M. Self-sampling for human papillomavirus (HPV) testing: a systematic review and meta-analysis. BMJ Glob Health 2019;4(3):e001351 [FREE Full text] [doi: 10.1136/bmjgh-2018-001351] [Medline: 31179035]
33. CASP Qualitative Checklist. Critical Appraisal Skills Programme (CASP). 2018. URL: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Qualitative-Checklist-2018.pdf [accessed 2020-08-17]
34. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
35. Ahmed-Little Y, Bothra V, Cordwell D, Freeman Powell D, Ellis D, Klapper P, et al. Attitudes towards HIV testing via home-sampling kits ordered online (RUClear pilots 2011-12). J Public Health (Oxf) 2016 Sep;38(3):585-590. [doi: 10.1093/pubmed/fdv075] [Medline: 26045472]
36. Friedman AL, Bloodgood B. Exploring the feasibility of alternative STD-testing venues and results delivery channels for a national screening campaign. Health Promot Pract 2013 Jan;14(1):96-104. [doi: 10.1177/1524839911404226] [Medline: 21965593]
37. Gaydos CA, Rizzo-Price PA, Barnes M, Dwyer K, Wood BJ, Hogan MT. The use of focus groups to design an internet-based program for chlamydia screening with self-administered vaginal swabs: what women want. Sex Health 2006 Dec;3(4):209-215. [doi: 10.1071/sh05059] [Medline: 17112429]
38. Gibbs J, Aicken CRH, Sutcliffe LJ, Gkatzidou V, Tickle LJ, Hone K, et al. Mixed-methods evaluation of a novel online STI results service. Sex Transm Infect 2018 Dec;94(8):622-624 [FREE Full text] [doi: 10.1136/sextrans-2017-053318] [Medline: 29326179]
39. Lorimer K, McDaid L. Young men's views toward the barriers and facilitators of Internet-based Chlamydia trachomatis screening: qualitative study. J Med Internet Res 2013;15(12):e265 [FREE Full text] [doi: 10.2196/jmir.2628] [Medline: 24300158]
40. Roth A, Van Der Pol B, Dodge B, Fortenberry JD, Zimet G. Future chlamydia screening preferences of men attending a sexually transmissible infection clinic. Sex Health 2011 Sep;8(3):419-426. [doi: 10.1071/SH10017] [Medline: 21851785]
41. Stahlman S, Plant A, Javanbakht M, Cross J, Montoya JA, Bolan R, et al. Acceptable interventions to reduce syphilis transmission among high-risk men who have sex with men in Los Angeles. Am J Public Health 2015 Mar;105(3):e88-e94. [doi: 10.2105/AJPH.2014.302412] [Medline: 25602881]
42. Tobin K, Edwards C, Flath N, Lee A, Tormohlen K, Gaydos CA. Acceptability and feasibility of a Peer Mentor program to train young Black men who have sex with men to promote HIV and STI home-testing to their social network members. AIDS Care 2018 Jul;30(7):896-902 [FREE Full text] [doi: 10.1080/09540121.2018.1442553] [Medline: 29482342]
43. Tomnay JE, Bourke L, Fairley CK. Exploring the acceptability of online sexually transmissible infection testing for rural young people in Victoria. Aust J Rural Health 2014 Feb;22(1):40-44. [doi: 10.1111/ajr.12077] [Medline: 24460999]

44. Wayal S, Llewellyn C, Smith H, Fisher M. Home sampling kits for sexually transmitted infections: preferences and concerns of men who have sex with men. Cult Health Sex 2011 Mar;13(3):343-353. [doi: 10.1080/13691058.2010.535018] [Medline: 21154069]
45. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
46. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther 2014 Nov;16(11):784-793. [doi: 10.1089/dia.2014.0078] [Medline: 24999599]
47. Irizarry T, DeVito DA, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
48. Kokolakis S. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. Computers & Security 2017 Jan;64:122-134 [FREE Full text] [doi: 10.1016/j.cose.2015.07.002]
49. Agosto D, Abbas J. “Don’t be dumb—that’s the rule I try to live by”: A closer look at older teens’ online privacy and safety attitudes. New Media & Society 2016 Jul 09;19(3):347-365 [FREE Full text] [doi: 10.1177/1461444815606121]
50. Dienlin T, Trepte S. Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors. Eur J Soc Psychol 2014 Jul 31;45(3):285-297 [FREE Full text] [doi: 10.1002/ejsp.2049]
51. Robards F, Kang M, Usherwood T, Sanci L. How Marginalized Young People Access, Engage With, and Navigate Health-Care Systems in the Digital Age: Systematic Review. J Adolesc Health 2018 Apr;62(4):365-381. [doi: 10.1016/j.jadohealth.2017.10.018] [Medline: 29429819]
52. Batra S, Baker RA, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. Med Devices (Auckl) 2017;10:237-251 [FREE Full text] [doi: 10.2147/MDER.S144158] [Medline: 29042823]
53. Banerjee P, Thorley N, Radcliffe K. A service evaluation comparing home-based testing to clinic-based testing for Chlamydia and gonorrhoea in Birmingham and Solihull. Int J STD AIDS 2018 Sep;29(10):974-979. [doi: 10.1177/0956462418767180] [Medline: 29690825]
54. Barnard S, Free C, Bakolis I, Turner KME, Looker KJ, Baraitser P. Comparing the characteristics of users of an online service for STI self-sampling with clinic service users: a cross-sectional analysis. Sex Transm Infect 2018 Aug;94(5):377-383 [FREE Full text] [doi: 10.1136/sextrans-2017-053302] [Medline: 29437985]
55. Turner KM, Zienkiewicz AK, Syred J, Looker KJ, de Sa J, Brady M, et al. Web-Based Activity Within a Sexual Health Economy: Observational Study. J Med Internet Res 2018 Mar 07;20(3):e74 [FREE Full text] [doi: 10.2196/jmir.8101] [Medline: 29514776]
56. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for communicating results of medical investigations. Cochrane Database Syst Rev 2012;6:CD007456. [doi: 10.1002/14651858.CD007456.pub2] [Medline: 22696369]
57. Pillemer F, Price RA, Paone S, Martich GD, Albert S, Haidari L, et al. Direct Release of Test Results to Patients Increases Patient Engagement and Utilization of Care. PLoS One 2016;11(6):e0154743 [FREE Full text] [doi: 10.1371/journal.pone.0154743] [Medline: 27337092]
58. Brown L, Copas A, Stephenson J, Gilleran G, Ross JDC. Preferred options for receiving sexual health screening results: a population and patient survey. Int J STD AIDS 2008 Mar;19(3):184-187. [doi: 10.1258/ijsa.2007.007172] [Medline: 18397559]
59. Martin L, Knight V, Read PJ, McNulty A. Clients' preferred methods of obtaining sexually transmissable infection or HIV results from Sydney Sexual Health Centre. Sex Health 2013 Mar;10(1):91-92. [doi: 10.1071/SH12062] [Medline: 23158695]
60. Lim MSC, Hocking JS, Hellard ME, Aitken CK. SMS STI: a review of the uses of mobile phone text messaging in sexual health. Int J STD AIDS 2008 May;19(5):287-290. [doi: 10.1258/ijsa.2007.007264] [Medline: 18482956]

Abbreviations

MSM: men who have sex with men
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 23.01.20; peer-reviewed by C Gaydos, D Linde, S Li; comments to author 10.03.20; revised version received 08.06.20; accepted 14.06.20; published 26.08.20.

Please cite as:
Spence T, Kander I, Walsh J, Griffiths F, Ross J
Perceptions and Experiences of Internet-Based Testing for Sexually Transmitted Infections: Systematic Review and Synthesis of Qualitative Research
J Med Internet Res 2020;22(8):e17667
URL: http://www.jmir.org/2020/8/e17667/
doi:10.2196/17667
PMID:32663151

Review

Connected Mental Health: Systematic Mapping Study

Nidal Drissi[1,2], MSc; Sofia Ouhbi[1], PhD; Mohammed Abdou Janati Idrissi[2], PhD; Luis Fernandez-Luque[3,4], PhD; Mounir Ghogho[5], PhD

[1]Department of Computer Science and Software Engineering, United Arab Emirates University, Al Ain, United Arab Emirates

[2]National School For Computer Science, Mohammed V University in Rabat, Rabat, Morocco

[3]Salumedia Labs, Seville, Spain

[4]Adhera Health Inc, Palo Alto, CA, United States

[5]TICLab, International University of Rabat, Rabat, Morocco

Corresponding Author:
Sofia Ouhbi, PhD
Department of Computer Science and Software Engineering
United Arab Emirates University
College of Information Technology, United Arab Emirates University
Al Ain, 15551
United Arab Emirates
Phone: 971 37135568
Email: sofia.ouhbi@uaeu.ac.ae

Abstract

Background: Although mental health issues constitute an increasing global burden affecting a large number of people, the mental health care industry is still facing several care delivery barriers such as stigma, education, and cost. Connected mental health (CMH), which refers to the use of information and communication technologies in mental health care, can assist in overcoming these barriers.

Objective: The aim of this systematic mapping study is to provide an overview and a structured understanding of CMH literature available in the Scopus database.

Methods: A total of 289 selected publications were analyzed based on 8 classification criteria: publication year, publication source, research type, contribution type, empirical type, mental health issues, targeted cohort groups, and countries where the empirically evaluated studies were conducted.

Results: The results showed that there was an increasing interest in CMH publications; journals were the main publication channels of the selected papers; exploratory research was the dominant research type; advantages and challenges of the use of technology for mental health care were the most investigated subjects; most of the selected studies had not been evaluated empirically; depression and anxiety were the most addressed mental disorders; young people were the most targeted cohort groups in the selected publications; and Australia, followed by the United States, was the country where most empirically evaluated studies were conducted.

Conclusions: CMH is a promising research field to present novel approaches to assist in the management, treatment, and diagnosis of mental health issues that can help overcome existing mental health care delivery barriers. Future research should be shifted toward providing evidence-based studies to examine the effectiveness of CMH solutions and identify related issues.

(J Med Internet Res 2020;22(8):e19950) doi:10.2196/19950

KEYWORDS

mental health; connected health; eHealth; mobile health; telehealth; mHealth; mobile phone; health informatics; review; interdisciplinary research; information technology; information systems

Introduction

Background

Mental health issues can decrease the quality of life [1,2], increase the risk of communicable and noncommunicable diseases, and contribute to both unintentional and intentional injury [3]. They may also cause, among other issues, lower educational achievements, substance abuse, and violence [4]. Mental illness is considered as "one of the main causes of unhappiness in the world. It produces nearly as much of the misery that exists as poverty does, and more than is caused by physical illness" [5]. On average, they reduce national income by 5% through unemployment, absenteeism, lowered productivity, and extra physical health care costs [6].

Although mental health issues constitute a global concern and threat, the mental health care industry is still struggling to overcome various barriers and reach people in need [7]. In low- and middle-income countries, more than 75% of people identified with serious anxiety, problematic mood changes, impulse control, or substance abuse disorders did not receive any care [8]. Among the barriers and challenges that threaten the mental health care industry are cost issues [9], shortage of mental health care providers, health plan barriers, lack of coverage or inadequate coverage [10], stigma, and poor mental health literacy [11]. Cultural orientations can, in some cases, be considered as a barrier for seeking mental help and access [12]. Mental health care delivery can also face barriers in complicated situations such as the recent coronavirus disease (COVID-19) global outbreak [13,14]. Such pandemics create global feelings of fear, worry, sadness, and anger and cause a global increase in stress and anxiety, especially for people with existing mental health problems [15,16], putting more load on health care institutions. They also create new obstacles to mental health care delivery as many people are in quarantine and several countries are in complete lockdown, making access to mental health care even more challenging [17].

Connected mental health (CMH), which refers to the use of information and communication technologies (ICT) to support and improve mental health conditions and mental health care, can help alleviate some of the aforementioned barriers [18]. This presents unlimited possibilities and opportunities to help overcome the challenges, barriers, and limitations of the mental health care industry [18].

CMH solutions are connected health solutions for mental health care. The term *connected health* is used to "encompass terms such as wireless, digital, electronic, mobile, and tele-health" [19]. Connected health has become an established research field in the past 5 years [20]. In this paper, we use the term CMH to refer to topics related to electronic mental (e-mental) health, mobile mental (m-mental) health, telemental health, and digital mental health, among others. CMH solutions have the potential to allow anonymous access to overcome care preventing stigma, access to care with minimal to no cost, reach remote areas, enhance patients' engagement, and access information [21,22]. They also have the potential to deliver cost-effective mental health services [23]; overcome geographical barriers [24]; and provide a better understanding of mental illnesses' development, recovery, symptom assessment, and monitoring through the use of technology [25]. CMH can use technology-based interventions such as video conferencing; telemental health [26]; mobile-based interventions [27,28]; internet-based solutions [29,30], which are necessary in situations such as global health pandemics; and novel digital data collection interventions [31,32]. In addition, CMH can improve several important areas of mental health care, including primarily information provision, screening, assessment, monitoring, intervention, and social support [22].

Objectives

This paper presents the results of a systematic mapping study conducted to analyze the current research landscape of CMH literature in the Scopus database. To conduct the analysis, 8 mapping questions (MQs) were addressed, which allowed us to classify 289 selected publications indexed in Scopus according to their publication year, publication source, research type, contribution, empirical type, mental health issues, and targeted cohort.

Methods

Overview

This study follows the mapping process proposed by Petersen et al [33]. This process covers the selection of relevant publications, the construction of a classification scheme, and a systematic mapping of publications. The principal objective of a systematic mapping study is to structure a research area and provide an overview of the available literature, primarily by investigating the covered topics and classifying the available contributions [34].

MQs

Textbox 1 presents the 8 MQs and their rationale. The MQs were defined to provide an overview and a structured understanding of the existing CMH literature in the Scopus database.

Textbox 1. Mapping questions.

MQ1 (mapping question)

- How has the frequency of publications addressing CMH (connected mental health) changed over time?

MQ2

- Which publication channels are the main target for CMH research?

MQ3

- What are the research types of studies addressing CMH?

MQ4

- What are the contributions of published CMH studies?

MQ5

- Are CMH studies empirically validated or evaluated?

MQ6

- What are the mental health disorders addressed in CMH literature?

MQ7

- Who are the target audience in CMH studies?

MQ8

- In which countries were the selected empirically evaluated studies conducted?

Search Strategy

The search was conducted in the Scopus database, which is considered to be the largest indexation database that includes articles from various disciplines such as engineering, medicine, business, and computer science [35].

The aim of the selection process was to identify the articles that are most relevant to the objective of this mapping study. To further focus the search and include relevant studies, the search was focused on the titles of the publications.

The following search strings were applied to the titles of the publications indexed in Scopus:

- "e-" AND "mental" AND "health"
- ("Mobile" OR "M-") AND "mental" AND "health"
- "Digital" AND "mental" AND "health"
- "Connected" AND "mental" AND "health"
- "Tele" AND "mental" AND "health"

The search strings were formulated to include a broad selection of literature and were not combined in one search string for the purpose of identifying the number of results for each term separately. The search for publications was conducted on January 23, 2020.

Papers Selection

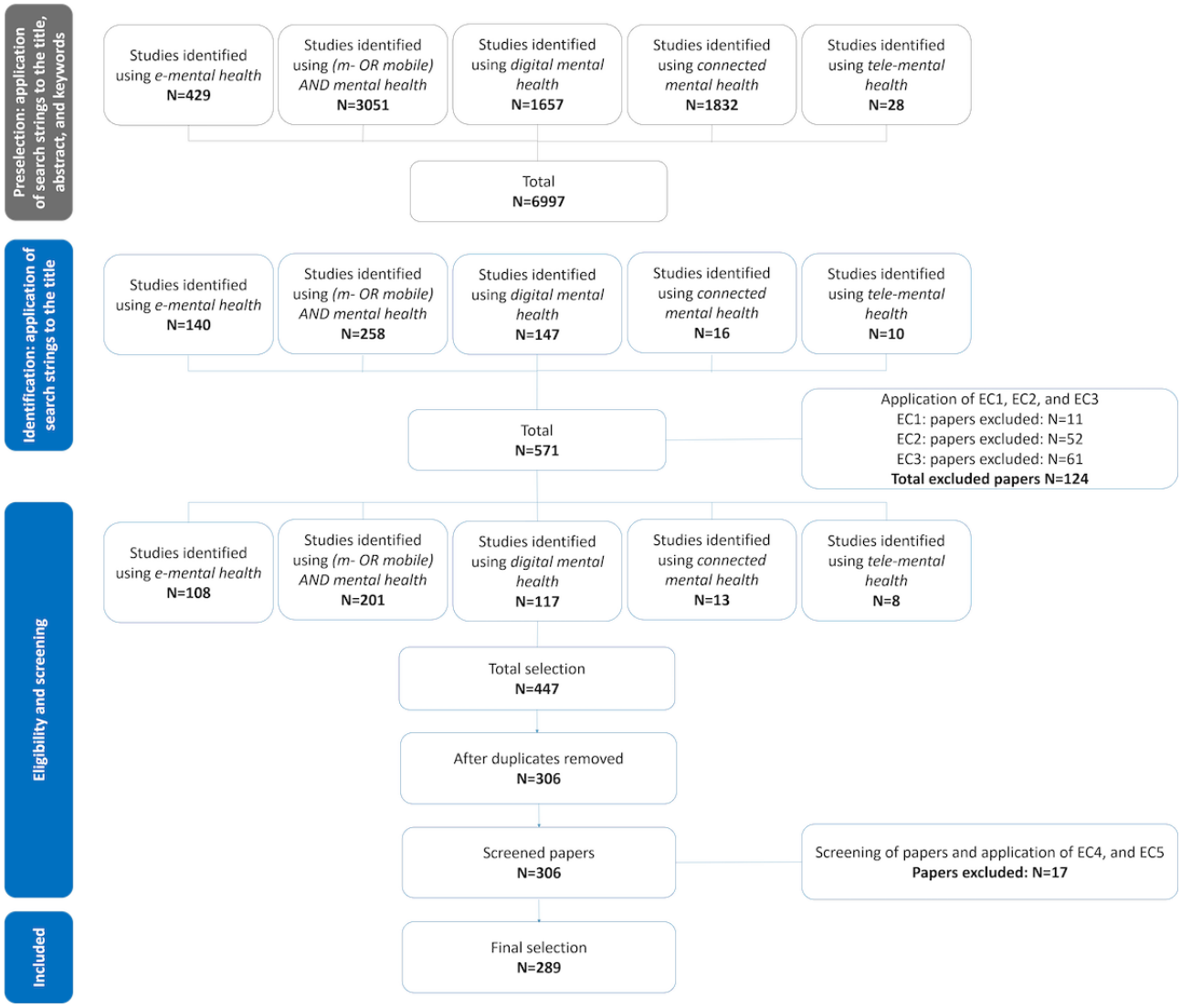
The first author (ND) retrieved candidate papers from the results of the search and entered information in an Excel (Microsoft Corporation) file that was shared with the second author (SO) for revision. The two authors examined the title, abstract, and keywords and then commented on whether the paper was to be included in or excluded from the selection according to the inclusion and exclusion criteria (EC).

The inclusion criterion was limited to publications that addressed an aspect of CMH, and the studies that satisfied any one of the following EC were eliminated:

- EC1: studies published after 2019 to construct a clear trend of publication
- EC2: studies not published in English, as there is a clear dominance of the English language in the international communications, science, and literature [36,37]
- EC3: papers published as notes, editorials, or letters
- EC4: publications not addressing mental health issues
- EC5: publications not addressing the use of technology for mental health.

Figure 1 shows the selection results. A total of 289 papers (out of 571 candidate studies) were included in the final selection. EC 1, 2, and 3 were applied in the Excel file. Checking for EC 4 and 5 and the screening of studies were conducted by inspecting the abstracts and, in some cases, the full texts of the candidate studies.

Figure 1. Selection process. EC: exclusion criteria; e-mental: electronic mental; tele-mental health: telehealth for mental health.



Data Extraction Strategy

The extraction of data from the selected studies was focused primarily on providing answers to the MQs according to the criteria presented in Textbox 2.

Textbox 2. Classification criteria.

MQ1 (mapping question)

- Identifying the publication year and term used can assist in suggesting the publication trend. The main term of each study was not based on the results of the selection process but was based on an analysis of the content of the publications

MQ2

- Identifying the publication channel and the publication source of each study

MQ3

- Research types can be classified as follows: (1) solution proposal: a solution to an existing problem, either novel solution or a significant extension of an existing solution; (2) review: analysis of the existing literature; (3) exploratory study: conducting research and investigating on a specific aspect; (4) opinion paper: these papers express the personal opinion of the author on a specific technique not relying on related work or research methodologies; (5) validation research: this type of paper investigates novel solutions that have not been implemented yet; (6) evaluation research: conducting a study on an implemented solution by identifying how the solution has been implemented in practice, the benefits and drawbacks of the implementation, and relevant problems in the field; (7) feasibility study: an analysis of relevant factors of success of an intervention to determine the likelihood of its implementation successfully; and (8) books and book chapters

MQ4

- The contributions of the selected studies can be classified as follows: (1) tool-based techniques: techniques based on software or devices used to accomplish a connected mental health (CMH)–related task; (2) model: a system that allows structured utilization of CMH solutions; (3) method: development or design of an approach by creating a series of steps that can assist in utilizing CMH solutions; (4) guidelines: a set of rules or recommendations to be followed during the course of an action related to CMH; (5) framework: a real or conceptual structure intended to support and guide the creation of CMH-related contributions; (6) protocol: a set of rules guiding use of technology for mental health care; (7) perspectives and attitudes toward CMH: results on perspectives, attitudes, acceptance, and preferences toward CMH in general or in a specific group of people; (8) advantages and challenges of CMH: views regarding the advantages, benefits, risks, challenges, and limitations of CMH; and (9) other: results of general research on the efficacy, effectiveness, and usability of CMH interventions

MQ5

- The empirical types can be classified as follows: (1) experiment: testing a hypothesis by creating a situation with controlled conditions; (2) case study: testing a solution in real life, with the purpose of gathering new information and detecting potential issues by analyzing real cases; (3) questionnaire: a set of specific, easy to answer questions targeting an audience to collect large amounts of data; (4) interview: a conversational method based on asking specific questions to gather in-depth, precise, and meaningful data; (5) mixed methods: using more than one empirical evaluation method; (6) focus group: investigating results and feedback in a relatively small group of participants; (7) other: all empirical evaluation methods that do not belong to any of the categories above; and (8) none: all solutions that were not validated empirically were classified as theories

MQ6

- Identifying the mental problem addressed in each study

MQ7

- Identifying the targeted cohort group

MQ8

- Identifying the country or region where the study has been conducted. Nonempirically evaluated studies are usually general, whereas empirically evaluated studies focus on specific groups and countries or regions. Therefore, for this question, we focus on the latest

Synthesis Method

The synthesis method used in this study consisted of the following steps:

- Reading and analyzing the 289 selected studies to extract information presented in the *Data Extraction Strategy* subsection.
- Classifying the studies by enumerating the number of publications per MQ. It should be noted that selected publications addressing more than one mental health issue (MQ6), more than one cohort group (MQ7), and more than one country (MQ8) were counted in each category.
- Presenting the classification results in figures and charts, such as bubble plots, as visualization of the results facilitated their analysis.
- Presenting a narrative summary to describe the principal findings.

Compliance With Ethical Standards

This paper does not contain any study with human participants or animals.

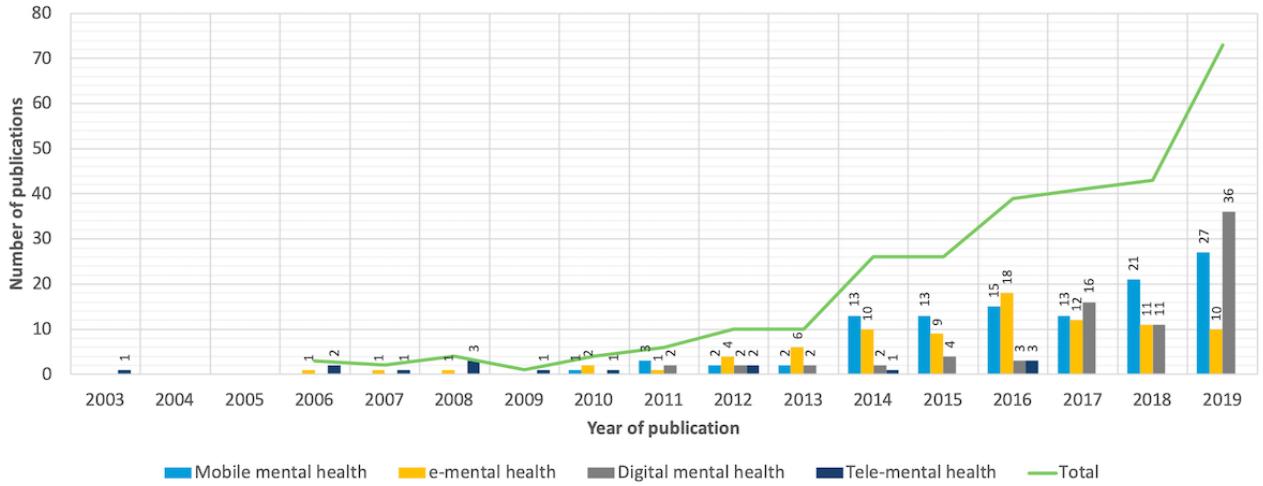
Results

This section summarizes the results of the MQs. The results of the mapping study can be found in Multimedia Appendix 1.

MQ1: How Has the Frequency of Publications Addressing CMH Changed Over Time?

Figure 2 shows the publication trend per term used in the selected papers: mobile or m-mental, e-mental, digital, and telemental health. This shows a significant increase in interest in the field of CMH research in the past decade. The publication trend began with fewer than 5 publications per year until 2011. We estimate that the trend of publication will continue to increase in 2020, particularly owing to the increasing interest in mental health during the recent COVID-19 pandemic.

Figure 2. Connected mental health publication trend. e-mental: electronic mental; tele-mental health: telehealth for mental health.



MQ2: Which Publication Channels Are the Main Target for CMH Research?

A total of 3 publication channels have been identified from the selected studies: journals, conferences, and books. Overall, 82.3% (238/289) of the selected papers were published in journals, 11.4% (33/289) were published in conference proceedings, and 6.2% (18/289) were published in books. A total of 39 of the selected studies were retrieved from the *Journal of Medical Internet Research*. Table 1 presents publication sources that published two or more selected studies.

It is worth noting that there might be relevant studies conducted in e-mental health and mobile health companies intended for internal use only and not published in scientific venues. The findings of such studies might offer new interesting insights and might help evolve the CMH field.

Table 1. Publication sources that published more than one selected publication.

Publication sources	Publications, n	References
Journal of Medical Internet Research	39	[38-76]
e-Mental Health	8	[77-84]
Journal of Mental Health	7	[18,25,85-89]
BMC Psychiatry	7	[90-96]
Psychiatric Services	7	[22,97-102]
Frontiers in Psychiatry	7	[103-109]
Studies in Health Technology and Informatics	6	[110-115]
Current Treatment Options in Psychiatry	6	[116-121]
Journal of Technology in Human Services	5	[122-126]
Professional Psychology: Research and Practice	4	[127-130]
Australian and New Zealand Journal of Psychiatry	4	[131-134]
Internet Interventions	3	[135-137]
Current Psychiatry Reports	3	[138-140]
Asian Journal of Psychiatry	3	[141-143]
Australian Family Physician	3	[144-146]
Evidence-Based Mental Health	3	[147-149]
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	3	[150-152]
PLoS ONE	3	[153-155]
International Journal of Mental Health Nursing	3	[156-158]
Telemedicine and e-Health	3	[159-161]
Psychiatric Times	3	[162-164]
Frontiers in Public Health	3	[165-167]
Health Informatics Journal	3	[168-170]
Indian Journal of Psychological Medicine	3	[171-173]
JAMA Psychiatry	2	[174,175]
Journal of Psychosocial Nursing and Mental Health Services	2	[176,177]
Journal of Physics: Conference Series	2	[178,179]
BMC Medical Informatics and Decision Making	2	[180,181]
Psychiatry (New York)	2	[182,183]
Conference on Human Factors in Computing Systems—Proceedings	2	[184,185]
Journal of Affective Disorders	2	[186,187]
Cyberpsychology, Behavior, and Social Networking	2	[188,189]
Journal of Psychiatric Research	2	[190,191]
European Journal of Psychotraumatology	2	[192,193]
IEEE International Symposium on Computer-Based Medical Systems—Proceedings	2	[194,195]
Advances in Mental Health	2	[196,197]
The Digitization of Healthcare: New Challenges and Opportunities	2	[198,199]
Australian Psychologist	2	[200,201]

MQ3: What Are the Research Types of Studies Addressing CMH?

Figure 3 presents the research types identified in the selected papers. The largest number of the selected publications were exploratory research (104/289, 35.9%), followed by reviews (61/289, 21.1%) and solution proposals (49/289, 16.9%); however, 10.0% (29/289) and only 1.3% (4/289) were evaluation studies and validation research, respectively.

Figure 3. Bubble graph associating the research types with the empirical types and the contribution types.

MQ4: What Are the Contributions of the Selected CMH Studies?

Figure 3 shows the different contributions of the selected papers: 36.6% (106/289) of the selected studies addressed the advantages, challenges, and limitations of CMH; 24.2% (70/289) proposed tool-based solutions; and 19.7% (57/289) investigated perspectives and attitudes toward CMH. Only 3.8% (11/289) of the studies presented guidelines, and only 2.4% (7/289) provided frameworks and protocols.

MQ5: What Are the Empirical Types of the Selected CMH Studies?

Figure 3 shows the identified empirical types of the selected papers. Overall, 52.5% (152/289) of the studies were not evaluated empirically, and 17.9% (52/289) and 12.4% (36/289) were evaluated empirically through experiments and questionnaires, respectively. Only a few publications used empirical evaluation methods such as case studies (7/289, 2.4%), focus groups (7/289, 2.4%), and interviews (6/289, 2.0%). In addition, 10.0% (29/289) of the studies implemented either a mixed method approach (17/289, 5.8%) or used a different, not largely used empirical evaluation approach (12/289, 4.1%).

MQ6: What Are the Mental Health Issues Addressed in the CMH Literature?

Figure 4 shows the mental health issues identified in the selected studies. Most of the selected studies (219 publications) addressed mental well-being in general. A total of 38 publications addressed depression, whereas 22 publications focused on anxiety. Furthermore, 37 papers addressed serious mental illnesses such as bipolar disorder, psychosis, schizophrenia, personality disorders, dementia, panic disorders, and major depressive disorders. Among the identified mental health issues addressed in the selected papers were addiction, suicide, cognitive disorders, sleeping disorders, emotional problems, mindfulness, and productivity influenced by mental state.

Figure 4. Mental health issue types versus targeted cohort. The color and the percentage represent the frequency of appearance of a problem in the selected studies dealing with a specific cohort.

Problems	No specific group (N=200)	Young people (N=33)	Adolescent (N=12)	Children (N=9)	Clinicians (N=14)	Workforce (N=8)	Veterans and partners (N=7)	Adults (N=4)	Women (N=4)	Elderly (N=3)	Parents (N=2)	Refugees (N=1)	Cancer patients (N=1)
General mental health (N=219)	74.5%	69.7%	75.0%	88.9%	78.6%	75.0%	57.1%	75.0%	25.0%	100.0%	50.0%	100.0%	0.0%
Depression (N=38)	11.5%	15.2%	25.0%	11.1%	7.1%	12.5%	0.0%	25.0%	50.0%	0.0%	0.0%	0.0%	100.0%
Serious mental illness (N=37)	15.0%	0.0%	25.0%	33.3%	7.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Anxiety (N=22)	7.0%	12.1%	8.3%	11.1%	7.1%	0.0%	0.0%	25.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Behavioral and emotional problems (N=15)	2.5%	9.1%	25.0%	11.1%	7.1%	0.0%	14.3%	0.0%	0.0%	0.0%	50.0%	0.0%	0.0%
Stress (N=13)	3.5%	9.1%	0.0%	0.0%	7.1%	12.5%	0.0%	0.0%	25.0%	0.0%	0.0%	0.0%	0.0%
Post-traumatic stress disorder (N=7)	1.5%	0.0%	0.0%	0.0%	0.0%	12.5%	42.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sleeping disorders (N=5)	1.0%	3.0%	8.3%	11.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Body image issues and eating disorders (N=5)	0.5%	0.0%	16.7%	22.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Addiction (N-4)	1.5%	3.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Suicide and self-harm (N=4)	1.0%	0.0%	8.3%	11.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Cognitive disorders (N=4)	1.0%	3.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	33.3%	0.0%	0.0%	0.0%

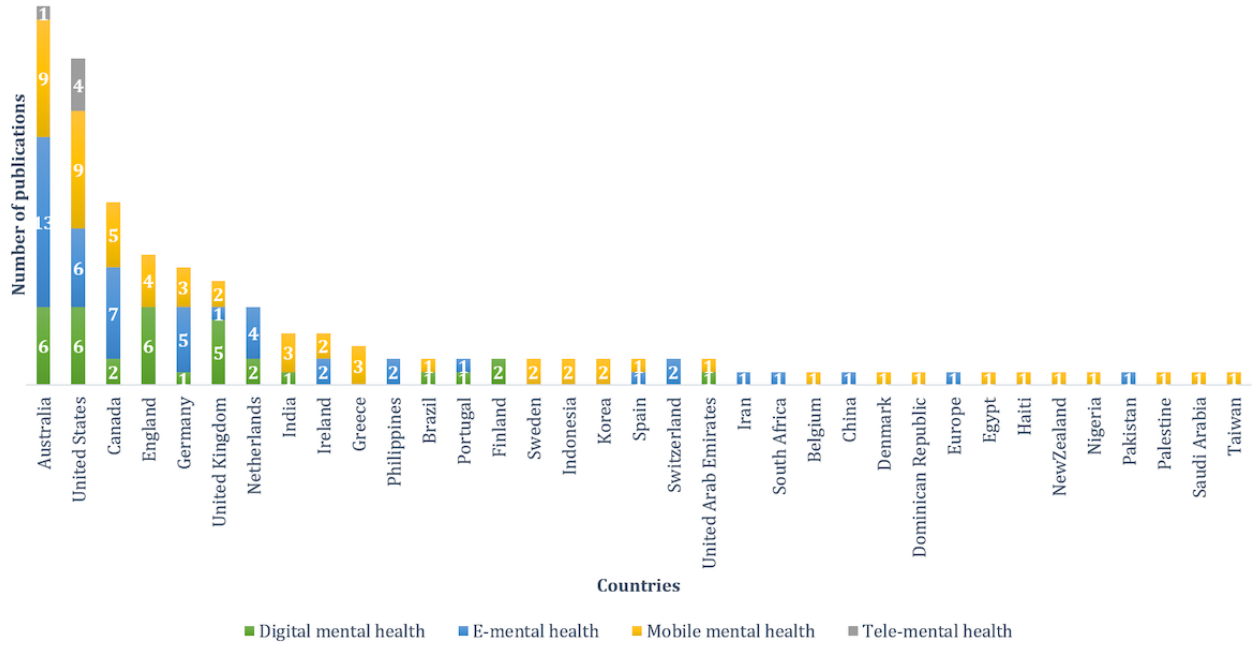
MQ7: Who Are the Targeted Cohorts in the Selected CMH Studies?

Figure 4 shows all groups identified in the studies. Most studies (200 selected publications) did not address a particular group of people. The remaining publications focused either on age-based groups, such as children, adolescents, and the elderly, or on specific groups, such as veterans, refugees, and cancer patients.

MQ8: In Which Countries Were the Selected Empirically Evaluated Studies Conducted?

Figure 5 shows the geographical distribution of the selected empirically evaluated studies. Most studies were conducted in Australia, followed by the United States and Canada. Some studies included more than one country and were counted for each one.

Figure 5. Geographical distribution of empirically evaluated studies. E-mental: electronic mental; tele-mental health: telehealth for mental health.



Discussion

Principal Findings

Trends in Publication and Usage of Terms

There has been an increasing interest in CMH publications in the past decade. This can be explained by the fact that CMH research has been influenced continuously by novel technological interventions and the growth in ICT usage, especially with the appearance of Web 2.0 between 2003 and 2004 [202]. Web 2.0 introduced an architecture of participation [203], focusing on social interactions and collective intelligence [204]. These changes were accompanied by a significant increase in the number of internet users worldwide [205] as well as ownership and usage of smartphones [206] and computers [207-209]. This has also impacted the use of ICT, which has been incorporated in many fields, including transport [210], tourism [211], education [212], virtual work teams [213], and health care [214].

An increase in the number of publications in the field of CMH could have also been influenced by the changes in the general literature. There has been a shift from paper copies to electronic copies, which significantly increased the accessibility to published studies, particularly open access publications [215]. Consequently, the research evaluations performed by peers became reliant on different bibliometrics [216]. The productivity of researchers has mainly been measured by bibliometrics (eg, journal's impact factor [217,218] and researcher's h-index [219]), which is impacting hiring processes, academic promotion, and funding decisions [219,220]. This puts more pressure on researchers to produce more publications in the literature, particularly publications indexed in Scopus.

The earliest term used in the selected CMH studies was *telemental health*, which falls in the same group of terms as *telemedicine* and *telepsychiatry*. These terms have been used to describe remote health care and informatization of some elements of the health care process since the 1990s [221,222]. The term *telemental health* has rarely been used in recent years as *e-mental health* began emerging since 2006.

The term *eHealth* was introduced in the late 1990s. It was defined at the 7th International Congress on Telemedicine and Telecare in London at the end of November 1999 by John Mitchell as "a new term needed to describe the combined use of electronic communication and information technology in the health sector" and as "the health industry's equivalent of e-commerce" [223]. Although previously used terms were linked to medical professionals, *eHealth* was found to be driven by nonprofessionals, namely, patients (or consumers) [224]. Many

journals after the introduction of *eHealth* changed their names to include it; for example, the *Telemedicine Journal* added an "and eHealth" to its title. Nevertheless, publications were not including the term *eHealth*, and researchers preferred to use more specific terms such as *medical informatics*, *telemedicine*, and *electronic patient records*, which are considered more meaningful than the generic term *eHealth* [223]. This might explain its late and delayed emergence in the literature.

The trend of publications also showed an increasing use of the term *m-mental health*, primarily since 2010. The expansion in the use and ownership of smartphones [225] brings new opportunities for the mental health care industry to face some of the most pressing global challenges and make mental health care more accessible, faster, less expensive, and better [226]. Studies have proven the interest and willingness of people with mental health issues to try mobile solutions [73,227]. Hence, it is expected that m-mental health will continue to attract the attention of CMH researchers and practitioners.

The term *digital mental health* has been used since 2011. This term is associated with the use of recent and advanced technologies, such as sensors [228] and machine learning techniques, and it has been used as a modern synonym for *e-mental health*, especially since 2017. It must be noted that researchers use different terms to refer to the same concept, and, in many studies, authors use more than one term, which might occasionally create confusion, especially in cases in which known definitions of each term overlap. Consequently, there is a need for a broad term such as *connected mental health* to encompass the different terms used. Although *connected health* has become an established field [20], we did not find the use of such a term in mental health care. To the best of our knowledge, this is the first study that uses this term to refer to the use of ICT in mental health.

Publication Channels

Most of the selected papers (238/289, 82.3%) were published in scientific journals. Publishing in a rigorous scientific journal is based on a set of ethics and rules [229] that can result in the production of high-quality content. This result could have been impacted also by the choice of digital library. More journals than conference proceedings in CMH might have satisfied the criteria of Scopus indexation. There are also many scientific journals that are categorized in health informatics. This encourages researchers to conduct studies suited for journals and indicates a particular level of maturity in the field from a quantitative perspective. Articles published in journals tend to be more detailed and extensive. Conferences are also an important venue for publications. Conferences with proceedings provide fast tracking of publication to keep researchers and practitioners up to date with the latest findings. The first identified publication in the selected studies, which appeared in 2003, was a book. Books are used to introduce or provide a detailed overview of a topic.

Research Types and Empirical Types

The multidisciplinary nature of the CMH field and its various subareas resulted in a high ratio of exploratory research and reviews compared with other types of research, such as solution proposals, validation, and evaluation research. The diversity of the identified empirical evaluation types points to the uncertainty regarding adequate methods to evaluate CMH solutions. There is a need for evidence-based studies that will verify the efficiency of CMH interventions, create accurate results for future research, and identify challenges that need to be addressed. Most of the selected studies (152/289, 52.5%) were not evaluated empirically, which can be linked to the fact that a large number of empirical evaluation studies in the health informatics field have not been published [230]. Most of the remaining studies were empirically evaluated through experiments (52/289, 17.9%), primarily owing to their controlled environments. Case studies are more suited for sensitive problems such as mental health issues and would be more beneficial for the growth of the field, as they help identify many issues that can be faced in real-life implementation of ICT interventions. However, only 2.4% (7/289) of the selected studies were evaluated empirically through case studies. Questionnaires represented an important part of the empirical evaluation methods of the selected studies (36/289, 12.4%) used mainly in exploratory research (104/289, 35.9%) to investigate perspectives and attitudes toward the use of technology for mental health and on advantages, challenges, and limitations of CMH. Questionnaires are useful tools for data collection as they can be formulated to collect significant amounts and various types of data and can be delivered through various means to reach large groups of people.

Contributions of the Selected Studies

The largest number of the selected studies focused on investigating the advantages, benefits, risks, challenges, and limitations of CMH (106/289, 36.6%) as well as perspectives and attitudes of patients, clinicians, and various stakeholders toward CMH (57/289, 19.7%). An important part of the identified contributions was about *tool-based techniques* (70/289, 24.2%), whereas only a few studies represented *frameworks*, *protocols*, and *models* (14/289, 4.8%). This points to a level of ambiguity in the field as frameworks, protocols, and models are typically signs of clear understanding of a concept, its benefits, and its limitations. Tool-based techniques are usually relatively specific to one area of implementation and target a specific need, for instance, a mobile app for tracking mood changes or a website for evaluating levels of depression. Despite an increasing number of CMH publications, the types of research identified are primarily theoretical and hypothetical, indicating that researchers are still proposing their conceptual ideas for CMH. Only 4.4% (13/289) of the studies, categorized as others in Figure 3, addressed topics such as effectiveness, efficiency, usability, and quality of technological solutions for mental health care. The lack of publications on these topics is related to the aforementioned lack of empirical evaluation studies.

Mental Disorders and the Targeted Cohort

Most of the selected studies did not target a specific group of people (200/289, 69.2%). The remaining studies focused primarily on young people, adolescents, and children (33/89, 37%; 12/89, 13%; and 9/89, 10%, respectively). There was an overlap of age ranges among the categories, primarily owing

to the existing overlapping of ages in the known definitions of the categories. The World Health Organization defines *adolescence* as the age range spanning from 10 to 19 years, *youth* as those aged between 15 and 24 years, and *young people* as the group covering the 2 categories from 10 to 24 years [231]. The adolescent group is further divided to include a subgroup of early adolescence (11-14 years) [232], which might be referred to by some authors as *children*. Targeting primarily the young category of patients can be explained by the fact that this category is characterized by the ownership of mobile devices, high usage of the internet and social networks, and familiarity with the use of technology, which makes it easy for them to become *customers* of CMH solutions.

On the basis of our findings, the *young people* category is the main target for studies on technological solutions for depression, anxiety, and related issues. These disorders are the third leading cause of death among adolescents and young adults, with a prevalence of approximately 4% in children and 10% to 20% in adolescents [233]. Other mental health issues such as body image issues, eating disorders, and sleep problems are also common among young people. Most mental health issues identified in this category could be linked to young people's extensive use of social networks and living a digital life, making them disconnected from real life and absorbed in following other people's lives and comparing them with their own lives. In many cases, this can cause them to feel worthless and disappointed with their lives, bodies, or achievements. Internet use has become an addiction for many young people, causing, among other problems, loss of sleep and high levels of loneliness [234] as well as many cyber-related mental problems [235].

Depression, followed by anxiety, was identified among most of the cohort groups, owing primarily to the increasing global expression of these disorders. The global population with depression increased by 18.4% between 2005 and 2015; similarly, the total number of people with anxiety increased by 14.9% between 2005 and 2015 [236]. Depression can occur as a result of other health problems, such as perinatal and postpartum depression among women [155,237] and common depression among cancer patients [238].

Many serious disorders such as schizophrenia, psychosis, and personality disorders were the subjects of many of the selected studies, pointing to the potential of using CMH not only for common mental disorders but also for more severe and complicated mental problems.

Most of the selected CMH publications addressed general mental well-being. Having more studies focusing on specific mental issues might help identify specific limitations and issues as well as highlight opportunities related to specific disorders. More studies should also focus on people in war zones as they are vulnerable to various mental problems and have more barriers to obtain mental health care, which CMH can help solve [239].

Countries in Which the Empirical Evaluations Were Conducted

Most empirical evaluations were conducted in developed English-speaking countries such as Australia, the United States, Canada, and England. This might be mainly owing to the advancement in research in these countries. In Australia, universities are considered *world-leading* in at least one area of research [240], and large investments are provided to research institutes [241]. There is an important research interest in health informatics in Australia [242]. Research in the United States is believed to be essential to the country's economic growth. Therefore, the federal government has devoted a significant amount of funding for research [243]. The United States has many of the most accomplished individuals in the field of health informatics working in research projects and institutions specialized in the field [244]. Several developing countries have good ICT infrastructures and can benefit from different CMH interventions to solve several existing issues; however, limited research is conducted in developing countries. The same goes for war zones as only 2 studies addressed people in war zones: one conducted for Palestinian children [245] and the other addressed Syrian refugees in Germany, Sweden, and Egypt [106]. CMH interventions are solutions to overcome mental care barriers in developing countries and war zones. Research in this field should be shifted to countries that are most in need of CMH solutions.

Comparison With Related Studies

The findings of this study are aligned with the findings of other studies related to CMH:

- The identified increase in interest in the CMH literature was also reported in a literature review of e-mental health [133].
- The identified lack of empirical evaluation of the proposed solutions was reflected in the findings of a meta-analysis on digital interventions for alcohol and substance use disorders, reporting limitations in the number of randomized control trials (RCTs) in that area [117]. Another review on RCTs on the effectiveness of occupational e-mental health concluded the need for more detailed and specific effectiveness research [246]. A narrative review on digital mental health in developing countries reported a shortage of studies on the effectiveness and cost-effectiveness of the investigated internet-based programs [166]. A review on the effectiveness of mobile apps reported a lack of clinically validated evidence on the efficiency of most available apps [191].
- There was a lack of use of the term *telemental health*. Related findings were described in a review of the directions for videoconferencing and telemental health [247]. The review reported a limited number of RCTs in the field as well as methodologically flawed and limited research studies, which resulted in the research agenda of telemental health not being fully maximized [247].
- Leading countries in the empirical studies selected were Australia and the United States, which are similar to findings from other reviews. A review reported that evidence for telemental health and web-based solutions are largely led by Australia [138]. Two reviews reported that most of their selected publications were found to have mostly originated in the United States and Australia [22,133], and a study [133] found that Australian e-mental health research was higher in diversity and quantity.

Limitations

This study has some limitations. This study did not include other digital databases. However, Scopus is widely considered to be the largest indexing digital library that includes publications from different disciplines [35]. This allowed us to identify different contributions in CMH research. It should be noted that Web of Science (WoS) was not considered as a recent study showed that slight differences exist between the scientific literature covered in Scopus and WoS Core Collection, which will result in a large number of duplicates [248].

The search strings were applied only to the titles to limit the sizes of the selection and to include only relevant and focused CMH publications, possibly omitting other candidate studies. However, we considered that if a paper's principal focus was CMH, then at least one of the terms from the search strings could be expected to appear in the title, alleviating the risk of discarding relevant publications for this mapping study.

Other terms, particularly terms related to different mental health issues, might have been relevant for inclusion in the search strings, possibly impacting the results. However, our aim was to provide a high-level overview of existing CMH literature.

Other classification criteria might have been relevant to extracting further information from the selected studies, possibly affecting our findings. However, the principal aim of this systematic mapping study was to provide an overview of CMH literature, and the process and criteria presented in the *Methods* section fit this purpose.

The final study selection depended on the first two authors, ND and SO, who conducted the search process. The first author conducted a preliminary search in Scopus and constructed a preliminary selection of publications. The second author revised the selection. If a disagreement arose between the two authors on the exclusion or inclusion of a paper, then a discussion took place until an agreement was reached. It should be noted that the two authors have experience in conducting mapping studies and systematic literature reviews [228,249,250].

Conclusions

This paper conducted a systematic mapping study of 289 publications indexed in Scopus and provided an overview of the available literature on CMH research. CMH has the potential to overcome some of the mental health care delivery barriers by introducing and exploiting ICT in the process of mental health care. A total of 289 publications were selected, analyzed, and classified. The results showed that CMH is a promising research field that has gained increasing attention from researchers over the years. The frequency of the selected publications was influenced by the continuous developments in digital media and the use of ICT as well as the changes in literature evaluation methods, which mainly relied on bibliometrics. The results also showed that most of the selected CMH literature addressed general mental well-being; depression and anxiety were the mental disorders most addressed in studies on specific mental issues, which is in accordance with their global prevalence; young people were the most targeted cohort group as they are more familiar with digital solutions; exploratory research and reviews were the dominant research types found in the selected literature, which indicates that researchers are more focused on exploring possible venues of implementation of ICT in mental health care and constructing an understanding of the field; and most of the selected studies were not empirically evaluated. Moreover, the results showed that selected studies that were empirically evaluated were mostly conducted in developed countries. Screening of the selected studies that were reviews showed that they targeted specific cohort groups, specific mental disorders, specific types of solutions, or specific terms. To the best of our knowledge, this is the first mapping study that addresses the field of CMH as a whole, including all relevant terms and without excluding any specific criteria. On the basis of our findings, we recommend the following to researchers:

- Shift attention to provide evidence-based solutions and studies, and empirically evaluate existing solutions in future CMH studies.
- Focus on specific mental issues and cohort groups to better identify possible issues and limitations of the field.
- Investigate CMH solution implementation in developing countries and war zones, which experience various mental care delivery barriers.
- Use *connected mental health* as an englobing term for the exploitation of the different types of ICT in mental health care.

Practitioners can use this study to find tool-based studies for specific cohorts and/or mental disorders. They can also find studies related to the attitudes and behaviors of CMH users.

We believe that our study will provide researchers and practitioners with relevant information regarding the existing CMH literature as well as recommendations for future publications. For future work, we intend to develop a conceptual framework for sustainable CMH solutions.

Acknowledgments

This work is part of the Startup project (number 31T131) funded by the United Arab Emirates University (2019-2021).

Authors' Contributions

All authors contributed to the creation of the manuscript. ND contributed to the design, conception, acquisition and interpretation of data, classification of selected papers, drafting of the manuscript, and revision. SO contributed to the design, conception, statistical support, interpretation of data, drafting of the manuscript, and critical revision. MJ, JL, and MG performed the critical revision. All authors have read and approved this manuscript.

Conflicts of Interest

LL is a shareholder of Salumedia Labs, which is a digital health company.

Multimedia Appendix 1
Classification results.
[XLSX File (Microsoft Excel File), 44 KB - jmir_v22i8e19950_app1.xlsx]

References

1. Westerhof GJ, Keyes CL. Mental illness and mental health: the two continua model across the lifespan. J Adult Dev 2010 Jun;17(2):110-119 [FREE Full text] [doi: 10.1007/s10804-009-9082-y] [Medline: 20502508]
2. Sawyer MG, Whaites L, Rey JM, Hazell PL, Graetz BW, Baghurst P. Health-related quality of life of children and adolescents with mental disorders. J Am Acad Child Adolesc Psychiatry 2002 May;41(5):530-537. [doi: 10.1097/00004583-200205000-00010] [Medline: 12014785]
3. Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, et al. No health without mental health. Lancet 2007 Sep 8;370(9590):859-877. [doi: 10.1016/S0140-6736(07)61238-0] [Medline: 17804063]
4. Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. Lancet 2007 Apr 14;369(9569):1302-1313. [doi: 10.1016/S0140-6736(07)60368-7] [Medline: 17434406]
5. Layard R. Mental Illness Destroys Happiness and is Costless to Treat. Global Happiness. 2018. URL: http://ghc-2018.s3.amazonaws.com/GHC_Ch3.pdf [accessed 2020-04-30]
6. Sachs SJ, Layard R, Helliwell J. World Happiness Report 2018. AWS. 2018. URL: https://s3.amazonaws.com/happiness-report/2018/WHR_web.pdf [accessed 2020-04-30]
7. Collins PY, Patel V, Joestl SS, March D, Insel TR, Daar AS, et al. Grand challenges in global mental health. Nature 2011 Jul 6;475(7354):27-30 [FREE Full text] [doi: 10.1038/475027a] [Medline: 21734685]
8. Patel V, Prince M. Global mental health: a new global health field comes of age. J Am Med Assoc 2010 May 19;303(19):1976-1977 [FREE Full text] [doi: 10.1001/jama.2010.616] [Medline: 20483977]
9. Rowan K, McAlpine DD, Blewett LA. Access and cost barriers to mental health care, by insurance status, 1999-2010. Health Aff (Millwood) 2013 Oct;32(10):1723-1730 [FREE Full text] [doi: 10.1377/hlthaff.2013.0133] [Medline: 24101061]
10. Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. Health Aff (Millwood) 2009;28(3):w490-w501. [doi: 10.1377/hlthaff.28.3.w490] [Medline: 19366722]
11. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
12. Snowden LR. Barriers to effective mental health services for African Americans. Ment Health Serv Res 2001 Dec;3(4):181-187. [doi: 10.1023/a:1013172913880] [Medline: 11859964]
13. Coronavirus. World Health Organization. URL: https://www.who.int/health-topics/coronavirus#tab=tab_1 [accessed 2020-04-30]
14. Coronavirus Disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-04-30]
15. Anxiety on Rise Due to Coronavirus, Say Mental Health Charities. The Guardian. URL: https://www.theguardian.com/world/2020/mar/13/anxiety-on-rise-due-to-coronavirus-say-mental-health-charities [accessed 2020-04-30]
16. Mental Health and Psychosocial Considerations During the Covid-19 Outbreak. World Health Organization. URL: https://www.who.int/docs/default-source/coronaviruse/mental-health-considerations.pdf?sfvrsn=6d3578af_2 [accessed 2020-04-30]
17. Coronavirus Disease (COVID-19) Advice for the Public. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [accessed 2020-04-30]
18. Schmidt U, Wykes T. E-mental health - a land of unlimited possibilities. J Ment Health 2012 Aug;21(4):327-331. [doi: 10.3109/09638237.2012.705930] [Medline: 22823092]
19. Caulfield BM, Donnelly SC. What is Connected Health and why will it change your practice? QJM 2013 Aug;106(8):703-707. [doi: 10.1093/qjmed/hct114] [Medline: 23676416]
20. Karampela M, Isomursu M, Porat T, Maramis C, Mountford N, Giunti G, et al. The extent and coverage of current knowledge of connected health: systematic mapping study. J Med Internet Res 2019 Sep 25;21(9):e14394 [FREE Full text] [doi: 10.2196/14394] [Medline: 31573915]
21. Nicholas J, Huckvale K, Larsen ME, Basu A, Batterham PJ, Shaw F, et al. Issues for ehealth in psychiatry: results of an expert survey. J Med Internet Res 2017 Feb 28;19(2):e55 [FREE Full text] [doi: 10.2196/jmir.6957] [Medline: 28246068]
22. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 1;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
23. Helen C, Griffiths K, Evans K. e-Mental health in Australia: Implications of the Internet and related technologies for policy. Canberra: Commonwealth Department of Health and Ageing; 2002.

24. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed J E Health 2007 Oct;13(5):573-590. [doi: 10.1089/tmj.2006.0073] [Medline: 17999619]
25. Bauer S, Moessner M. Technology-enhanced monitoring in psychotherapy and e-mental health. J Ment Health 2012 Aug;21(4):355-363. [doi: 10.3109/09638237.2012.667886] [Medline: 22548363]
26. Wade VA, Karnon J, Elshaug AG, Hiller JE. A systematic review of economic analyses of telehealth services using real time video communication. BMC Health Serv Res 2010 Aug 10;10:233 [FREE Full text] [doi: 10.1186/1472-6963-10-233] [Medline: 20696073]
27. Mitrpanont J, Phandhu-Fung J, Klubdee N, Ratanalaor S, Mitrpanont T. iCare-Stress: An Integrated Mental Health Software. In: 2nd International Conference on Information Technology. 2017 Presented at: INCIT'17; November 2-3, 2017; Nakhonpathom, Thailand. [doi: 10.1109/incit.2017.8257889]
28. Owen JE, Kuhn E, Jaworski BK, McGee-Vincent P, Juhasz K, Hoffman JE, et al. VA mobile apps for PTSD and related problems: public health resources for veterans and those who care for them. Mhealth 2018;4:28 [FREE Full text] [doi: 10.21037/mhealth.2018.05.07] [Medline: 30148141]
29. Batterham PJ, Calear AL, Farrer L, McCallum SM, Cheng VW. Randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. Internet Interv 2018 Jun;12:91-99 [FREE Full text] [doi: 10.1016/j.invent.2017.08.002] [Medline: 30135773]
30. Jani HM. Benefiting From Online Mental Status Examination System and Mental Health Diagnostic System. In: 3rd International Conference on Information Sciences and Interaction Sciences. 2010 Presented at: ICICIS'10; June 23-25, 2010; Chengdu, China. [doi: 10.1109/icicis.2010.5534712]
31. Sahu SS, Sharma A. Detecting Brainwaves to Evaluate Mental Health Using Labview and Applications. In: International Conference on Emerging Technological Trends. 2016 Presented at: ICETT'16; October 21-22, 2016; Kollam, India. [doi: 10.1109/icett.2016.7873738]
32. Pham TD, Thang T, Oyama-Higa M. Toward the Development of a Cost-effective E-depression Detection System. In: International Conference on Computerized Healthcare. 2012 Presented at: ICCH'12; December 17-18, 2012; Hong Kong, China. [doi: 10.1109/icch.2012.6724465]
33. Peterson K, Feldt R, Mujtaba S, Mattsson M. Systematic Mapping Studies in Software Engineering. In: 12th International Conference on Evaluation and Assessment in Software Engineering. 2008 Presented at: EASE'08; June 26-27, 2008; Bari, Italy. [doi: 10.14236/ewic/ease2008.8]
34. Petersen K, Vakkalanka S, Kuzniarz L. Guidelines for conducting systematic mapping studies in software engineering: an update. Inform Software Tech 2015 Aug;64:1-18. [doi: 10.1016/j.infsof.2015.03.007]
35. Montoya FG, Alcayde A, Baños R, Manzano-Agugliaro F. A fast method for identifying worldwide scientific collaborations using the Scopus database. Telematics and Informatics 2018 Apr;35(1):168-185. [doi: 10.1016/j.tele.2017.10.010]
36. Hamel RE. The dominance of English in the international scientific periodical literature and the future of language use in science. Aila Rev 2008 Mar 7;20:53-71. [doi: 10.1075/aila.20.06ham]
37. Ammon U. The Dominance of English as a Language of Science: Effects on Other Languages and Language Communities. New York, USA: Walter de Gruyter; 2011.
38. Liverpool S, Webber H, Matthews R, Wolpert M, Edbrooke-Childs J. A mobile app to support parents making child mental health decisions: protocol for a feasibility cluster randomized controlled trial. JMIR Res Protoc 2019 Aug 14;8(8):e14571 [FREE Full text] [doi: 10.2196/14571] [Medline: 31414665]
39. Schueller SM, Glover AC, Rufa AK, Dowdle CL, Gross GD, Karnik NS, et al. A mobile phone-based intervention to improve mental health among homeless young adults: pilot feasibility trial. JMIR Mhealth Uhealth 2019 Jul 2;7(7):e12347 [FREE Full text] [doi: 10.2196/12347] [Medline: 31267980]
40. Coelhoso CC, Tobo PR, Lacerda SS, Lima AH, Barrichello CR, Amaro E, et al. A new mental health mobile app for well-being and stress reduction in working women: randomized controlled trial. J Med Internet Res 2019 Nov 7;21(11):e14269 [FREE Full text] [doi: 10.2196/14269] [Medline: 31697244]
41. Garabiles MR, Harper Shehadeh M, Hall BJ. Cultural adaptation of a scalable world health organization e-mental health program for overseas Filipino workers. JMIR Form Res 2019 Mar 29;3(1):e11600 [FREE Full text] [doi: 10.2196/11600] [Medline: 30924784]
42. Greer B, Robotham D, Simblett S, Curtis H, Griffiths H, Wykes T. Digital exclusion among mental health service users: qualitative investigation. J Med Internet Res 2019 Jan 9;21(1):e11696 [FREE Full text] [doi: 10.2196/11696] [Medline: 30626564]
43. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
44. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related internet usage and design feature preference for e-mental health programs among men and women. J Med Internet Res 2019 Mar 18;21(3):e11224 [FREE Full text] [doi: 10.2196/11224] [Medline: 30882361]

45. Birk MV, Mandryk RL. Improving the efficacy of cognitive training for digital mental health interventions through avatar customization: crowdsourced quasi-experimental study. J Med Internet Res 2019 Jan 8;21(1):e10133 [FREE Full text] [doi: 10.2196/10133] [Medline: 30622095]
46. Thomas K, Bendtsen M. Mental health promotion among university students using text messaging: protocol for a randomized controlled trial of a mobile phone-based intervention. JMIR Res Protoc 2019 Aug 15;8(8):e12396 [FREE Full text] [doi: 10.2196/12396] [Medline: 31418426]
47. Anastasiadou D, Folkvord F, Serrano-Troncoso E, Lupiañez-Villanueva F. Mobile health adoption in mental health: user experience of a mobile health app for patients with an eating disorder. JMIR Mhealth Uhealth 2019 May 31;7(6):e12920 [FREE Full text] [doi: 10.2196/12920] [Medline: 31199329]
48. Andrews JA, Brown L, Hawley M, Astell A. Older adults' perspectives on using digital technology to maintain good mental health: interactive group study. J Med Internet Res 2019 Feb 13;21(2):e11694 [FREE Full text] [doi: 10.2196/11694] [Medline: 30758292]
49. Mandryk RL, Birk MV. The potential of game-based digital biomarkers for modeling mental health. JMIR Ment Health 2019 Apr 23;6(4):e13485 [FREE Full text] [doi: 10.2196/13485] [Medline: 31012857]
50. Scholten H, Granic I. Use of the principles of design thinking to address limitations of digital mental health interventions for youth: viewpoint. J Med Internet Res 2019 Jan 14;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]
51. Thornton L, Kay-Lambkin F, Tebbutt B, Hanstock TL, Baker AL. A mobile phone-based healthy lifestyle monitoring tool for people with mental health problems (MyHealthPA): development and pilot testing. JMIR Cardio 2018 Oct 1;2(2):e10228 [FREE Full text] [doi: 10.2196/10228] [Medline: 31758772]
52. Kaur A, Isaranuwatchai W, Jaffer A, Ferguson G, Abi-Jaoude A, Johnson A, et al. A web- and mobile-based map of mental health resources for postsecondary students (Thought Spot): protocol for an economic evaluation. JMIR Res Protoc 2018 Mar 29;7(3):e83 [FREE Full text] [doi: 10.2196/resprot.9478] [Medline: 29599106]
53. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An assessment framework for e-mental health apps in Canada: results of a modified delphi process. JMIR Mhealth Uhealth 2018 Jul 9;6(7):e10016 [FREE Full text] [doi: 10.2196/10016] [Medline: 29986846]
54. March S, Day J, Ritchie G, Rowe A, Gough J, Hall T, et al. Attitudes toward e-mental health services in a community sample of adults: online survey. J Med Internet Res 2018 Feb 19;20(2):e59 [FREE Full text] [doi: 10.2196/jmir.9109] [Medline: 29459357]
55. Carolan S, de Visser RO. Employees' perspectives on the facilitators and barriers to engaging with digital mental health interventions in the workplace: qualitative study. JMIR Ment Health 2018 Jan 19;5(1):e8 [FREE Full text] [doi: 10.2196/mental.9146] [Medline: 29351900]
56. Hatch A, Hoffman JE, Ross R, Docherty JP. Expert consensus survey on digital health tools for patients with serious mental illness: optimizing for user characteristics and user support. JMIR Ment Health 2018 Jun 12;5(2):e46 [FREE Full text] [doi: 10.2196/mental.9777] [Medline: 29895514]
57. Palmius N, Saunders KE, Carr O, Geddes JR, Goodwin GM, De Vos M. Group-personalized regression models for predicting mental health scores from objective mobile phone data streams: observational study. J Med Internet Res 2018 Oct 22;20(10):e10194 [FREE Full text] [doi: 10.2196/10194] [Medline: 30348626]
58. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. J Med Internet Res 2018 Jun 8;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]
59. di Matteo D, Fine A, Fotinos K, Rose J, Katzman M. Patient willingness to consent to mobile phone data collection for mental health apps: structured questionnaire. JMIR Ment Health 2018 Aug 29;5(3):e56 [FREE Full text] [doi: 10.2196/mental.9539] [Medline: 30158102]
60. Boonstra TW, Nicholas J, Wong QJ, Shaw F, Townsend S, Christensen H. Using mobile phone sensor technology for mental health research: integrated analysis to identify hidden challenges and potential solutions. J Med Internet Res 2018 Jul 30;20(7):e10131 [FREE Full text] [doi: 10.2196/10131] [Medline: 30061092]
61. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153 [FREE Full text] [doi: 10.2196/jmir.7725] [Medline: 28490417]
62. Bidargaddi N, Musiat P, Winsall M, Vogl G, Blake V, Quinn S, et al. Efficacy of a web-based guided recommendation service for a curated list of readily available mental health and well-being mobile apps for young people: randomized controlled trial. J Med Internet Res 2017 May 12;19(5):e141 [FREE Full text] [doi: 10.2196/jmir.6775] [Medline: 28500020]
63. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
64. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
65. Mandryk RL, Birk MV. Toward game-based digital mental health interventions: player habits and preferences. J Med Internet Res 2017 Apr 20;19(4):e128 [FREE Full text] [doi: 10.2196/jmir.6906] [Medline: 28428161]

66. Bennett-Levy J, Singer J, DuBois S, Hyde K. Translating e-mental health into practice: what are the barriers and enablers to e-mental health implementation by aboriginal and Torres Strait Islander health professionals? J Med Internet Res 2017 Jan 11;19(1):e1 [FREE Full text] [doi: 10.2196/jmir.6269] [Medline: 28077347]
67. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
68. Robotham D, Satkunanathan S, Doughty L, Wykes T. Do we still have a digital divide in mental health? A five-year survey follow-up. J Med Internet Res 2016 Nov 22;18(11):e309 [FREE Full text] [doi: 10.2196/jmir.6511] [Medline: 27876684]
69. Meurk C, Leung J, Hall W, Head BW, Whiteford H. Establishing and governing e-mental health care in Australia: a systematic review of challenges and a call for policy-focussed research. J Med Internet Res 2016 Jan 13;18(1):e10 [FREE Full text] [doi: 10.2196/jmir.4827] [Medline: 26764181]
70. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mhealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016 Jun 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
71. Kahn JR, Collinge W, Soltysik R. Post-9/11 veterans and their partners improve mental health outcomes with a self-directed mobile and web-based wellness training program: a randomized controlled trial. J Med Internet Res 2016 Sep 27;18(9):e255 [FREE Full text] [doi: 10.2196/jmir.5800] [Medline: 27678169]
72. Younes N, Chollet A, Menard E, Melchior M. E-mental health care among young adults and help-seeking behaviors: a transversal study in a community sample. J Med Internet Res 2015 May 15;17(5):e123 [FREE Full text] [doi: 10.2196/jmir.4254] [Medline: 25979680]
73. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014 Jan 21;2(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2994] [Medline: 25098314]
74. Ellis LA, Collin P, Davenport TA, Hurley PJ, Burns JM, Hickie IB. Young men, mental health, and technology: implications for service design and delivery in the digital age. J Med Internet Res 2012 Nov 22;14(6):e160 [FREE Full text] [doi: 10.2196/jmir.2291] [Medline: 23171827]
75. Mohr DC, Duffecy J, Jin L, Ludman EJ, Lewis A, Begale M, et al. Multimodal e-mental health treatment for depression: a feasibility trial. J Med Internet Res 2010 Dec 19;12(5):e48 [FREE Full text] [doi: 10.2196/jmir.1370] [Medline: 21169164]
76. Bennett K, Bennett AJ, Griffiths KM. Security considerations for e-mental health interventions. J Med Internet Res 2010 Dec 19;12(5):e61 [FREE Full text] [doi: 10.2196/jmir.1468] [Medline: 21169173]
77. Mucic D, editor. E-Mental Health. Cham, UK: Springer; 2015.
78. Shoemaker EZ, Hilty DM. E-mental health improves access to care, facilitates early intervention, and provides evidence-based treatments at a distance. In: E-Mental Health. Cham, UK: Springer; 2016:43-57.
79. Mucic D, Hilty DM, Yellowlees PM. E-Mental health toward cross-cultural populations worldwide. In: E-Mental health. Cham, UK: Springer; 2016:77-91.
80. Yellowlees PM, Hilty. DM, Mucic D. Global/worldwide e-mental health: international and futuristic perspectives of telepsychiatry and the future. In: E-mental health. Cham, UK: Springer; 2016:233-249.
81. Donald H, Barb J, Robert MM. How e-mental health adds to traditional outpatient and newer models of integrated care for patients, providers, and systems. In: E-Mental Health. Cham, UK: Springer; 2016:129-149.
82. Donald MH, Erica ZS, Jay S. How to build, evaluate, and increase your e-mental health program efficiency. In: E-Mental Health. Cham, UK: Springer; 2016:59-76.
83. Donald MH, Davor M. Technology, health, and contemporary practice: How does e-mental health fit it and what does it offer? In: E-Mental Health. Cham, UK: Springer; 2016:3-27.
84. Hilty DM, Yellowlees P, Myers K, Parish M, Rabinowitz T. The effectiveness of e-mental healthvidence base, how to choose the model based on ease/cost/strength, and future areas of research. In: E-mental Health. Cham, UK: Springer; 2016:95-127.
85. Clough BA, Zarean M, Ruane I, Mateo NJ, Aliyeva TA, Casey LM. Going global: do consumer preferences, attitudes, and barriers to using e-mental health services differ across countries? J Ment Health 2019 Feb;28(1):17-25. [doi: 10.1080/09638237.2017.1370639] [Medline: 28857650]
86. Kirst M, Pridham KF, Narrandes R, Matheson F, Young L, Niedra K, et al. Examining implementation of mobile, police-mental health crisis intervention teams in a large urban center. J Ment Health 2015 Dec;24(6):369-374. [doi: 10.3109/09638237.2015.1036970] [Medline: 26383041]
87. Ennis L, Rose D, Denis M, Pandit N, Wykes T. Can't surf, won't surf: the digital divide in mental health. J Ment Health 2012 Aug;21(4):395-403 [FREE Full text] [doi: 10.3109/09638237.2012.689437] [Medline: 22712756]
88. Musiat P, Hoffmann L, Schmidt U. Personalised computerised feedback in e-mental health. J Ment Health 2012 Aug;21(4):346-354. [doi: 10.3109/09638237.2011.648347] [Medline: 22315961]
89. Harrison V, Proudfoot J, Wee PP, Parker G, Pavlovic DH, Manicavasagar V. Mobile mental health: review of the emerging field and proof of concept study. J Ment Health 2011 Dec;20(6):509-524. [doi: 10.3109/09638237.2011.608746] [Medline: 21988230]

90. Berry N, Lobban F, Bucci S. A qualitative exploration of service user views about using digital health interventions for self-management in severe mental health problems. BMC Psychiatry 2019 Jan 21;19(1):35 [FREE Full text] [doi: 10.1186/s12888-018-1979-1] [Medline: 30665384]
91. Dreier M, Ludwig J, Härter M, von dem Knesebeck O, Baumgardt J, Bock T, et al. Development and evaluation of e-mental health interventions to reduce stigmatization of suicidality - a study protocol. BMC Psychiatry 2019 May 17;19(1):152 [FREE Full text] [doi: 10.1186/s12888-019-2137-0] [Medline: 31101103]
92. Metz MJ, Elfeddali I, Krol DG, Veerbeek MA, de Beurs E, Beekman AT, et al. A digital intake approach in specialized mental health care: study protocol of a cluster randomised controlled trial. BMC Psychiatry 2017 Mar 7;17(1):86 [FREE Full text] [doi: 10.1186/s12888-017-1247-9] [Medline: 28270129]
93. Agyapong VIO, Mrklas K, Juhás M, Omeje J, Ohinmaa A, Dursun SM, et al. Cross-sectional survey evaluating Text4Mood: mobile health program to reduce psychological treatment gap in mental healthcare in Alberta through daily supportive text messages. BMC Psychiatry 2016 Nov 8;16(1):378 [FREE Full text] [doi: 10.1186/s12888-016-1104-2] [Medline: 27821096]
94. Clarke J, Proudfoot J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Effects of mental health self-efficacy on outcomes of a mobile phone and web intervention for mild-to-moderate depression, anxiety and stress: secondary analysis of a randomised controlled trial. BMC Psychiatry 2014 Sep 26;14:272 [FREE Full text] [doi: 10.1186/s12888-014-0272-1] [Medline: 25252853]
95. Krieger T, Meyer B, Sude K, Urech A, Maercker A, Berger T. Evaluating an e-mental health program ('deprexis') as adjunctive treatment tool in psychotherapy for depression: design of a pragmatic randomized controlled trial. BMC Psychiatry 2014 Oct 8;14:285 [FREE Full text] [doi: 10.1186/s12888-014-0285-9] [Medline: 25298158]
96. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014 Apr 11;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
97. Ben-Zeev D, Buck B, Hallgren K, Drake RE. Effect of mobile health on in-person service use among people with serious mental illness. Psychiatr Serv 2019 Jun 1;70(6):507-510. [doi: 10.1176/appi.ps.201800542] [Medline: 30947636]
98. Fortuna KL, Walker R, Fisher DB, Mois G, Allan S, Deegan PE. Enhancing standards and principles in digital mental health with recovery-focused guidelines for mobile, online, and remote monitoring technologies. Psychiatr Serv 2019 Dec 1;70(12):1080-1081 [FREE Full text] [doi: 10.1176/appi.ps.201900166] [Medline: 31615369]
99. Torous J, Haim A. Dichotomies in the development and implementation of digital mental health tools. Psychiatr Serv 2018 Dec 1;69(12):1204-1206 [FREE Full text] [doi: 10.1176/appi.ps.201800193] [Medline: 30256182]
100. Ben-Zeev D, Brian RM, Jonathan G, Razzano L, Pashka N, Carpenter-Song E, et al. Mobile health (mhealth) versus clinic-based group intervention for people with serious mental illness: a randomized controlled trial. Psychiatr Serv 2018 Sep 1;69(9):978-985. [doi: 10.1176/appi.ps.201800063] [Medline: 29793397]
101. Ben-Zeev D. Mobile health for mental health in West Africa: the case for Ghana. Psychiatr Serv 2018 Jul 1;69(7):741-743. [doi: 10.1176/appi.ps.201700555] [Medline: 29540120]
102. van der Krieke L, Wunderink L, Emerencia AC, de Jonge P, Sytema S. E-mental health self-management for psychotic disorders: state of the art and future perspectives. Psychiatr Serv 2014 Jan 1;65(1):33-49. [doi: 10.1176/appi.ps.201300050] [Medline: 24129842]
103. Iorfino F, Cross SP, Davenport T, Carpenter JS, Scott E, Shiran S, et al. A digital platform designed for youth mental health services to deliver personalized and measurement-based care. Front Psychiatry 2019;10:595 [FREE Full text] [doi: 10.3389/fpsyt.2019.00595] [Medline: 31507465]
104. Bucci S, Berry N, Morris R, Berry K, Haddock G, Lewis S, et al. 'They are not hard-to-reach clients. We have just got hard-to-reach services.' staff views of digital health tools in specialist mental health services. Front Psychiatry 2019;10:344 [FREE Full text] [doi: 10.3389/fpsyt.2019.00344] [Medline: 31133906]
105. Seifert A, Reinwand DA, Schlomann A. Designing and using digital mental health interventions for older adults: being aware of digital inequality. Front Psychiatry 2019;10:568 [FREE Full text] [doi: 10.3389/fpsyt.2019.00568] [Medline: 31447716]
106. Burchert S, Alkneme MS, Bird M, Carswell K, Cuijpers P, Hansen P, et al. User-centered app adaptation of a low-intensity e-mental health intervention for Syrian refugees. Front Psychiatry 2019;9:663 [FREE Full text] [doi: 10.3389/fpsyt.2018.00663] [Medline: 30740065]
107. Garrido S, Millington C, Cheers D, Boydell K, Schubert E, Meade T, et al. What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 2019;10:759 [FREE Full text] [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
108. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017;8:116. [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
109. Rehm IC, Foenander E, Wallace K, Abbott JM, Kyrios M, Thomas N. What role can avatars play in e-mental health interventions? Exploring new models of client-therapist interaction. Front Psychiatry 2016;7:186 [FREE Full text] [doi: 10.3389/fpsyt.2016.00186] [Medline: 27917128]

110. Bucci S, Ainsworth J, Barrowclough C, Lewis S, Haddock G, Berry K, et al. A theory-informed digital health intervention in people with severe mental health problems. Stud Health Technol Inform 2019 Aug 21;264:526-530. [doi: 10.3233/SHTI190278] [Medline: 31437979]
111. Househ M, Alam T, Al-Thani D, Schneider J, Siddig M, Fernandez-Luque L, et al. Developing a Digital Mental Health Platform for the Arab World: From Research to Action. 2019 Jul 04 Presented at: ICIMTH; July 2019; Athens, Greece p. 392-395 URL: http://ebooks.iospress.nl/publication/51763 [doi: 10.3233/SHTI190101]
112. Peyton D, Hiscock H, Sciberras E. Do digital health interventions improve mental health literacy or help-seeking among parents of children aged 2-12 years? A scoping review. Stud Health Technol Inform 2019 Aug 8;266:156-161. [doi: 10.3233/SHTI190788] [Medline: 31397317]
113. Strudwick G, Booth R, Strauss J. Identifying how patient portals may be effectively used among mental health populations to support digital inclusion: a study protocol. Stud Health Technol Inform 2018;250:30. [Medline: 29857361]
114. Wang J, Lam RW, Ho K, Attridge M, Lashewicz BM, Patten SB, et al. Preferred features of e-mental health programs for prevention of major depression in male workers: results from a Canadian national survey. J Med Internet Res 2016 Jun 6;18(6):e132 [FREE Full text] [doi: 10.2196/jmir.5685] [Medline: 27267782]
115. Perlich A, von Thienen J, Meinel C. Supporting talk-based mental health interventions with digital whiteboards. Stud Health Technol Inform 2014;205:433-437. [Medline: 25160221]
116. Naslund JA, Gonsalves PP, Gruebner O, Pendse SR, Smith SL, Sharma A, et al. Digital innovations for global mental health: opportunities for data science, task sharing, and early intervention. Curr Treat Options Psychiatry 2019 Dec;6(4):337-351. [doi: 10.1007/s40501-019-00186-8] [Medline: 32457823]
117. Boumparis N, Schulte MH, Riper H. Digital mental health for alcohol and substance use disorders. Curr Treat Options Psych 2019 Nov 26;6(4):352-366. [doi: 10.1007/s40501-019-00190-y]
118. Becker TD, Torous JB. Recent developments in digital mental health interventions for college and university students. Curr Treat Options Psych 2019 Jun 14;6(3):210-220. [doi: 10.1007/s40501-019-00178-8]
119. Wykes T, Lipshitz J, Schueller SM. Towards the design of ethical standards related to digital mental health and all its applications. Curr Treat Options Psych 2019 Jul 5;6(3):232-242. [doi: 10.1007/s40501-019-00180-0]
120. Schueller SM, Hunter JF, Figueroa C, Aguilera A. Use of digital mental health for marginalized and underserved populations. Curr Treat Options Psych 2019 Jul 5;6(3):243-255. [doi: 10.1007/s40501-019-00181-z]
121. Magee JC, Adut S, Brazill K, Warnick S. Mobile app tools for identifying and managing mental health disorders in primary care. Curr Treat Options Psychiatry 2018 Sep;5(3):345-362 [FREE Full text] [doi: 10.1007/s40501-018-0154-0] [Medline: 30397577]
122. Johnson KF, Kalkbrenner MT. The utilization of technological innovations to support college student mental health: mobile health communication. J Tech Hum Serv 2017 Sep 11;35(4):314-339. [doi: 10.1080/15228835.2017.1368428]
123. East ML, Havard B, Hastings NB. Mental health mobile apps' instruction: technology adoption theories applied in a mixed methods study of counseling faculty. J Tech Hum Serv 2016 Oct 21;34(4):301-325. [doi: 10.1080/15228835.2016.1233842]
124. Mar MY, Neilson EK, Torchalla I, Werker GR, Laing A, Krausz M. Exploring e-mental health preferences of generation Y. J Tech Hum Serv 2014 Nov 26;32(4):312-327. [doi: 10.1080/15228835.2014.943457]
125. Feng XL, Campbell A. Understanding e-mental health resources: personality, awareness, utilization, and effectiveness of e-mental health resources amongst youth. J Tech Hum Serv 2011 Apr;29(2):101-119. [doi: 10.1080/15228835.2011.595276]
126. Zack JS. How sturdy is that digital couch? Legal considerations for mental health professionals who deliver clinical services via the internet. J Tech Hum Serv 2008 Jul 3;26(2-4):333-359. [doi: 10.1080/15228830802097083]
127. Marshall JM, Dunstan DA, Bartik W. Smartphone psychology: new approaches towards safe and efficacious mobile mental health apps. Prof Psychol Res Pr 2020 Jun;51(3):214-222. [doi: 10.1037/pro0000278]
128. Miller KE, Kuhn E, Yu J, Owen JE, Jaworski BK, Taylor K, et al. Use and perceptions of mobile apps for patients among VA primary care mental and behavioral health providers. Prof Psychol Res Pr 2019 Jun;50(3):204-209. [doi: 10.1037/pro0000229]
129. Lui JH, Marcus DK, Barry CT. Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. Prof Psychol Res Pr 2017 Jun;48(3):199-210. [doi: 10.1037/pro0000122]
130. Jones N, Moffitt M. Ethical guidelines for mobile app development within health and mental health fields. Professional Psychology: Research and Practice 2016 Apr;47(2):155-162 [FREE Full text] [doi: 10.1037/pro0000069]
131. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. The first 30 months of the MindSpot clinic: evaluation of a national e-mental health service against project objectives. Aust N Z J Psychiatry 2017 Dec;51(12):1227-1239. [doi: 10.1177/0004867416671598] [Medline: 27733709]
132. Batterham PJ, Sunderland M, Calear AL, Davey CG, Christensen H, Teesson M, et al. Developing a roadmap for the translation of e-mental health services for depression. Aust N Z J Psychiatry 2015 Sep;49(9):776-784. [doi: 10.1177/0004867415582054] [Medline: 25907269]
133. Christensen H, Petrie K. State of the e-mental health field in Australia: where are we now? Aust N Z J Psychiatry 2013 Feb;47(2):117-120. [doi: 10.1177/0004867412471439] [Medline: 23297367]
134. Jorm AF, Morgan AJ, Malhi GS. The future of e-mental health. Aust N Z J Psychiatry 2013 Feb;47(2):104-106. [doi: 10.1177/0004867412474076] [Medline: 23382504]

135. Strauss P, Morgan H, Toussaint DW, Lin A, Winter S, Perry Y. Trans and gender diverse young people's attitudes towards game-based digital mental health interventions: a qualitative investigation. Internet Interv 2019 Dec;18:100280 [FREE Full text] [doi: 10.1016/j.invent.2019.100280] [Medline: 31890628]
136. Apolinário-Hagen J, Fritsche L, Bierhals C, Salewski C. Improving attitudes toward e-mental health services in the general population via psychoeducational information material: a randomized controlled trial. Internet Interv 2018 Jun;12:141-149. [doi: 10.1016/j.invent.2017.12.002] [Medline: 30135778]
137. Schueller SM, Washburn JJ, Price M. Exploring mental health providers' interest in using web and mobile-based tools in their practices. Internet Interv 2016 May;4(2):145-151 [FREE Full text] [doi: 10.1016/j.invent.2016.06.004] [Medline: 28090438]
138. Hensel JM, Ellard K, Koltek M, Wilson G, Sareen J. Digital health solutions for indigenous mental well-being. Curr Psychiatry Rep 2019 Jul 1;21(8):68 [FREE Full text] [doi: 10.1007/s11920-019-1056-6] [Medline: 31263971]
139. Armontrout J, Torous J, Fisher M, Drogin E, Gutheil T. Mobile mental health: navigating new rules and regulations for digital tools. Curr Psychiatry Rep 2016 Oct;18(10):91. [doi: 10.1007/s11920-016-0726-x] [Medline: 27553979]
140. Torous J, Staples P, Onnela J. Realizing the potential of mobile mental health: new methods for new data in psychiatry. Curr Psychiatry Rep 2015 Aug;17(8):602 [FREE Full text] [doi: 10.1007/s11920-015-0602-0] [Medline: 26073363]
141. Zheng X, Woo BK. E-mental health in ethnic minority: a comparison of YouTube and talk-based educational workshops in dementia. Asian J Psychiatr 2017 Feb;25:246-248. [doi: 10.1016/j.ajp.2016.12.002] [Medline: 28262161]
142. Brian RM, Ben-Zeev D. Mobile health (mHealth) for mental health in Asia: objectives, strategies, and limitations. Asian J Psychiatr 2014 Aug;10:96-100. [doi: 10.1016/j.ajp.2014.04.006] [Medline: 25042960]
143. Aggarwal NK. Applying mobile technologies to mental health service delivery in South Asia. Asian J Psychiatr 2012 Sep;5(3):225-230. [doi: 10.1016/j.ajp.2011.12.009] [Medline: 22981050]
144. Orman J, Clarke J, Whittle E, Anonuevo C, Proudfoot J. A diabetes-specific e-mental health tool: development, acceptability and outcomes of a feasibility study. Aust Fam Physician 2016 Aug;45(8):600-605 [FREE Full text] [Medline: 27610452]
145. Orman J, O'Dea B, Shand F, Berk M, Proudfoot J, Christensen H. e-Mental health for mood and anxiety disorders in general practice. Aust Fam Physician 2014 Dec;43(12):832-837. [Medline: 25705730]
146. Reid SC, Kauer S, Khor A, Hearps S, Sanci L, Kennedy A, et al. Using a mobile phone application in youth mental health - an evaluation study. Aust Fam Physician 2012 Sep;41(9):711-714. [Medline: 22962650]
147. Punukollu M, Marques M. Use of mobile apps and technologies in child and adolescent mental health: a systematic review. Evid Based Ment Health 2019 Nov;22(4):161-166. [doi: 10.1136/ebmental-2019-300093] [Medline: 31358537]
148. Sheehan R, Hassiotis A. Digital mental health and intellectual disabilities: state of the evidence and future directions. Evid Based Ment Health 2017 Nov;20(4):107-111. [doi: 10.1136/eb-2017-102759] [Medline: 28947677]
149. Davies EB, Craven MP, Martin JL, Simons L. Proportionate methods for evaluating a simple digital mental health tool. Evid Based Ment Health 2017 Nov;20(4):112-117 [FREE Full text] [doi: 10.1136/eb-2017-102755] [Medline: 28993317]
150. Damen K, Van DS. Virtual Reality as e-Mental Health to Support Starting with Mindfulness-Based Cognitive Therapy. In: International Conference on Entertainment Computing. 2018 Presented at: ICEC'18; September 18-20, 2018; Poznan, Poland. [doi: 10.1007/978-3-319-99426-0_24]
151. Chiou W, Kao C, Lo L, Huang D, Wang M, Chen B. Feasibility of Utilizing E-Mental Health with Mobile APP Interface for Social Support Enhencement: A Conceptional Solution for Postpartum Depression in Taiwan. In: International Conference of Design, User Experience, and Usability. 2017 Presented at: DUXU'17; July 9-14, 2017; Vancouver, Canada. [doi: 10.1007/978-3-319-58637-3_15]
152. Martinez SG, Badillo-Urquiola K, Leis R, Chavez J, Green T, Clements T. Investigation of Multimodal Mobile Applications for Improving Mental Health. In: International Conference on Augmented Cognition. 2016 Presented at: AC'16; July 17-22, 2016; Toronto, Canada. [doi: 10.1007/978-3-319-39952-2_32]
153. Fils-Aimé JR, Grelotti DJ, Thérosmé T, Kaiser BN, Raviola G, Alcindor Y, et al. A mobile clinic approach to the delivery of community-based mental health services in rural Haiti. PLoS One 2018;13(6):e0199313 [FREE Full text] [doi: 10.1371/journal.pone.0199313] [Medline: 29924866]
154. Sprenger M, Mettler T, Osma J. Health professionals' perspective on the promotion of e-mental health apps in the context of maternal depression. PLoS One 2017;12(7):e0180867 [FREE Full text] [doi: 10.1371/journal.pone.0180867] [Medline: 28704442]
155. Ketelaar SM, Nieuwenhuijsen K, Gärtner FR, Bolier L, Smeets O, Sluiter JK. Effect of an e-mental health approach to workers' health surveillance versus control group on work functioning of hospital employees: a cluster-RCT. PLoS One 2013;8(9):e72546 [FREE Full text] [doi: 10.1371/journal.pone.0072546] [Medline: 24069148]
156. Neilsen A, Wilson RL. Combining e-mental health intervention development with human computer interaction (HCI) design to enhance technology-facilitated recovery for people with depression and/or anxiety conditions: an integrative literature review. Int J Ment Health Nurs 2019 Feb;28(1):22-39. [doi: 10.1111/inm.12527] [Medline: 30133096]
157. de Vecchi N, Kenny A, Dickson-Swift V, Kidd S. How digital storytelling is used in mental health: a scoping review. Int J Ment Health Nurs 2016 Jun;25(3):183-193. [doi: 10.1111/inm.12206] [Medline: 26900000]

158. Lee SJ, Thomas P, Doulis C, Bowles D, Henderson K, Keppich-Arnold S, et al. Outcomes achieved by and police and clinician perspectives on a joint police officer and mental health clinician mobile response unit. Int J Ment Health Nurs 2015 Dec;24(6):538-546. [doi: 10.1111/inm.12153] [Medline: 26597480]
159. Naslund JA, Aschbrenner KA, Barre LK, Bartels SJ. Feasibility of popular m-health technologies for activity tracking among individuals with serious mental illness. Telemed J E Health 2015 Mar;21(3):213-216 [FREE Full text] [doi: 10.1089/tmj.2014.0105] [Medline: 25536190]
160. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. Telemed J E Health 2015 Dec;21(12):1038-1041. [doi: 10.1089/tmj.2015.0002] [Medline: 26171663]
161. Khasanshina EV, Wolfe WL, Emerson EN, Stachura ME. Counseling center-based tele-mental health for students at a rural university. Telemed J E Health 2008 Feb;14(1):35-41. [doi: 10.1089/tmj.2006.0038]
162. Torous J, Kossowsky J. Digital Mental Health for Youth: New Evidence but Still Much Unknown. Psychiatric Times. 2018. URL: https://www.psychiatrictimes.com/view/digital-mental-health-youth-new-evidence-still-much-unknown [accessed 2020-08-19]
163. Peek H. Harnessing Social Media and Mobile Apps for Mental Health. Psychiatric Times. 2015. URL: https://www.psychiatrictimes.com/view/harnessing-social-media-and-mobile-apps-mental-health [accessed 2020-08-19]
164. Brooks L, Schirmer T. Mobile Apps for Mental Health. Psychiatric Times. 2014. URL: https://www.psychiatrictimes.com/view/mobile-apps-mental-health
165. Aschbrenner KA, Naslund JA, Tomlinson EF, Kinney A, Pratt SI, Brunette MF. Adolescents' use of digital technologies and preferences for mobile health coaching in public mental health settings. Front Public Health 2019;7:178 [FREE Full text] [doi: 10.3389/fpubh.2019.00178] [Medline: 31312629]
166. Rojas G, Martínez V, Martínez P, Franco P, Jiménez-Molina A. Improving mental health care in developing countries through digital technologies: a mini narrative review of the Chilean case. Front Public Health 2019;7:391 [FREE Full text] [doi: 10.3389/fpubh.2019.00391] [Medline: 31921754]
167. Moock J. Support from the internet for individuals with mental disorders: advantages and disadvantages of e-mental health service delivery. Front Public Health 2014;2:65 [FREE Full text] [doi: 10.3389/fpubh.2014.00065] [Medline: 24967221]
168. Aryana B, Brewster L. Design for mobile mental health: exploring the informed participation approach. Health Informatics J 2019;26(2):1208-1224. [doi: 10.1177/1460458219873540] [Medline: 31566470]
169. Kenny R, Fitzgerald A, Segurado R, Dooley B. Is there an app for that? A cluster randomised controlled trial of a mobile app-based mental health intervention. Health Informatics J 2019;26(3):1538-1559. [doi: 10.1177/1460458219884195] [Medline: 31702409]
170. Kenny R, Dooley B, Fitzgerald A. Developing mental health mobile apps: exploring adolescents' perspectives. Health Informatics J 2016 Jun;22(2):265-275. [doi: 10.1177/1460458214555041] [Medline: 25385165]
171. Sreejith G, Menon V. Mobile phones as a medium of mental health care service delivery: perspectives and barriers among patients with severe mental illness. Indian J Psychol Med 2019;41(5):428-433 [FREE Full text] [doi: 10.4103/IJPSYM.IJPSYM_333_18] [Medline: 31548765]
172. Basavarajappa C, Chand P. Digital platforms for mental health-care delivery. Indian J Psychol Med 2017;39(5):703-706 [FREE Full text] [doi: 10.4103/IJPSYM.IJPSYM_209_17] [Medline: 29200576]
173. Singh G, Manjunatha N, Rao S, Shashidhara H, Moirangthem S, Madegowda R, et al. Use of mobile phone technology to improve follow-up at a community mental health clinic: a randomized control trial. Indian J Psychol Med 2017;39(3):276-280 [FREE Full text] [doi: 10.4103/0253-7176.207325] [Medline: 28615760]
174. Graham AK, Lattie EG, Mohr DC. Experimental therapeutics for digital mental health. JAMA Psychiatry 2019 Aug 21:-epub ahead of print. [doi: 10.1001/jamapsychiatry.2019.2075] [Medline: 31433448]
175. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry 2017 May 1;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
176. Wynn SD, Sherrod RA. Providing mental health care to veterans in rural areas: using telehealth in mobile clinics. J Psychosoc Nurs Ment Health Serv 2012 Jun;50(6):22-28. [doi: 10.3928/02793695-20120508-01] [Medline: 22589227]
177. Tschirch P, Walker G, Calvacca LT. Nursing in tele-mental health. J Psychosoc Nurs Ment Health Serv 2006 May;44(5):20-27. [doi: 10.3928/02793695-20060501-05] [Medline: 16719184]
178. Sukmawati I, Ardi Z, Ifdil I, Zikra Z. Development and validation of acceptability of mental-health mobile app survey (AMMS) for android-based online counseling service assessment. J Phys: Conf Ser 2019 Dec 16;1339:012124. [doi: 10.1088/1742-6596/1339/1/012124]
179. Ardi Z, Sukmawati I, Ifdil I, Afdal A, Rangka IB, Suranata K. Exploring the acceptability of internet-based mental health mobile app services using network psychometrics analysis. J Phys: Conf Ser 2018 Dec 7;1114:012106. [doi: 10.1088/1742-6596/1114/1/012106]
180. Tielman ML, Neerincx MA, Pagliari C, Rizzo A, Brinkman W. Considering patient safety in autonomous e-mental health systems - detecting risk situations and referring patients back to human care. BMC Med Inform Decis Mak 2019 Mar 18;19(1):47 [FREE Full text] [doi: 10.1186/s12911-019-0796-x] [Medline: 30885190]

181. Pelletier J, Rowe M, François N, Bordeleau J, Lupien S. No personalization without participation: on the active contribution of psychiatric patients to the development of a mobile application for mental health. BMC Med Inform Decis Mak 2013 Jul 27;13:78. [doi: 10.1186/1472-6947-13-78] [Medline: 23890085]
182. Williams A, Fossey E, Farhall J, Foley F, Thomas N. Going online together: the potential for mental health workers to integrate recovery oriented e-mental health resources into their practice. Psychiatry 2018;81(2):116-129. [doi: 10.1080/00332747.2018.1492852] [Medline: 30273103]
183. Morganstein J. Mobile applications for mental health providers. Psychiatry 2016;79(4):358-363. [doi: 10.1080/00332747.2016.1237753] [Medline: 27997318]
184. Rooksby J, Morrison A, Murray-Rust D. Student Perspectives on Digital Phenotyping: The Acceptability of Using Smartphone Data to Assess Mental Health. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI'19; May 4-19, 2019; Glasgow, Scotland, UK. [doi: 10.1145/3290605.3300655]
185. Sakata J, Zhang M, Pu S, Xing J, Versha K. Beam: a Mobile Application to Improve Happiness and Mental Health. In: CHI '14 Extended Abstracts on Human Factors in Computing Systems. 2014 Presented at: CHI EA'14; April 26-May 1, 2014; Toronto, Canada. [doi: 10.1145/2559206.2580931]
186. Adewuya AO, Momodu O, Olibamoyo O, Adegbaju A, Adesoji O, Adegbokun A. The effectiveness and acceptability of mobile telephone adherence support for management of depression in the mental health in primary care (MeHPriC) project, Lagos, Nigeria: a pilot cluster randomised controlled trial. J Affect Disord 2019 Jun 15;253:118-125. [doi: 10.1016/j.jad.2019.04.025] [Medline: 31035212]
187. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. J Affect Disord 2018 Feb;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
188. Seko Y, Kidd S, Wiljer D, McKenzie K. Youth mental health interventions via mobile phones: a scoping review. Cyberpsychol Behav Soc Netw 2014 Sep;17(9):591-602. [doi: 10.1089/cyber.2014.0078] [Medline: 25007383]
189. Casey LM, Joy A, Clough BA. The impact of information on attitudes toward e-mental health services. Cyberpsychol Behav Soc Netw 2013 Aug;16(8):593-598. [doi: 10.1089/cyber.2012.0515] [Medline: 23679567]
190. Hendrikoff L, Kambeitz-Ilankovic L, Pryss R, Senner F, Falkai P, Pogarell O, et al. Prospective acceptance of distinct mobile mental health features in psychiatric patients and mental health professionals. J Psychiatr Res 2019 Feb;109:126-132. [doi: 10.1016/j.jpsychires.2018.11.025] [Medline: 30530207]
191. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
192. Schellong J, Lorenz P, Weidner K. Proposing a standardized, step-by-step model for creating post-traumatic stress disorder (PTSD) related mobile mental health apps in a framework based on technical and medical norms. Eur J Psychotraumatol 2019;10(1):1611090 [FREE Full text] [doi: 10.1080/20008198.2019.1611090] [Medline: 31143412]
193. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6:27882 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
194. Teles A, Rodrigues I, Viana D, Silva F, Coutinho L, Endler M, et al. Mobile Mental Health: A Review of Applications for Depression Assistance. In: 32nd International Symposium on Computer-Based Medical Systems. 2019 Presented at: CBMS'19; June 5-7, 2019; Cordoba, Spain. [doi: 10.1109/cbms.2019.00143]
195. Teles AS, Silva F, Rocha A, Lopes J, O'Sullivan D, Van DV, et al. Towards Situation-Aware Mobile Applications in Mental Health. In: 29th International Symposium on Computer-Based Medical Systems. 2016 Presented at: CBMS'16; June 20-24, 2016; Dublin, Ireland. [doi: 10.1109/cbms.2016.7]
196. Gindidis S, Stewart S, Roodenburg J. A systematic scoping review of adolescent mental health treatment using mobile apps. Adv Mental Health 2019;17(2):161-177. [doi: 10.1080/18387357.2018.1523680]
197. Dragovic M, Davison S, Morgan VA, Chiu VW, Richards N, Vatskalis T, et al. 'Validated, easy to use and free': top three requests for mobile device applications ('apps') from mental health consumers and clinicians. Adv Mental Health 2018 Dec 11:1-9. [doi: 10.1080/18387357.2018.1557014]
198. Apolinário-Hagen J. Current perspectives on e-mental-health self-help treatments: Exploring the 'black box' Of public views, perceptions, and attitudes toward the digitalization of mental health care. In: The Digitization of Healthcare. London, UK: Palgrave Macmillan; 2017:205-223.
199. Eftychiou L, El Morr C. Mobile mental health virtual communities: challenges and opportunities. In: The Digitization of Healthcare. London, UK: Palgrave Macmillan; 2017:257-275.
200. Bakker D, Rickard N. Engagement with a cognitive behavioural therapy mobile phone app predicts changes in mental health and wellbeing: MoodMission. Aust Psychol 2019 Mar 13;54(4):245-260. [doi: 10.1111/ap.12383]
201. Dingwall KM, Puszka S, Sweet M, Nagel T. 'Like drawing into sand': acceptability, feasibility, and appropriateness of a new e-mental health resource for service providers working with aboriginal and Torres Strait Islander people. Aus Psychol 2015 Jan 12;50(1):60-69. [doi: 10.1111/ap.12100]
202. Cormode G, Krishnamurthy B. Key Differences Between Web 1.0 and Web 2.0. First Monday. 2008. URL: https://firstmonday.org/ojs/index.php/fm/article/view/2125 [accessed 2020-08-19]

203. O'Reilly T. Web 2.0: Compact Definition? Radar – O'Reilly. 2005. URL: http://radar.oreilly.com/2005/10/web-20-compact-definition.html [accessed 2020-04-30]
204. Murugesan S. Understanding Web 2.0. IEEE Computer Society. 2007. URL: https://www.computer.org/csdl/magazine/it/2007/04/f4034/13rRUyYSWoO [accessed 2020-08-19]
205. Number of Internet Users Worldwide From 2005 to 2019. Statista. 2020. URL: https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/
206. Number of Smartphones Sold to End Users Worldwide From 2007 to 2020. Statista. 2020. URL: https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/
207. Share of Households That Use Desktop or Portable Computers to Access the Internet in Germany From 2003 to 2014. Statista. 2014. URL: https://www.statista.com/statistics/379251/internet-usage-via-desktop-or-mobile-computer-germany/ [accessed 2020-04-30]
208. Share of Households That Use Desktop or Portable Computers to Access the Internet in Portugal From 2003 to 2014. Statista. 2014. URL: https://www.statista.com/statistics/380431/internet-usage-via-desktop-or-mobile-computer-portugal/ [accessed 2020-04-30]
209. Share of Households That Use Desktop or Portable Computers to Access the Internet in Lithuania From 2003 to 2014. Statista. 2014. URL: https://www.statista.com/statistics/379295/internet-usage-via-desktop-or-mobile-computer-lithuania/ [accessed 2020-04-30]
210. Giannopoulos G. The application of information and communication technologies in transport. Eur J Oper Res 2004 Jan;152(2):302-320. [doi: 10.1016/s0377-2217(03)00026-2]
211. Buhalis D, O'Connor P. Information communication technology revolutionizing tourism. Tour Recreat Res 2015 Jan 12;30(3):7-16. [doi: 10.1080/02508281.2005.11081482]
212. Yusuf MO. Information and communication technology and education: analysing the Nigerian national policy for information technology. Int Educ J 2005;6(3):316-321 [FREE Full text]
213. Shachaf P. Cultural diversity and information and communication technology impacts on global virtual teams: An exploratory study. Information & Management 2008 Mar;45(2):131-142. [doi: 10.1016/j.im.2007.12.003]
214. Shiferaw F, Zolfo M. The role of information communication technology (ICT) towards universal health coverage: the first steps of a telemedicine project in Ethiopia. Glob Health Action 2012;5:1-8 [FREE Full text] [doi: 10.3402/gha.v5i0.15638] [Medline: 22479235]
215. Kurtz MJ, Bollen J. Usage bibliometrics. Ann Rev Info Sci Tech 2012 Jan 18;44(1):1-64. [doi: 10.1002/aris.2010.1440440108]
216. Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. Bibliometrics: the Leiden Manifesto for research metrics. Nature 2015 Apr 23;520(7548):429-431. [doi: 10.1038/520429a] [Medline: 25903611]
217. Stephan P, Veugelers R, Wang J. Reviewers are blinkered by bibliometrics. Nature 2017 Apr 26;544(7651):411-412. [doi: 10.1038/544411a] [Medline: 28447652]
218. Haustein S, Larivière V. The use of bibliometrics for assessing research: possibilities, limitations and adverse effects. In: Incentives and Performance. Cham, UK: Springer; 2015:121-139.
219. Agarwal A, Durairajanayagam D, Tatagari S, Esteves S, Harlev A, Henkel R, et al. Bibliometrics: tracking research impact by selecting the appropriate metrics. Asian J Androl 2016;18(2):296-309 [FREE Full text] [doi: 10.4103/1008-682X.171582] [Medline: 26806079]
220. Werner R. The focus on bibliometrics makes papers less useful. Nature 2015 Jan 15;517(7534):245. [doi: 10.1038/517245a] [Medline: 25592498]
221. Nickelson DW. Telehealth and the evolving health care system: strategic opportunities for professional psychology. Prof Psychol Res Pr 1998;29(6):527-535. [doi: 10.1037/0735-7028.29.6.527]
222. Alberta Heritage Foundation for Medical Research, Simpson J. An Assessment of Routine Telepsychiatry Services. Ponoka, Alberta: Alberta Heritage Foundation for Medical Research; 1999.
223. Della Mea V. What is e-health (2): the death of telemedicine? J Med Internet Res 2001;3(2):E22 [FREE Full text] [doi: 10.2196/jmir.3.2.e22] [Medline: 11720964]
224. Allen A. Morphing telemedicine-telecare-telehealth-ehealth. Telemed Today 2000;1(43):-.
225. Number of Smartphone Users Worldwide From 2016 to 2021. Statista. 2020. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2020-04-30]
226. Emerging mHealth: Paths for Growth. PWC. 2014. URL: https://www.pwc.com/gx/en/healthcare/mhealth/assets/pwc-emerging-mhealth-full.pdf [accessed 2020-04-30]
227. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
228. Drissi N, Ouhbi S, García-Berná J, Janati Idrissi MA, Ghogho M. Sensor-based Solutions for Mental Healthcare: A Systematic Literature Review. In: 13th International Conference on Health Informatics. 2020 Presented at: HEALTHINF'20; February 24-26, 2020; Valletta, Malta p. 593-600. [doi: 10.5220/0009100805930600]
229. Shewan LG, Coats AJ. Ethics in the authorship and publishing of scientific articles. Int J Cardiol 2010 Sep;144(1):1-2. [doi: 10.1016/j.ijcard.2010.07.030]

230. Ammenwerth E, de Keizer N. A viewpoint on evidence-based health informatics, based on a pilot survey on evaluation studies in health care informatics. J Am Med Inform Assoc 2007;14(3):368-371 [FREE Full text] [doi: 10.1197/jamia.M2276] [Medline: 17329724]
231. Sunitha S, Gururaj G. Health behaviours & problems among young people in India: cause for concern & call for action. Indian J Med Res 2014 Aug;140(2):185-208 [FREE Full text] [Medline: 25297351]
232. Stang J, Story M. Guidelines for Adolescent Nutrition Services. University of Minnesota. 2008. URL: https://books.google.ae/books/about/Guidelines_for_Adolescent_Nutrition_Serv.html?id=t-zJDAEACAAJ&redir_esc=y
233. Garber J, Frankel SA, Herrington CG. Developmental demands of cognitive behavioral therapy for depression in children and adolescents: cognitive, social, and emotional processes. Annu Rev Clin Psychol 2016;12:181-216 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032814-112836] [Medline: 27019397]
234. Nalwa K, Anand AP. Internet addiction in students: a cause of concern. Cyberpsychol Behav 2003 Dec;6(6):653-656. [doi: 10.1089/109493103322725441] [Medline: 14756932]
235. Young K, Pistner M, O'Mara J, Buchanan J. Cyber disorders: the mental health concern for the new millennium. Cyberpsychol Behav 1999;2(5):475-479. [doi: 10.1089/cpb.1999.2.475] [Medline: 19178220]
236. Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf
237. Fonseca A, Gorayeb R, Canavarro MC. Women's use of online resources and acceptance of e-mental health tools during the perinatal period. Int J Med Inform 2016 Oct;94:228-236. [doi: 10.1016/j.ijmedinf.2016.07.016] [Medline: 27573331]
238. Kim J, Lim S, Lee B, Lee J. Detecting Depression of Cancer Patients with Daily Mental Health Logs from Mobile Applications. Association for Information Systems (AIS) eLibrary. 2015. URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1551&context=icis2015 [accessed 2020-08-19]
239. Drissi N, Ouhbi S, Janati Idrissi MA, Ghogho M. Mobile Apps for Post Traumatic Stress Disorder. In: Conf Proc IEEE Eng Med Biol Soc. 2019 Jul Presented at: EMBC'19; July 23-27, 2019; Berlin, Germany p. 4279-4282 URL: https://ieeexplore.ieee.org/abstract/document/8857197 [doi: 10.1109/EMBC.2019.8857197]
240. Excellence in Research for Australia. Australian Research Council. 2019. URL: https://www.arc.gov.au/excellence-research-australia [accessed 2020-04-30]
241. Research Opportunities. Australian Government. 2020. URL: https://www.studyinaustralia.gov.au/English/Australian-Education/Universities-Higher-Education/research-opportunities
242. About the Australasian Institute of Digital Health. Australasian Institute of Digital Health. 2020. URL: https://digitalhealth.org.au/about/ [accessed 2020-04-30]
243. Hather GJ, Haynes W, Higdon R, Kolker N, Stewart EA, Arzberger P, et al. The United States of America and scientific research. PLoS One 2010 Aug 16;5(8):e12203 [FREE Full text] [doi: 10.1371/journal.pone.0012203] [Medline: 20808949]
244. 50 Leaders in Health IT. Becker's Hospital Review - Healthcare News. 2015. URL: https://www.beckershospitalreview.com/healthcare-information-technology/50-leaders-in-health-it.html [accessed 2020-04-30]
245. Hashemi B, Ali S, Awaad R, Soudi L, Housel L, Sosebee SJ. Facilitating mental health screening of war-torn populations using mobile applications. Soc Psychiatry Psychiatr Epidemiol 2017 Jan;52(1):27-33. [doi: 10.1007/s00127-016-1303-7] [Medline: 27815623]
246. Phillips EA, Gordeev VS, Schreyögg J. Effectiveness of occupational e-mental health interventions: a systematic review and meta-analysis of randomized controlled trials. Scand J Work Environ Health 2019 Nov 1;45(6):560-576. [doi: 10.5271/sjweh.3839] [Medline: 31184758]
247. Richardson LK, Frueh BC, Grubaugh A, Egede L, Elhai J. Current directions in videoconferencing tele-mental health research. Clin Psychol (New York) 2009 Sep 1;16(3):323-338 [FREE Full text] [doi: 10.1111/j.1468-2850.2009.01170.x] [Medline: 20161010]
248. 249 AD, Sivertsen G. A criteria-based assessment of the coverage of Scopus and web of science. J Data Inf Sci 2019;4(1):1-21. [doi: 10.2478/jdis-2019-0001]
249. Karampela M, Ouhbi S, Isomursu M. Personal health data: a systematic mapping study. Int J Med Inform 2018 Oct;118:86-98. [doi: 10.1016/j.ijmedinf.2018.08.006] [Medline: 30153927]
250. Souza-Pereira L, Pombo N, Ouhbi S, Felizardo V, Garcia N. Clinical decision support systems for chronic diseases: a systematic literature review. Comput Methods Programs Biomed 2020 May 23;195:105565. [doi: 10.1016/j.cmpb.2020.105565] [Medline: 32480191]

Abbreviations

CMH: connected mental health
COVID-19: coronavirus disease
EC: exclusion criteria
e-mental: electronic mental
ICT: information and communication technologies
m-mental: mobile mental

MQ: mapping question
RCTs: randomized control trials
WoS: Web of Science

Edited by T Rashid Soron; submitted 07.05.20; peer-reviewed by S Chen, S Brigitte, P Matthews, D Hidalgo-Mazzei; comments to author 14.06.20; revised version received 02.07.20; accepted 28.07.20; published 28.08.20.

Please cite as:
Drissi N, Ouhbi S, Janati Idrissi MA, Fernandez-Luque L, Ghogho M
Connected Mental Health: Systematic Mapping Study
J Med Internet Res 2020;22(8):e19950
URL: http://www.jmir.org/2020/8/e19950/
doi:10.2196/19950
PMID:32857055

Review

Web-Based Interventions for Dietary Behavior in Adults With Type 2 Diabetes: Systematic Review of Randomized Controlled Trials

Jedha Dening[1], MNutr; Sheikh Mohammed Shariful Islam[1], PhD; Elena George[1], PhD; Ralph Maddison[1], PhD

Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Geelong, Victoria, Australia

Corresponding Author:
Sheikh Mohammed Shariful Islam, PhD
Institute for Physical Activity and Nutrition
School of Exercise and Nutrition Sciences
Deakin University
Locked Bag 20000
Geelong, Victoria, 3220
Australia
Phone: 61 3 924 68393
Email: shariful.islam@deakin.edu.au

Abstract

Background: Type 2 diabetes mellitus (T2DM) is among the most prevalent noncommunicable health conditions worldwide, affecting over 500 million people globally. Diet is a key aspect of T2DM management with dietary modification shown to elicit clinically meaningful outcomes such as improved glycemic control, and reductions in weight and cardiovascular disease risk factors. Web-based interventions provide a potentially convenient and accessible method for delivering dietary education, but its effects on dietary behavior in people with T2DM are unknown.

Objective: The objective of this review was to determine the effectiveness of web-based interventions on dietary behavior change and glycemic control in people with T2DM.

Methods: Per PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, systematic literature searches were performed using Medline, Embase, The Cochrane Library, and CINAHL to retrieve papers from January 2013 to May 2019. Randomized controlled trials of web-based interventions in adults with T2DM with reported dietary assessment were included. Population and intervention characteristics, dietary guidelines and assessments, and significant clinical outcomes were extracted. Differences between groups and within groups were assessed for dietary behavior and clinical outcomes.

Results: There were 714 records screened, and five studies comprising 1056 adults were included. Studies measured dietary changes by assessing overall diet quality, changes in specific dietary components, or dietary knowledge scores. Significant improvements in dietary behavior were reported in four out of the five studies, representing healthier food choices, improvements in eating habits, reductions in carbohydrates, added sugar, sodium, saturated fat and overall fat intake, and/or increases in dietary knowledge. Three studies found significant mean reductions for hemoglobin A1c ranging from –0.3% to –0.8%, and/or weight ranging from –2.3 kg to –12.7 kg, fasting blood glucose (–1 mmol/L), waist circumference (–1 cm), and triglycerides (–60.1 mg/dL). These studies provided varied dietary recommendations from standard dietary guidelines, national health program guidelines, and a very low carbohydrate ketogenic diet.

Conclusions: This review provided evidence that web-based interventions may be an effective way to support dietary behavior change in people with T2DM, potentially leading to changes in glycemic control and other clinical outcomes. However, the evidence should be viewed as preliminary as there were only five studies included with considerable heterogeneity in terms of the diets recommended, the dietary assessment measures used, the complexity of the interventions, and the modes and methods of delivery.

(J Med Internet Res 2020;22(8):e16437) doi:10.2196/16437

KEYWORDS

type 2 diabetes; dietary behavior; diet; glycemic control; self-management; eHealth; web-based; HbA1c

Introduction

Type 2 diabetes mellitus (T2DM) is among the most prevalent noncommunicable diseases worldwide, estimated in 2018 to affect more than 500 million people across 45 countries [1]. T2DM is a metabolic disorder characterized by hyperglycemia; therefore, obtaining glycemic control is the overarching goal of T2DM treatment [2]. Glycemic control is generally defined as a hemoglobin A_{1c} (HbA_{1c}) level of 7% or less. However, it is recognized that tighter glycemic control of 6.5% or less may further reduce the risk of macro and microvascular complications such as cardiovascular disease (CVD), neuropathy, nephropathy, and retinopathy [2], along with decreasing the risk of mortality [3]. The American Diabetes Association (ADA) [2] recommends education aimed at lifestyle modifications, including diet and physical activity, and prescription of medications as necessary for the management of T2DM.

It is well recognized that improving diet can optimize glycemic control, with research demonstrating improved diet quality can reduce HbA_{1c} to a similar or greater level than the provision of medication to patients with T2DM [2,4]. Even where medication is necessary, diet remains an important component of the overall treatment plan for people with T2DM [2]. Additionally, improved diet quality may optimize weight management, blood pressure, and lipid profile, which in turn may decrease the risk of CVD and stroke in people with T2DM [4]. Previous qualitative studies including people with T2DM, have confirmed that dietary support is one of their leading preferences for self-management education [5-8]. Nevertheless, there remains a substantial gap between the need for and the provision of dietary education for those with T2DM. Initial and ongoing diabetes education is the role of an individuals' multidisciplinary care team and/or provided by health care professionals or organizations through structured diabetes education programs [9]. However, rates of receiving any type of diabetes education are reported to be low globally, 23-66% in the United States [10], 11% in the United Kingdom [11], and 40% in Australia [9]. One explanation is access to and availability of health care professionals [6,12], the number of people with T2DM outweighing the number of health care professionals available to provide diabetes education [12]. Additionally, the cost and labor of delivering diabetes education programs face-to-face represent a significant challenge to organizations [13,14].

For the past two decades, researchers have become increasingly interested in providing diabetes education via technological means, as it represents a delivery method that has greater reach and access for people with T2DM [15], and is potentially more cost-effective [16]. Compared to usual care, web-based programs in people with T2DM have been shown to reduce HbA_{1c} by 0.47%-1.49% [17], while mobile health (mHealth) interventions reduced HbA_{1c} by an average of 0.8% [18]. Interventions using mobile apps show reductions in HbA_{1c} of 0.4%-1.9% [19], and the provision of telehealth has been associated with an average HbA_{1c} reduction of 0.17% [20]. In terms of the influence of intervention features on HbA_{1c} outcomes in people with T2DM, when compared to mHealth and telehealth interventions, statistically significant results were only found for web-based interventions, which may indicate that web-based interventions are particularly useful for eliciting behavior change in people with T2DM [21].

Although people with T2DM need and want dietary education, web-based interventions to date have overwhelmingly focused on overall self-management [10,21,22]. While some interventions have included a healthy eating component within the intervention package, assessment of dietary adherence or behavior remains scarce. To date, we are not aware of any review that has investigated the effects of web-based interventions on change in dietary behavior in people with T2DM. Therefore, the primary aim of this systematic review was to identify and synthesize the available evidence from randomized controlled trials (RCTs) and determine the effectiveness of web-based interventions on dietary behavior change and glycemic control in people with T2DM.

Methods

This systematic review was conducted per the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [23]. The review protocol was registered at PROSPERO (International Prospective Register of Systematic Reviews) (2018 #CRD42018109312).

Eligibility Criteria

Web-based interventions were included and defined as web-based if participants received information and directly interfaced with the internet, but they were not required to input data to a website [24]. In addition, studies were included if they were published in English; were RCTs or pilot RCTs; included an assessment of nutrition, diet, or dietary behavior; included adult participants (≥18 years) with diagnosed T2DM. Exclusion criteria were studies using non–web-based digital interventions; studies including participants with prediabetes or type 1 diabetes; studies including a combination of T2DM and participants with other types of diabetes or where outcomes for multiple chronic diseases were assessed; and studies focused on diabetes prevention.

Information Sources

A systematic literature search of four electronic databases, Medline, Embase, The Cochrane Library, and CINAHL for relevant papers published between January 2013 and May 2019, was conducted. Papers published before January 2013 were excluded as previous systematic reviews have reported on web-based, computer-based, and digital interventions in people with T2DM up to this date [22,25,26]. From these reviews, we extracted papers that met our inclusion criteria. Additionally, we conducted an in-depth exploration of reference lists for related papers and searched grey literature, including Google Scholar.

Search Strategy

Keywords used in the search were ("Type 2 Diabet*" OR "diabetes mellitus, type 2" OR "T2DM" OR "T2D") AND ("web-based" OR "internet" OR "online" OR "digital" OR "information technology" OR "IT" OR "computer-assisted"

OR “computer-based” OR “computer interface” OR “ehealth”) AND (“diet*” OR “nutrition” OR “self-management” OR “lifestyle modification”).

Data Extraction

Search results were merged using reference management software (Endnote 8) and duplicate papers removed. Screening of the titles and abstracts for individual studies were conducted in duplicate, independently, by two authors (JD, SI) with disagreements resolved by consensus. Articles deemed eligible for full-text review were assessed for eligibility by two authors, independently (JD, SI), and disagreements for inclusion were reached via consensus. Corresponding papers from the same study were merged to extract relevant outcomes. The following parameters were extracted from included studies: author/date, study design, sample size, total study period and length of follow up, population characteristics (including location, age, and comorbidities), behavioral change theory or model, digital intervention characteristics (including the type of digital intervention and duration of exposure), intervention providers, type of dietary intervention or guidelines administered, the dietary assessment used, dietary behavior change outcomes, and significant T2DM clinical outcomes including glycemic control as indicated by blood glucose levels or HbA_{1c}, and biomarkers weight, waist circumference, and CVD biomarkers. We also extracted intervention components, website usability rates, modes and methods of delivery, and process evaluation measures.

Assessment of Study Risk of Bias

The risk of bias was assessed by two researchers independently using the Cochrane risk-of-bias tool for randomized trials (RoB 2) [27]. RoB 2 provides an in-depth framework structured into five domains to assess the risk of bias in RCTs. These five domains assess the risk of bias arising from the randomization process, deviations from intended interventions, missing outcome data, outcome measurements, and selection of reported results. During the assessment, each domain is given a rating of low, high, or unclear. Conflicting assessment ratings were resolved by consensus. Results were then calculated to reach a quality assessment rating of poor, fair, or good quality.

Data Analysis

For qualitative analysis, differences in end intervention measures between groups and change between groups were reported, depending on the outcomes reported for individual studies. We also reported significant within-group changes. Data were considered statistically significant if the reported *P* value was <.05.

Results

The search identified a total of 714 papers, of which 15 full-text articles were assessed for eligibility, and five studies met the inclusion criteria and were included in this review (Figure 1) [28-32]. Corresponding papers from the same studies [33-37] provided supporting information extracted for this review. Four of the studies reviewed were RCTs, and one was a pilot RCT. Each study described a web-based intervention where participants received information that directly interfaced with the internet and included some form of dietary assessment. All the studies included only adults with T2DM; a summary of the included studies is shown in Table 1.

Figure 1. Prisma Flow Chart.

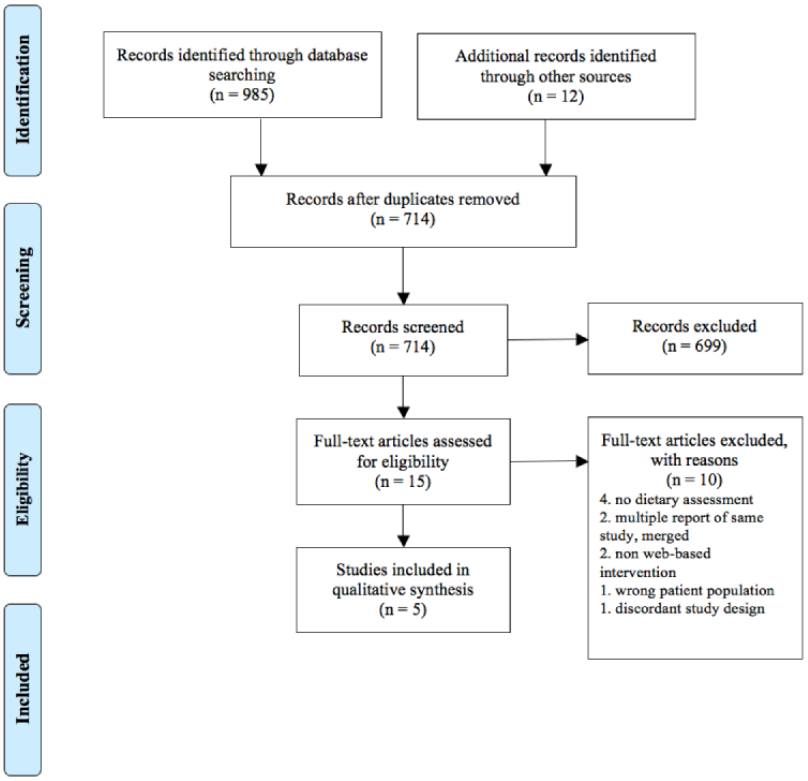


Table 1. Summary of study characteristics for included randomized controlled trials.

Author/date	Study design	Sample size, N	Population characteristics	Theory/model	Intervention participants, N, program name, digital intervention characteristics	Type of diets prescribed	Intervention/follow-up period
Ramadas et al, 2018 [29]	RCT[a]	128	Adults with T2DM[b], mean age 50.5 years, most with a family history of T2DM. Malaysia	Transtheoretical Model Stages of Change, user-centered design	N=62, myDIDeA[c], received personalized intensive dietary intervention via website + standard care	IG[d]: standard dietary guidelines CG[e]: no prescribed diet	6-month intervention, 12-month follow-up
Hansel et al, 2017 [28]	RCT	120	Adults with abdominal obesity and T2DM, mean age 57 years. Paris	No theory reported	N=60, ANODE[f], fully automated web-based nutritional support program	IG: National Nutrition & Health Program guidelines CG: received general diet advice	16-week intervention
Saslow et al, 2017 [30]	Pilot RCT	25	Overweight adults with T2DM, mean age 55.6 years. California	No theory reported	N=12, dietary instruction provided via email lessons including mindfulness training and lifestyle advice	IG: VLKCD[g] (20-50 g/carbs per day) CG: ADA[h] "create your plate" diet	16-week intervention, 32-week follow-up
Glasgow et al, 2003 [31]	RCT	320	Adults with T2DM, mean age 59 years. Colorado	Self-efficacy theory, social support theory	N=N/A[i], D-Net[j], two intervention arms: tailored self-management training, peer support	All groups received general healthy eating advice to decrease fat and increase fruit and vegetable intake	3-month intervention, 10-month follow-up
Glasgow et al, 2012 [32]	RCT	463	Adults with T2DM, mean age 58.4 years. Colorado	Social-ecological theory, social cognitive theory and the "5 As" self-management model	N=189, two intervention arms: CASM[k]: self-administered computer-assisted self-management; N=182, CASM+[l] with enhanced social support	All groups received general advice to decrease fat and eat a healthy diet	4-month intervention, 12-month follow-up

[a]RCT: randomized controlled trial.

[b]T2DM: type 2 diabetes mellitus.

[c]myDIDeA: Malaysian Dietary Intervention for People with Type 2 Diabetes: An e-Approach.

[d]IG: intervention group.

[e]CG: control group.

[f]ANODE: Accompagnement Nutritionnel de l'Obésité et du Diabète par Ecoaching.

[g]VLCKD: very low carbohydrate ketogenic diet.

[h]ADA: American Diabetes Association.

[i]N/A: not applicable.

[j]D-Net: Diabetes Network.

[k]CASM: computer-assisted self-management.

[l]CASM+: computer-assisted self-management plus social support.

Participants and Intervention Characteristics

The included studies provided results from a total of 1056 participants representing three countries—France [28], Malaysia [29], and the United States [30-32]. The number of participants ranged from ≤26 in the pilot study [30] to ≤463 in the RCTs [32]. The mean age of participants ranged from 50.5 years [29] to 59 years [32]. Three studies had an intervention period of 4 months [28,30,32], one was 3 months [31], and one was 6 months [29]. Two studies included a 12-month follow up [29,32], one had a 10-month follow up [31], and another had a 32-week follow up [30]. The Malaysian Dietary Intervention for People with Type 2 Diabetes: An e-Approach (myDIDeA) [29] focused solely on dietary behavior by providing participants with a personalized intensive dietary intervention. The remaining studies assessed dietary behavior but included dietary recommendations alongside physical activity [28,30], mindfulness training [30], social support and self-management information [31,32], and face-to-face contact [32]. Three studies included health practitioner-assistance [29,31,32], ranging from contact with a nutritionist [29,32] through to availability of health care coaches and professionals [31,32], and physicians [32]. The Accompagnement Nutritionnel de l'Obésité et du Diabète par Ecoaching (ANODE) study [28] was a fully

automated intervention aside from providing technical assistance. Three studies [29,31,32] utilized behavioral theories or models, which varied widely. Only myDIDeA [29] applied user-centered design theory to support intervention development.

Dietary Recommendations and Measurement

Three of the studies indicated the type of dietary recommendations prescribed [28-30], which were standard diabetes dietary guidelines modified to suit the Malaysian population [29], the National Nutrition and Health Program guidelines of France [28], and a very low carbohydrate ketogenic diet (VLCKD) [30]. The remaining two studies, computer-assisted self-management/computer-assisted self-management plus social support (CASM/+) [32] and the Diabetes Network (D-Net) [31], provided self-administered and/or personalized tailored dietary instruction from health professionals with general healthy eating recommendations and goal setting to reduce fast foods, fried foods, or sugar-sweetened beverages, and increase fruit and vegetable consumption. The tools and scales used for dietary assessment varied widely across all five studies (Table 2). myDIDeA [29] used a validated Malaysian 36-item Dietary Knowledge, Attitude, and Behavior (DKAB) score. ANODE [28] used the International Diet Quality Index (DQI-I) with food frequency questionnaires from both 24-hour and 3-day diet recalls. CASM/+ [32] used dietary questionnaires measured by two validated scales, the 20-item Kristal Fat and Fiber Behavior scale (FFB) and the 15-item Block/National Cancer Institute (NCI) Fat Screener. D-Net [31] also used the Block/NCI Fat Screener along with the 8-item “Starting the Conversation” scale. The VLCKD [30] used unvalidated self-reported dietary intake obtained from participants’ entries on a consumer-based website.

Table 2. An overview of dietary behavioral outcomes.

Reference	Dietary assessment	Baseline dietary assessment results	Postintervention dietary changes
Ramadas et al, 2018 [29]	DKAB[a] score, DSOC[b]	IG[c]: DKAB 34.2 (5.2), DSOC 193.3 (14.6) CG[d]: DKAB 33.7 (5.5), DSOC 191.2 (16.2)	IG: DKAB 54.0 (8.7), DSOC 199.7 (18.2) CG:DKAB 41.3 (7.7), DSOC 191.5 (15.1) DKAB score significantly improved in both groups with the margin of improvement higher in the IG, a difference of 12.2 points between groups. No difference between groups in DSOC at 6 months. IG showed improved DCOS at 12-month follow-up and no change in CG
Hansel et al, 2017 [28]	DQI-I[e]	IG: DQI-I 54.0 CG: DQI-I 52.8	IG: Significant increase in the DQI-I score of 4.55, total 58.55 CG: Decrease in DQI-I score of –1.68, total 51.12 Dietary changes towards healthier foods were noted in the IG, particularly for saturated fat (P=.02) and sodium (P<.001)
Saslow et al, 2017 [30]	Self-reported dietary intake (MyFitnessPal) and self-reported subjective experience of diets	IG: nonfiber carbohydrates (g) 163.6 (86.7), fat (g) 77.1 (41.4), protein (g) 83.3 (18.0), sugar (g) 50.6 (33.8) CG: nonfiber carbohydrates (g) 152.0 (58.9), fat (g) 81.3 (27.3), protein (g) 74.5 (17.2), sugar (g) 44.9 (23.8)	Self-reported dietary intake showed the IG ate fewer grams of nonfiber carbohydrates and sugar compared to CG. No differences in protein and fat between groups. Change in mean carbohydrate intake in IG from 39.6% of calories to 15.5%. Compared to CG, IG rated themselves as less likely to cheat on their diet, with a large effect size of at least Cohen d=–1.0
Glasgow et al, 2003 [31]	FFB[f] and the NCI[g] Fat Screener	Not reported	Trending improvements in FFB in both IGs but no significant differences between groups
Glasgow et al, 2012 [32]	"Starting the Conversation" scale and NCI Fat Screener	IG: eating habits 2.18 (0.2), fat intake 34.86 (28) CG: eating habits 2.13 (0.3), fat intake 35.18 (40)	IG: eating habits 2.32 (0.2), fat intake 33.22 (24) CG: eating habits 2.23 (0.3), fat intake 33.91 (37) The combined IG CASM[h]/CASM+[i] significantly improved eating habits more than CG over 12 months (chi-square = 9.01), fat intake (chi-square = 6.01)

[a]DKAB: Dietary Knowledge, Attitude, and Behavior.

[b]DSOC: Dietary Stages of Change.

[c]IG: intervention group.

[d]CG: control group.

[e]DQI-I: International Diet Quality Index.

[f]FFB: Kristal Fat and Fiber Behavior scale.

[g]NCI: National Cancer Institute.

[h]CASM: computer-assisted self-management.

[i]CASM+: computer-assisted self-management plus social support.

Dietary Behavior Change and Clinical Outcomes

Four of the studies reported a statistically significant dietary behavior change in the intervention group (Table 2) [28-30,32]. Compared to a control group receiving usual care, ANODE [28] found significant improvements in dietary quality with participants choosing healthier foods overall and improvements in saturated fat and sodium intake. Similarly, compared to an enhanced care control group, CASM/+ [32] found that participants' overall eating habits improved along with reductions in overall fat intake. Compared to a control group prescribed the ADA's "create-your-plate" diet, participants following a VLCKD [30] demonstrated adherence with decreased consumption of carbohydrates within the prescribed range of 20-50 grams per day and decreased consumption of added sugar. Compared to a control group prescribed usual care, myDIDeA [29] showed that web-based interventions could be a feasible option for supporting dietary behavior change for people with T2DM in developing countries such as Malaysia, with participants in the intervention group achieving a 12.2-point difference in DKAB score. CASM/+ [32] demonstrated dietary behavior change in individuals with lower literacy and numeracy and diverse and higher-risk populations such as American Indian/Alaska Native, Asian, Black and African American, and Latino.

Changes in clinical outcomes were inconsistent and differed across studies (Table 3). Two studies reported statistically significant improvements in glycemic control for HbA_{1c} between

groups compared to usual care [28] and the ADA's "create-your-plate" diet [30]. One study found significant between-group differences in fasting blood glucose and HbA_{1c} [29] compared to usual care, benefits that were only observed in the intervention group at 12-months follow-up. However, clinical outcomes were not reflected across all four studies that reported significant changes in dietary behavior. CASM/+ [32] reported no significant clinical improvements. Reductions in weight or waist circumference were seen in two studies [28,30]. Twenty percent of the ANODE study's [28] intervention participants achieved >5% weight loss, and 90% achieved at least 5% weight loss in the VLCKD intervention [30]. Weight reductions could be explained by improved overall diet quality and decreased fat intake [28], decreased sugar and carbohydrate consumption [30], greater calorie deficits in both intervention groups compared to control groups, along with recommendations for physical activity included in both interventions. Additionally, following a VLCKD yielded significant reductions in triglycerides [30].

Table 3. Significant clinical outcomes for dietary intervention groups. Data were considered statistically significant if $P<.05$.

Author/date/reported mean and outcomes measured	Baseline	Timepoint	Outcome	Within-group changes (*P* value)	Between-group changes (*P* value)
Ramadas et al, 2018 [29], mean (SD)					
HbA_{1c}[a] (%)	9.1 (2.0)	6 months = 8.7 (1.9)	12 months = 8.5 (1.8)	.004	—[b]
Fasting blood glucose (mmol/L)	8.9 (3.9)	6 months = 8.1 (2.7)	12 months = 7.9 (2.5)	.015	—
Hansel et al, 2017 [28], mean (SD)					
HbA_{1c} (%)	7.16 (0.78)	—	6.86 (0.94)	—	<.001
Weight (kg)	93.3 (16.2)	—	91 (3.0)	—	.01
Waist circumference (cm)	110 (10)	—	109.1 (4.7)	—	.01
Saslow et al, 2017 [30], mean (SD) and mean (EMM[c])					
HbA_{1c} (%)	7.1 (0.4)	16 weeks = 6.2 (–1.1, –0.6)	32 weeks = 6.3 (–1.1, –0.6)	—	.002
Weight (kg)	109 (24.9)	16 weeks = 100.5 (–11.9, -5.2)	32 weeks = 96.3 (–7.3, 1.3)	—	<.001
Triglycerides (mg/dL)	183 (135)	16 weeks = 147.5 (–65.7, –5.2 EMM)	32 weeks = 122.9 (–46.0, 33.6 EMM)	—	.01

[a]HbA_{1c}: hemoglobin A_{1c}

[b]Not applicable

[c]EMM: estimated marginal means.

Attrition, Website Features, and Usability

An overview of intervention components, attrition, usability, and modes and methods used to deliver the interventions is shown in Multimedia Appendix 1. The attrition rate varied among studies, with the highest dropout rate seen in the CASM/+ interventions [32], losing 34.2% of intervention participants randomized to two complex intervention arms, compared to 19.5% of the control group. The lowest attrition rate was seen in myDIDeA, losing 4.8% of intervention participants and 10.6% of the control group, perhaps because the intervention focused solely on providing a structured dietary intervention and implemented a user-centered approach. Across all web-based interventions, the highest website usage was reported in the first month, followed by declined usage over time. Three studies reported login rates as a measure for usability [28,31,32]. Only one study [29] reported both login rates and time spent on site. Various modes and methods of delivering intervention components were reported, including providing content on a website, which was used in four of the five studies [28,29,31,32], while one sent content via email with text, videos, and links to various web resources [30]. Four studies provided some form of self-monitoring and feedback, whether automated or assisted by a health care professional [28,29,31,32]. Three of the five interventions [29-31] provided updated intervention materials to participants ranging from biweekly to bimonthly. No associations could be drawn between these methods and participant adherence and intervention outcomes due to heterogeneity in the methods of delivery and a lack of detailed reporting.

Postintervention Process Evaluation

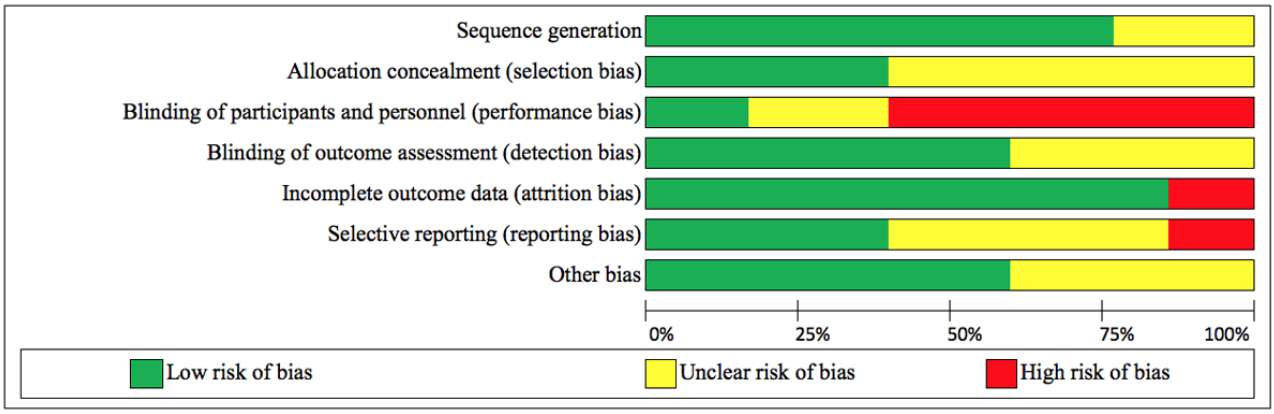
Postintervention process evaluation measured levels of adherence, usability, acceptability, and program satisfaction (Multimedia Appendix 2). The most common statistic used to measure adherence or intervention usage was website login rates [28,29,31,32], with studies consistently demonstrating higher login rates early in the intervention (Multimedia Appendix 1). Two studies included a more comprehensive process evaluation [28,29], both of which provided participants

with self-reported feedback questionnaires. myDIDeA [29] usability rates were 72%, acceptability 62%, and program satisfaction 64%. ANODE's [28] satisfaction rate was 70%. D-Net [31] indicated an implementation percentage of 100% related to their dietary assessment component but did not provide overall satisfaction rates.

Overall Study Quality

Figure 2 summarizes the risk-of-bias assessment [27]. Four studies had a reasonable sequence generation described [28,30-32]. It was unclear in three studies if allocation concealment was adequate [29,30,32]. Only one study sufficiently blinded participants and personnel [29], one study was unclear [31], while the remainder did not blind [28,30,32]. However, blinding in dietary behavior studies is often not feasible. Three studies sufficiently blinded outcome assessors [28-30], while two remained unclear [31,32]. Only one study failed to provide complete outcome data [30]. Two studies reported full datasets [28,31], two were unclear [30,32], while one [29] had several outcomes included in study protocols [36] that were missing from the final study outcomes. Other biases were low [28,31,32] or unclear [29,30]. Based on these assessments, only one study [28] was assessed to be of fair quality, while the remaining four were assessed to be of poor quality.

Figure 2. Risk of bias. Judgments about each risk of bias item presented as percentages across all included studies.



Discussion

This systematic review is the first to demonstrate the potential for web-based interventions to achieve significant improvements in dietary behavior in people with T2DM. Moreover, this review showed that improved glycemic control could be achieved using web-based dietary interventions. However, improvements in dietary behavior did not consistently result in improvements in glycemic control across studies.

One explanation for this inconsistency may be the varied dietary recommendations provided to participants across the studies (Table 1). While it is recognized that various dietary patterns are suitable for the management of T2DM [2], it appears that providing participants with more direct dietary recommendations may facilitate greater clinical outcomes. The ANODE study prescribed the National Nutrition and Health Program guidelines of France, myDIDeA prescribed evidence-based standard diabetes dietary guidelines modified to suit the Malaysian population, and the VLCKD prescribed specific guidelines for carbohydrate consumption. In contrast, CASM/+ only provided participants with general healthy eating information. Another explanation may be intervention complexity, as web-based dietary interventions in this review varied widely in terms of the number of components provided to participants (Multimedia Appendix 1). Previous literature [6,38] indicated overly complex interventions might provoke a lack of motivation due to confusion, provision of irrelevant content, and technical difficulties.

Regardless of the setting, patient engagement and adherence to their recommended care plan is an important issue [39]. Our review confirms previous observations [22,25] that usage of web and computer-based interventions decreases over time, with the highest usage seen in the first month. This pattern is particularly concerning as the studies in this review had intervention periods of only 3-6 months, highlighting the challenge of engaging participants in a web-based environment even over the short term. The majority of studies in this review reported only login rates as a measure for usability [28,31,32]. Login rates capture only a broad measure of website usage, providing little information about engagement with an intervention [15]. Only one study [29] provided evidence of engagement, reporting both login rates and time spent on the site, with a consistent level of engagement observed across the 6-month intervention for all participants. This consistency may have been due to the provision of structured dietary education modules. In face-to-face settings, structured education has been demonstrated to be effective for assisting people with T2DM to improve glycemic control and overall health [2]. Various authors suggest social support may increase participation [10,21]; however, this was not demonstrated in our results, as even the more complex web-based interventions offering contact with peers [31,32], and/or health care professionals [29,31,32] contributed no difference in terms of achieving clinical outcomes.

myDIDeA [29] was the first study to focus solely on providing a web-based dietary behavior change intervention for people with T2DM in Malaysia, providing the longest intervention period of 6-months with a 12-month follow-up. The intervention group showed a 12.2-point greater DKAB score compared to the control group (Table 2). While no between-group changes

were found for clinical outcomes (Table 3), within-group changes in fasting blood glucose and HbA_{1c} were only found in the intervention group at 12-months follow-up, which may indicate that web-based dietary interventions could continue to influence behavior change beyond the intervention period. Furthermore, myDIDeA had the lowest overall attrition rate, which may be explained by the provision of an intervention solely focused on diet. Since determining what to eat is a significant challenge for people with T2DM [2], this specific dietary focus and the structured nature of delivering the components may have encouraged clarity, fostering longer-term commitment. Additionally, myDIDeA integrated a user-centered approach during the development of the intervention. User-centered design is a human-factor engineering strategy for designing user-friendly platforms [40], a method that has been described in other web-based T2DM self-management interventions [41,42].

The Medical Research Council [39] suggests that best practice for complex interventions is using the best available evidence and appropriate theory. According to a systematic review and meta-analysis [43] of diet behavior change techniques in people with T2DM, the only intervention feature associated with significant reductions in HbA_{1c} was the application of a theoretical model or framework. However, the relationship between theoretical application and clinical outcomes has yet to be confirmed in web-based dietary behavior change interventions. For instance, the ANODE study [28] demonstrated statistically significant improvements in both dietary behavior and clinical outcomes, yet there was no theoretical basis for the development of the program. Rather it was a nutritional support tool developed by a private company.

Changes in glycemic control ranged from HbA_{1c} reductions of 0.3% [28] to 0.8% [30]. These results are clinically meaningful as previous research has demonstrated that for every 1% reduction in HbA_{1c}, there is an associated risk reduction for heart attacks, microvascular complications, and deaths related to diabetes [3]. Achieving modest weight loss of ≥5% has been shown to improve glycemic control [2], a goal that was achieved by participants in the two studies that found statistically significant clinical outcomes between groups [28,30]. The VLCKD [30] produced the most significant results overall, with reductions in HbA_{1c}, weight, and triglycerides. However, these results should be interpreted with caution due to the small study size. Furthermore, systematic reviews and meta-analysis of RCTs prescribing low-carbohydrate diets in people with T2DM [44], suggest that adherence to VLCKD interventions (<50 grams carbohydrates per day) are frequently poor and more difficult for people to maintain than a low-carbohydrate diet (50-130 grams carbohydrates per day), with no additional clinical benefits over a prescribed low-carbohydrate diet [45]. Overall, the results of this review demonstrated that, regardless of diet characteristics, participants who adhered to dietary recommendations showed improvements in their food choices and overall dietary quality, which improved clinical outcomes in most cases.

Limitations and Future Research

Only five studies met the eligibility criteria, and these were heterogeneous in terms of dietary recommendations, focus on diet alone or with additional behavioral components, behavioral theories and models applied, and target population. The modes and methods of delivering the web-based interventions also differed, and the duration and follow up periods varied widely. While most studies report statistically significant improvements in dietary behavior change, the use of different dietary assessment measures made it difficult to compare study outcomes and generalize results. Web-based interventions could improve and expand reporting of website statistics to help inform patterns of participant behavior, and consideration of measuring adherence to diet as a means of determining if greater adherence leads to greater improvement in clinical outcomes is warranted. Web-based interventions could be a cost-effective way to provide education and support to individuals with T2DM, broadening access for a greater number of people, including those who have location or mobility constraints and cannot access face-to-face services. More research is needed to explore web-based dietary interventions for these diverse populations, including younger adults and the elderly. Web-based dietary interventions must be studied in larger cohorts, for longer durations, and with more clearly defined dietary recommendations. Studies must also explore intervention content, modes and methods of delivery, and whether these enhance participant engagement or contribute meaningfully to the expected outcomes.

Conclusion

This review provided evidence that web-based interventions may be an effective way to improve dietary behavior in people with T2DM. The results also suggest improvements in glycemic control and clinical outcomes may be possible, although the studies in this review yielded inconsistent results. While this preliminary evidence showed promise of a positive effect, the small number of studies and the fact they are highly heterogeneous makes it difficult to draw any firm conclusions. The field requires more well-designed web-based dietary interventions that report dietary prescription and adherence in people with T2DM to confirm their effectiveness in optimizing dietary behavior and improving clinical outcomes.

Acknowledgments

SI designed and registered this review and supervised study procedures. JD conducted data extraction and analyses under the supervision of SI. JD wrote the manuscript with critical input from SI, EG, and RM. We would like to thank Reza Daryabeygi-Khotbehsara for research assistance with the risk-of-bias assessment.

Conflicts of Interest

JD is a co-owner of Diabetes Meal Plans (DMP), a web-based low carbohydrate nutrition support service for people with T2DM. No study reviewed in this manuscript was associated with DMP.

Multimedia Appendix 1
Table 4. Intervention components, attrition, usability and modes and methods of delivery.
[PDF File (Adobe PDF File), 88 KB - jmir_v22i8e16437_app1.pdf]

Multimedia Appendix 2
Table 5. Process evaluation measures.
[PDF File (Adobe PDF File), 66 KB - jmir_v22i8e16437_app2.pdf]

References

1. Kaiser AB, Zhang N, Der Pluijm WV. Global Prevalence of Type 2 Diabetes over the Next Ten Years (2018-2028). Diabetes 2018 May;67(Supplement 1):202-LB. [doi: 10.2337/db18-202-lb]
2. American Diabetes Association. Standards of Medical Care in Diabetes 2019. Diabetes Care (Supplement 1). 2019. URL: https://care.diabetesjournals.org/content/diacare/suppl/2018/12/17/42.Supplement_1.DC1/DC_42_S1_2019_UPDATED.pdf [accessed 2019-07-15]
3. Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ 2000 Aug 12;321(7258):405-412 [FREE Full text] [doi: 10.1136/bmj.321.7258.405] [Medline: 10938048]
4. Evert AB, Dennison M, Gardner CD, Garvey WT, Lau KHK, MacLeod J, et al. Nutrition Therapy for Adults With Diabetes or Prediabetes: A Consensus Report. Diabetes Care 2019 May 18;42(5):731-754 [FREE Full text] [doi: 10.2337/dci19-0014] [Medline: 31000505]
5. Lopez JMS, Katic BJ, Fitz-Randolph M, Jackson RA, Chow W, Mullins CD. Understanding preferences for type 2 diabetes mellitus self-management support through a patient-centered approach: a 2-phase mixed-methods study. BMC Endocr Disord 2016 Jul 18;16(1):41 [FREE Full text] [doi: 10.1186/s12902-016-0122-x] [Medline: 27430259]
6. Pal K, Dack C, Ross J, Michie S, May C, Stevenson F, et al. Digital Health Interventions for Adults With Type 2 Diabetes: Qualitative Study of Patient Perspectives on Diabetes Self-Management Education and Support. J Med Internet Res 2018 Jan 29;20(2):e40 [FREE Full text] [doi: 10.2196/jmir.8439] [Medline: 29463488]
7. Cassimatis M, Kavanagh DJ, Smith AC. Perceived Needs for Supported Self-management of Type 2 Diabetes: A Qualitative Investigation of the Potential for a Web-based Intervention. Australian Psychologist 2014 Mar 10;49(2):75-85. [doi: 10.1111/ap.12050]
8. Booth AO, Lowis C, Dean M, Hunter SJ, McKinley MC. Diet and physical activity in the self-management of type 2 diabetes: barriers and facilitators identified by patients and health professionals. Prim Health Care Res Dev 2013 Jul;14(3):293-306. [doi: 10.1017/S1463423612000412] [Medline: 23739524]
9. Kennedy M, Dunning T. Diabetes education: essential but underfunded in Australia. Diabetes & Primary Care Australia 2017;2(1):e [FREE Full text]
10. Pereira K, Phillips B, Johnson C, Vorderstrasse A. Internet delivered diabetes self-management education: a review. Diabetes Technol Ther 2015 Jan;17(1):55-63. [doi: 10.1089/dia.2014.0155] [Medline: 25238257]
11. Managing diabetes: improving services for people with diabetes. Commission for Healthcare Audit Inspection, London, UK. URL: https://www.yearofcare.co.uk/sites/default/files/pdfs/Managing_diabetes.pdf [accessed 2019-08-19]
12. Australian Diabetes Educators Association. Workforce in Diabetes Education (Online). URL: https://members.adea.com.au/resources-2/workforce-in-diabetes-education/ [accessed 2020-08-19]
13. Chatterjee S, Davies MJ, Heller S, Speight J, Snoek FJ, Khunti K. Diabetes structured self-management education programmes: a narrative review and current innovations. Lancet Diabetes Endocrinol 2018 Dec;6(2):130-142. [doi: 10.1016/S2213-8587(17)30239-5] [Medline: 28970034]
14. Odgers-Jewell K, Isenring EA, Thomas R, Reidlinger DP. Group-based education for patients with type 2 diabetes: a survey of Australian dietitians. Aust J Prim Health 2017 Sep;23(4):364-372. [doi: 10.1071/PY16156] [Medline: 28566113]
15. Karekla M, Kasinopoulos O, Neto DD, Ebert DD, Van Daele T, Nordgreen T, et al. Best Practices and Recommendations for Digital Interventions to Improve Engagement and Adherence in Chronic Illness Sufferers. European Psychologist 2019 Jan;24(1):49-67. [doi: 10.1027/1016-9040/a000349]
16. Li J, Parrott S, Sweeting M, Farmer A, Ross J, Dack C, et al. Cost-Effectiveness of Facilitated Access to a Self-Management Website, Compared to Usual Care, for Patients With Type 2 Diabetes (HeLP-Diabetes): Randomized Controlled Trial. J Med Internet Res 2018 Jun 08;20(6):e201 [FREE Full text] [doi: 10.2196/jmir.9256] [Medline: 29884608]
17. Ramadas A, Quek KF, Chan CKY, Oldenburg B. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. Int J Med Inform 2011 Jun;80(6):389-405. [doi: 10.1016/j.ijmedinf.2011.02.002] [Medline: 21481632]

18. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: An overview of systematic reviews. PLoS One 2017;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
19. Fu H, McMahon SK, Gross CR, Adam TJ, Wyman JF. Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: A systematic review. Diabetes Res Clin Pract 2017 Sep;131:70-81. [doi: 10.1016/j.diabres.2017.06.016] [Medline: 28692830]
20. Nangrani N, Malabu U, Vangaveti V. Outcomes of Telehealth in the Management of Type 2 Diabetes—A Systematic Review and Meta-analysis of Randomised Controlled Trials. Diabetes 2018 May;67(Supplement 1):1298-P. [doi: 10.2337/db18-1298-p]
21. Kebede MM, Zeeb H, Peters M, Heise TL, Pischke CR. Effectiveness of Digital Interventions for Improving Glycemic Control in Persons with Poorly Controlled Type 2 Diabetes: A Systematic Review, Meta-analysis, and Meta-regression Analysis. Diabetes Technol Ther 2018 Nov;20(11):767-782. [doi: 10.1089/dia.2018.0216] [Medline: 30257102]
22. Cotterez AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
23. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
24. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
25. Pal K, Eastwood SV, Michie S, Farmer A, Barnard ML, Peacock R, et al. Computer-based interventions to improve self-management in adults with type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2014 Jun;37(6):1759-1766. [doi: 10.2337/dc13-1386] [Medline: 24855158]
26. Pal K, Eastwood SV, Michie S, Farmer AJ, Barnard ML, Peacock R, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev 2013;3:CD008776. [doi: 10.1002/14651858.CD008776.pub2] [Medline: 23543567]
27. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
28. Hansel B, Giral P, Gambotti L, Lafourcade A, Peres G, Filipecki C, et al. A Fully Automated Web-Based Program Improves Lifestyle Habits and HbA1c in Patients With Type 2 Diabetes and Abdominal Obesity: Randomized Trial of Patient E-Coaching Nutritional Support (The ANODE Study). J Med Internet Res 2017 Nov 08;19(11):e360 [FREE Full text] [doi: 10.2196/jmir.7947] [Medline: 29117929]
29. Ramadas A, Chan CKY, Oldenburg B, Hussein Z, Quek KF. Randomised-controlled trial of a web-based dietary intervention for patients with type 2 diabetes: changes in health cognitions and glycemic control. BMC Public Health 2018 Dec 08;18(1):716 [FREE Full text] [doi: 10.1186/s12889-018-5640-1] [Medline: 29884161]
30. Saslow LR, Mason AE, Kim S, Goldman V, Ploutz-Snyder R, Bayandorian H, et al. An Online Intervention Comparing a Very Low-Carbohydrate Ketogenic Diet and Lifestyle Recommendations Versus a Plate Method Diet in Overweight Individuals With Type 2 Diabetes: A Randomized Controlled Trial. J Med Internet Res 2017 Feb 13;19(2):e36 [FREE Full text] [doi: 10.2196/jmir.5806] [Medline: 28193599]
31. Glasgow RE, Boles SM, McKay HG, Feil EG, Barrera M. The D-Net diabetes self-management program: long-term implementation, outcomes, and generalization results. Prev Med 2003 Apr;36(4):410-419. [doi: 10.1016/s0091-7435(02)00056-7] [Medline: 12649049]
32. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Twelve-month outcomes of an Internet-based diabetes self-management support program. Patient Educ Couns 2012 Apr;87(1):81-92 [FREE Full text] [doi: 10.1016/j.pec.2011.07.024] [Medline: 21924576]
33. Ramadas A, Chan CKY, Oldenburg B, Hussien Z, Quek KF. A web-based dietary intervention for people with type 2 diabetes: development, implementation, and evaluation. Int J Behav Med 2015 Jun;22(3):365-373 [FREE Full text] [doi: 10.1007/s12529-014-9445-z] [Medline: 25274015]
34. Glasgow RE, Strycker LA, Kurz D, Faber A, Bell H, Dickman JM, et al. Recruitment for an internet-based diabetes self-management program: scientific and ethical implications. Ann Behav Med 2010 Aug;40(1):40-48. [doi: 10.1007/s12160-010-9189-1] [Medline: 20411443]
35. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]
36. Ramadas A, Quek KF, Chan CKY, Oldenburg B, Hussein Z. Randomised-controlled trial of a web-based dietary intervention for patients with type 2 diabetes mellitus: study protocol of myDIDeA. BMC Public Health 2011 May 21;11:359 [FREE Full text] [doi: 10.1186/1471-2458-11-359] [Medline: 21599990]

37. McKay HG, Glasgow RE, Feil EG, Boles SM, Barrera MJ. Internet-based diabetes self-management and support: Initial outcomes from the Diabetes Network project. Rehabilitation Psychology 2002;47(1):31-48. [doi: 10.1037/0090-5550.47.1.31]
38. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout From an eHealth Intervention for Adults With Type 2 Diabetes: A Qualitative Study. J Med Internet Res 2017 May 30;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]
39. Medical Research Council. Developing and evaluating complex interventions: Following considerable development in the field since 2006, MRC and NIHR have jointly commissioned an update of this guidance to be published in 2019 (Online). URL: https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/ [accessed 2020-08-29]
40. LeRouge C, Wickramasinghe N. A review of user-centered design for diabetes-related consumer health informatics technologies. J Diabetes Sci Technol 2013 Jul 01;7(4):1039-1056 [FREE Full text] [doi: 10.1177/193229681300700429] [Medline: 23911188]
41. Murray E, Sweeting M, Dack C, Pal K, Modrow K, Hudda M, et al. Web-based self-management support for people with type 2 diabetes (HeLP-Diabetes): randomised controlled trial in English primary care. BMJ Open 2017 Sep 27;7(9):e016009 [FREE Full text] [doi: 10.1136/bmjopen-2017-016009] [Medline: 28954789]
42. Yu CH, Parsons JA, Hall S, Newton D, Jovicic A, Lottridge D, et al. User-centered design of a web-based self-management site for individuals with type 2 diabetes - providing a sense of control and community. BMC Med Inform Decis Mak 2014;14:60 [FREE Full text] [doi: 10.1186/1472-6947-14-60] [Medline: 25056379]
43. Cradock KA, ÓLaighin G, Finucane FM, McKay R, Quinlan LR, Martin GKA, et al. Diet Behavior Change Techniques in Type 2 Diabetes: A Systematic Review and Meta-analysis. Diabetes Care 2017 Dec;40(12):1800-1810. [doi: 10.2337/dc17-0462] [Medline: 29162585]
44. Huntriss R, Campbell M, Bedwell C. The interpretation and effect of a low-carbohydrate diet in the management of type 2 diabetes: a systematic review and meta-analysis of randomised controlled trials. Eur J Clin Nutr 2018 Mar;72(3):311-325. [doi: 10.1038/s41430-017-0019-4] [Medline: 29269890]
45. McArdle PD, Greenfield SM, Rilstone SK, Narendran P, Haque MS, Gill PS. Carbohydrate restriction for glycaemic control in Type 2 diabetes: a systematic review and meta-analysis. Diabet Med 2019 Mar;36(3):335-348. [doi: 10.1111/dme.13862] [Medline: 30426553]

Abbreviations

ADA: American Diabetes Association
ANODE: Accompagnement Nutritionnel de l'Obésité et du Diabète par Ecoaching
CASM/+: computer-assisted self-management/computer-assisted self-management plus social support
CINAHL: Cumulative Index of Nursing and Allied Health Literature
CVD: cardiovascular disease
DKAB: Dietary Knowledge, Attitude and Behavior
D-Net: Diabetes Network
HbA_{1c}: hemoglobin A_{1c}
mHealth: mobile health
myDIDeA: Malaysian Dietary Intervention for People with Type 2 Diabetes: An e-Approach
NCI: National Cancer Institute
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
RoB 2: Cochrane risk-of-bias tool for randomized trials
T2DM: type 2 diabetes mellitus
VLCKD: very low carbohydrate ketogenic diet

Edited by G Eysenbach; submitted 29.09.19; peer-reviewed by K Johnson, R Meseri; comments to author 19.05.20; revised version received 30.05.20; accepted 03.06.20; published 28.08.20.

Please cite as:
Dening J, Islam SMS, George E, Maddison R
Web-Based Interventions for Dietary Behavior in Adults With Type 2 Diabetes: Systematic Review of Randomized Controlled Trials
J Med Internet Res 2020;22(8):e16437
URL: http://www.jmir.org/2020/8/e16437/
doi:10.2196/16437
PMID:32857059

Review

Effectiveness of Mobile Health Interventions Promoting Physical Activity and Lifestyle Interventions to Reduce Cardiovascular Risk Among Individuals With Metabolic Syndrome: Systematic Review and Meta-Analysis

Irene Sequi-Dominguez[1], MSc; Celia Alvarez-Bueno[1,2], PhD; Vicente Martinez-Vizcaino[1,3], MD; Rubén Fernandez-Rodriguez[1,4], MSc; Alicia del Saz Lara[1], MSc; Iván Cavero-Redondo[1,2], PhD

[1]Universidad de Castilla-La Mancha, Health and Social Research Center, Cuenca, Spain

[2]Universidad Politécnica y Artística del Paraguay, Asunción, Paraguay

[3]Universidad Autónoma de Chile, Facultad de Ciencias de la Salud, Talca, Chile

[4]Movi-Fitness SL, Universidad de Castilla-La Mancha, Cuenca, Spain

Corresponding Author:
Celia Alvarez-Bueno, PhD
Universidad de Castilla-La Mancha
Health and Social Research Center
Edificio Melchor Cano Campus Universitario de Cuenca
Camino de Pozuelo, s/n
Cuenca, 16071
Spain
Phone: 34 969 179 100 ext 4659
Email: celia.alvarezbueno@uclm.es

Abstract

Background: Physical activity and lifestyle interventions, such as a healthy diet, have been proven to be effective approaches to manage metabolic syndrome. However, these interventions require great commitment from patients and clinicians owing to their economic costs, time consumption, and lack of immediate results.

Objective: The aim of this systematic review and meta-analysis was to analyze the effect of mobile-based health interventions for reducing cardiometabolic risk through the promotion of physical activity and healthy lifestyle behaviors.

Methods: PubMed, Scopus, Web of Science, Cochrane Central Register of Controlled Trials, and SPORTdiscus databases were searched for experimental studies evaluating cardiometabolic risk indicators among individuals with metabolic syndrome who were included in technology-assisted physical activity and lifestyle interventions. Effect sizes, pooled mean changes, and their respective 95% CIs were calculated using the DerSimonian and Laird method. Outcomes included the following clinical and biochemical parameters: body composition (waist circumference [WC] and BMI), blood pressure (systolic blood pressure [SBP] and diastolic blood pressure [DBP]), glucose tolerance (fasting plasma glucose [FPG] and glycated hemoglobin A1c [HbA_{1c}]), and lipid profile (total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol [HDL-C], and triglycerides).

Results: A total of nine studies were included in the meta-analysis. Owing to the scarcity of studies, only pooled mean pre-post changes in the intervention groups were estimated. Significant mean changes were observed for BMI (−1.70 kg/m2, 95% CI −3.20 to −0.20; effect size: −0.46; *P*=.03), WC (−5.77 cm, 95% CI −9.76 to −1.77; effect size: −0.54; *P*=.005), SBP (−7.33 mmHg, 95% CI −13.25 to −1.42; effect size: −0.43; *P*=.02), DBP (−3.90 mmHg, 95% CI −7.70 to −0.11; effect size: −0.44; *P*=.04), FPG (−3.65 mg/dL, 95% CI −4.79 to −2.51; effect size: −0.39; *P*<.001), and HDL-C (4.19 mg/dL, 95% CI 2.43-5.95; effect size: 0.23; *P*<.001).

Conclusions: Overall, mobile-based health interventions aimed at promoting physical activity and healthy lifestyle changes had a strong positive effect on cardiometabolic risk indicators among individuals with metabolic syndrome. Nevertheless, further research is required to compare this approach with usual care in order to support the incorporation of these technologies in health systems.

Trial Registration: PROSPERO CRD42019125461; https://tinyurl.com/y3t4wog4.

(J Med Internet Res 2020;22(8):e17790) doi:10.2196/17790

KEYWORDS

mobile health; mobile technology; telemedicine; metabolic syndrome; physical activity; lifestyle intervention; systematic review; meta-analysis

Introduction

Metabolic syndrome (MetS) is a cluster of cardiometabolic risk factors that include abdominal obesity, dyslipidemia, hypertension, and insulin resistance [1,2]. MetS has become a worldwide epidemic in parallel with the increase in unhealthy behaviors, such as high rates of physical inactivity and energy dense diets, which have led to alarming obesity prevalence rates in wealthy countries, as well as in developing countries, but to a lesser extent [3]. MetS increases the risk of diabetes mellitus and cardiovascular disease (CVD) in patients with or without a history of cardiovascular events [4]; thus, its early detection may be an important strategy to improve patients' future cardiometabolic risk.

Traditionally, MetS has not been clinically addressed as a single entity but has been managed by treating each of its individual components separately by recommending lifestyle changes (healthy diet and exercise) and pharmacological or even surgical approaches (specifically bariatric surgery, when required). Physical activity interventions have been proven to be effective in reducing CVD risk factors by increasing cardiorespiratory fitness, and dietary interventions have been proven to be effective in decreasing adiposity [5]. In addition, physical activity interventions have been shown to be effective at 12 weeks or more for cardiometabolic parameters [6]. So far, randomized controlled trials (RCTs) of these interventions have required intensive one-on-one or group lifestyle recommendations, raising questions about the feasibility and scalability of implementing these interventions outside of research settings [7].

Mobile-based health (mHealth) technologies can be conceptualized as the remote delivery of health care and exchange of health care information [8]. These technologies can be seen as a complement for some traditional health care methods that, by enabling remote health consultations and monitoring, improve accessibility to health services and the efficiency of some health interventions [8]. Since mobile apps play a key role in everyday life, lifestyle interventions based on these technologies may increase the potential for scalability of interventions and improve their long-term effects and sustainability. In fact, it is expected that the prevention and management of the most common health disorders, which traditionally place a large burden on personnel and resources, will gradually shift to a disease management model in the near future, introducing the use of mHealth [9].

Thus, the aim of this systematic review and meta-analysis was to analyze the effect of lifestyle interventions, including physical activity recommendations through mHealth technologies, on CVD risk factors among individuals with MetS.

Methods

Design

This systematic review and meta-analysis was registered in PROSPERO (registration number: CRD42019125461) and was reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [10]. The recommendations of the Cochrane Handbook for Systematic Reviews of Interventions [11] were followed to conduct this systematic review and meta-analysis.

Search Strategy

PubMed (via Medline), EMBASE (via Scopus), Web of Science, Cochrane Central Register of Controlled Trials, and SPORTdiscus databases were searched from their inception to August 2019 following the same PICO (population, intervention, comparison, and outcome) strategy (Figure 1) that included the following: ((“metabolic syndrome”) AND (“physical activity” OR “lifestyle intervention” OR “health coaching” OR “technology assisted” OR “mobile technology” OR “health technology” OR “internet based” OR “mobile health” OR “mobile phone-based”) AND (effectiveness OR utility OR effect OR “cardiometabolic risk factors” OR “cardio-metabolic markers” OR weight OR “body mass index” OR “waist circumference” OR “blood pressure” OR “hemoglobin A1c” OR “fasting plasma glucose” OR “total cholesterol” OR HDL-C OR LDL-C OR triglyceride)).

Figure 1. PICO (population, intervention, comparison, and outcome) search strategy.

	Keywords
Population	"metabolic syndrome"
Intervention	"physical activity", "lifestyle intervention", "health coaching", "technology assisted", "mobile technology", "health technology", "internet based", "mobile health", "mobile phone-based"
Comparison	
Outcome	effectiveness, utility, effect, "cardiometabolic risk factors", "cardiometabolic markers", weight, "body mass index", "waist circumference", "blood pressure", "haemoglobin A1c", "fasting plasma glucose", "total cholesterol", HDL-C, LDL-C, triglyceride

Selection of Studies

Eligible articles were experimental studies (RCTs or nonrandomized experimental studies and single-arm pre-post studies), which aimed to measure the effectiveness of lifestyle and physical activity recommendations, using mHealth technologies to reduce cardiometabolic risk factors in individuals with MetS. Studies not written in English or Spanish, including patients with diabetes, or not reporting pre- and postcardiometabolic risk factor values were excluded.

Interventions were classified according to their main characteristics as follows: (1) performing data monitoring or not; (2) carrying out lifestyle and/or physical activity recommendations; and (3) including goal setting tools or not. Outcomes were measured as mean changes in the following cardiometabolic risk indicators: body composition (BMI and waist circumference [WC]), blood pressure (systolic blood pressure [SBP] and diastolic blood pressure [DBP]), glucose tolerance (fasting plasma glucose [FPG] and glycated hemoglobin A1c [HbA_{1c}]), and lipid profile (total cholesterol, high-density lipoprotein cholesterol [HDL-C], low-density lipoprotein cholesterol [LDL-C], and triglycerides).

The literature search was independently conducted by two reviewers (ISD and ICR), and disagreements were solved by consensus or discussion with a third researcher (CAB).

Data Extraction and Quality Assessment

The following information was extracted from the included studies: (1) year of publication, (2) country, (3) type of study, (4) sample characteristics (sample size and mean age), (5) intervention characteristics (design and length of intervention), and (6) MetS indicators.

The Cochrane Collaborations tool was used for assessing risk of bias in randomized trials [12], which scores six domains as low, high, or unclear risk. The Quality Assessment Tool for Quantitative Studies [13] was used for nonrandomized experimental and single-arm pre-post studies. It consists of seven domains of risk of bias that are rated as strong, moderate, or weak. Both tools assessed the risk of bias of each study as low (with no high/weak ratings), moderate (with one high/weak rating), or high (with two or more high/weak ratings) [14].

Data extraction and quality assessment were independently performed by two reviewers (ISD and ICR), and inconsistencies were solved by consensus or discussion with a third researcher (CAB). The agreement rate between reviewers was calculated using the kappa statistic.

Statistical Analysis

The DerSimonian and Laird method [15] was used to compute the pooled mean change estimates for BMI, WC, SBP, DBP, FPG, HbA_{1c}, total cholesterol, HDL-C, LDL-C, and triglycerides, with their respective 95% CIs. Because of the scarcity of RCTs, in which the difference in change between intervention and control groups for the outcome variable was calculated, we calculated the pooled mean pre-post change in the outcome variable for all the interventions (not for the control group). In multiarm trials (two or more intervention groups), we calculated separately the pooled mean pre-post change in each arm, and the common control group was not included in the analysis. Additionally, standardized mean difference scores for the pooled mean change estimates were calculated using the effect size of Cohen *d*, in which the effect was considered weak for values around 0.2, moderate for values around 0.5, strong for values around 0.8, and very strong for values greater than 1.0. When studies reported pre- and postmean values, effect size estimates were calculated for each parameter.

The heterogeneity of results across studies was evaluated using the I^2 statistic [16]. I^2 values were assessed as follows: 0%-30%, might not be important; 30%-50%, moderate heterogeneity; 50%-75%, substantial heterogeneity; and 75%-100%, considerable heterogeneity. The corresponding *P* values were also taken into account [11].

Sensitivity analyses were conducted to assess the robustness of the summary estimates and to detect if any particular study accounted for a large proportion of heterogeneity. Random-effects meta-regression models were used to evaluate whether pooled estimates were influenced by the mean age of participants and the percentage of women [17]. Finally, publication bias was evaluated through visual inspection of funnel plots, as well as using the method proposed by Egger [18].

The significance value of the pooled mean change was estimated based on the 95% CI. Statistical analyses were performed using STATA SE software, version 15 (StataCorp).

Results

Systematic Review

After removing duplicate studies, a total of 47 articles were selected for full-text review following title and abstract screening. Finally, nine studies [19-27] were included in this systematic review (Figure 2).

Of the included studies, five were RCTs [19,23-26] and four were single-arm pre-post studies [20-22,27]. Studies were published between 2013 and 2018, and conducted in four different countries (two in Canada [26,27], one in Germany [23], three in the Republic of Korea [21,22,25], and three in the United States [19,20,24]).

The sample size of the included studies ranged from 12 to 421 participants (51.7% females, although two studies included men only [21,22]), and the mean age varied between 38.4 and 59.7 years. All participants met the diagnostic criteria for MetS (according to the Adult Treatment Panel III guidelines or the International Diabetes Federation) and were able to access and use the technology required for each intervention.

The interventions were mainly based on physical activity and lifestyle recommendations, with personalization in some cases [20-23], and were delivered through a website, videoconferencing, or an app. The effects of the recommendations were assessed using telemonitoring through mobile devices. In three of the included studies, the interventions were strengthened using self-goal setting tools such as a behavioral strategy for patients to help them visualize their accomplishments and objectives [24,26,27]. The duration of interventions ranged from 8 to 48 weeks, with the number of clinical encounters varying between 2 and 24, and most of them were in-person encounters to perform periodic clinical evaluations (Table 1).

Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram.

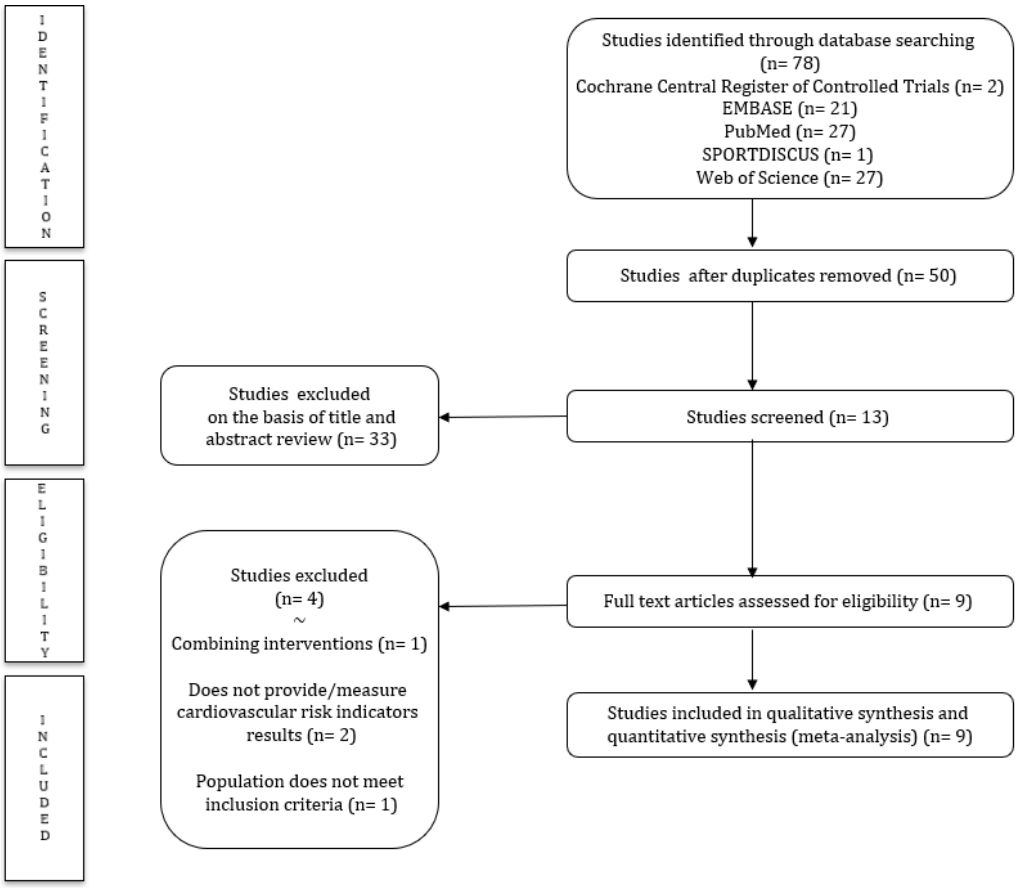


Table 1. Characteristics of the included studies.

First author, year of publication	Country	Study design	Intervention characteristics					Cardiometabolic risk outcomes
			Sample size (n [%] female)	Mean age (years)	Intervention design	Number of clinical encounters	Duration	
Azar et al, 2016 [19]	USA	RCT[a]	n=74 (44 [59%]) IG[b]: n=37 CG[c]: n=37	59.7 (SD 11.2)	IG: Data monitoring, PA[d] and lifestyle web advice, and weekly videoconferencing CG: Intervention 3 months delayed	24 virtual group sessions Seven in-person PA sessions	24 weeks	Weight, weight change, BMI, WC[e], SBP[f], DBP[g], TC[h], HDL-C[i], LDL-C[j], TC/HDL ratio, and triglyceride
Everett et al, 2018 [20]	USA	Pre-post study	n=38 (24 [63%])	57.2 (SD 9.1)	Data monitoring and PA, weight reduction, and diet personalized advice through a smartphone	Two face-to-face sessions	12 weeks	Weight, percentage weight change, BMI, WC, SBP, DBP, HbA_{1c}[k], and FPG[l]
Kim and Kang, 2013 [21]	Republic of Korea	Pre-post study	n=18 (0 [0%])	43.1 (SD 7.4)	PA and weight control personalized advice through a website and SMS text messages	Weekly web visits	8 weeks	Weight, visceral fat mass, WC, SBP, DBP, HDL-C, TG, FPG, and CVD[m] risk
Kim et al, 2014 [22]	Republic of Korea	Pre-post study	n=48 (0 [0%]) IG: n=24 CG: n=24	IG: 40.88 (SD 7.70) CG: 38.38 (SD 6.82)	PA and weight control personalized advice through a website and SMS text messages	Weekly online sessions	16 weeks	Weight, body fat, VFM[n], WC, SBP, DBP, HDL-C, TG, FPG, and CVD risk
Luley et al, 2014 [23]	Germany	RCT	n=178 (73 [41%]) IG1: n=60 (18 [30%]) IG2: n=58 (27 [47%]) CG: n=60 (28 [47%])	IG1: 50.3 (SD 7.8) IG2: 50.3 (SD 8.0) CG: 50.1 (SD 8.1)	IG1: PA and diet recommendations, data telemonitoring, and weekly feedback letters IG2: PA and diet recommendations, data telemonitoring, and monthly feedback calls CG: PA and diet in-person recommendations	Four in-person sessions	48 weeks	Weight loss; BMI, WC, SBP, DBP, TC, HDL-C, LDL-C, TG, apolipoprotein B, uric acid, alanine aminotransferase, aspartate aminotransferase, high-sensitivity CRP[o], FPG, HbA_{1c}, and HOMA-IR[p]
Mann et al, 2016 [24]	USA	RCT	n=54 (45 [83%]) IG: n=27 CG: n=27	IG: 47.5 (SD 11.99) CG: 43.67 (SD 9.28)	IG: Data monitoring, PA and diet recommendations, and goal setting using electronic medical records CG: Traditional recommendations and follow-up	Two compulsory in-person sessions	24 weeks	Weight, BMI, TC, HCL-C, LDL-C, TG, HbA_{1c}, REAP-S[q] score, risk knowledge, risk perception, total step average, and 7-day step average

First author, year of publication	Country	Study design	Intervention characteristics					
			Sample size (n [%] female)	Mean age (years)	Intervention design	Number of clinical encounters	Duration	Cardiometabolic risk outcomes
Oh et al, 2015 [25]	Republic of Korea	RCT	IG: n=212 (113 [53%]) CG: n=209 (99 [47%])	IG: 46.78 (SD 13.11) CG: 50.35 (SD 14.24)	IG: Body composition and pedometer data remote monitoring, and personalized PA and health online advice CG: Data records and PA and diet recommendations	Four in-person sessions	24 weeks	Weight and BMI
Petrella et al, 2014 [26]	Canada	RCT	IG: n=75 (55 [73%]) CG: n=74 (56 [76%])	IG: 55.7 (SD 10.1) CG: 57.8 (SD 8.7)	IG: Data telemonitoring, PA prescription, and goal setting CG: PA prescription and goal setting	Four in-person sessions	12 weeks	WC, SBP, DBP, TC, HDL-C, LDL-C, TG, FPG, HbA_{1c}, HOMA-IR, and high-sensitivity CRP
Stuckey et al, 2013 [27]	Canada	Pre-post study	n=12 (9 [75%])	56.9 (SD 7.0)	PA prescription, goal setting, and data telemonitoring	Two in-person sessions	8 weeks	WC, SBP, DBP, TG, HDL-C, FPG, VO_2 max[r], and steps

[a]RCT: randomized controlled trial.

[b]IG: intervention group.

[c]CG: control group.

[d]PA: physical activity.

[e]WC: waist circumference.

[f]SBP: systolic blood pressure.

[g]DBP: diastolic blood pressure.

[h]TC: total cholesterol.

[i]HDL-C: high-density lipoprotein cholesterol.

[j]LDL-C: low-density lipoprotein cholesterol.

[k]HbA_{1c}: glycated hemoglobin A_{1c}.

[l]FPG: fasting plasma glucose.

[m]CVD: cardiovascular disease.

[n]VFM: visceral fat mass.

[o]CRP: C-reactive protein.

[p]HOMA-IR: homeostatic model assessment of insulin resistance.

[q]REAP-S: rapid eating and activity assessment for patients.

[r]VO_2 max: predicted maximal oxygen capacity.

Risk of Bias

Seven out of nine studies were assessed as having a high risk of bias (including all single-arm pre-post studies), and the other two were assessed as having a moderate risk of bias. Analyzing each study individually, all single-arm pre-post studies had the lowest scores in the confounders and blinding domains (Table 2). All RCTs had a high risk of bias in the performance and detection bias domains (Table 3).

Table 2. Quality assessment of the included pre-post studies.

First author, year of publication	Selection bias	Study design	Confounders	Blinding	Data collection	Withdrawals	Risk of bias
Everett et al, 2018 [20]	Moderate	Moderate	Weak	Weak	Weak	Strong	High
Kim and Kang, 2013 [21]	Moderate	Moderate	Weak	Weak	Strong	Strong	High
Kim et al, 2014 [22]	Strong	Moderate	Weak	Weak	Strong	Strong	High
Stuckey et al, 2013 [27]	Moderate	Weak	Weak	Weak	Strong	Strong	High

Table 3. Quality assessment of the included randomized controlled trials.

First author, year of publication	Selection bias	Performance bias	Detection bias	Attrition bias	Reporting bias	Other bias	Risk of bias
Azar et al, 2016 [19]	Low	High	Unclear	Low	Low	Low	Moderate
Luley et al, 2014 [23]	Unclear	High	High	Low	Unclear	Low	High
Mann et al, 2016 [24]	High	High	High	Low	Low	Unclear	High
Oh et al, 2015 [25]	Low	Unclear	Unclear	Low	Low	Low	Low
Petrella et al, 2014 [26]	High	High	High	Low	Low	Unclear	High

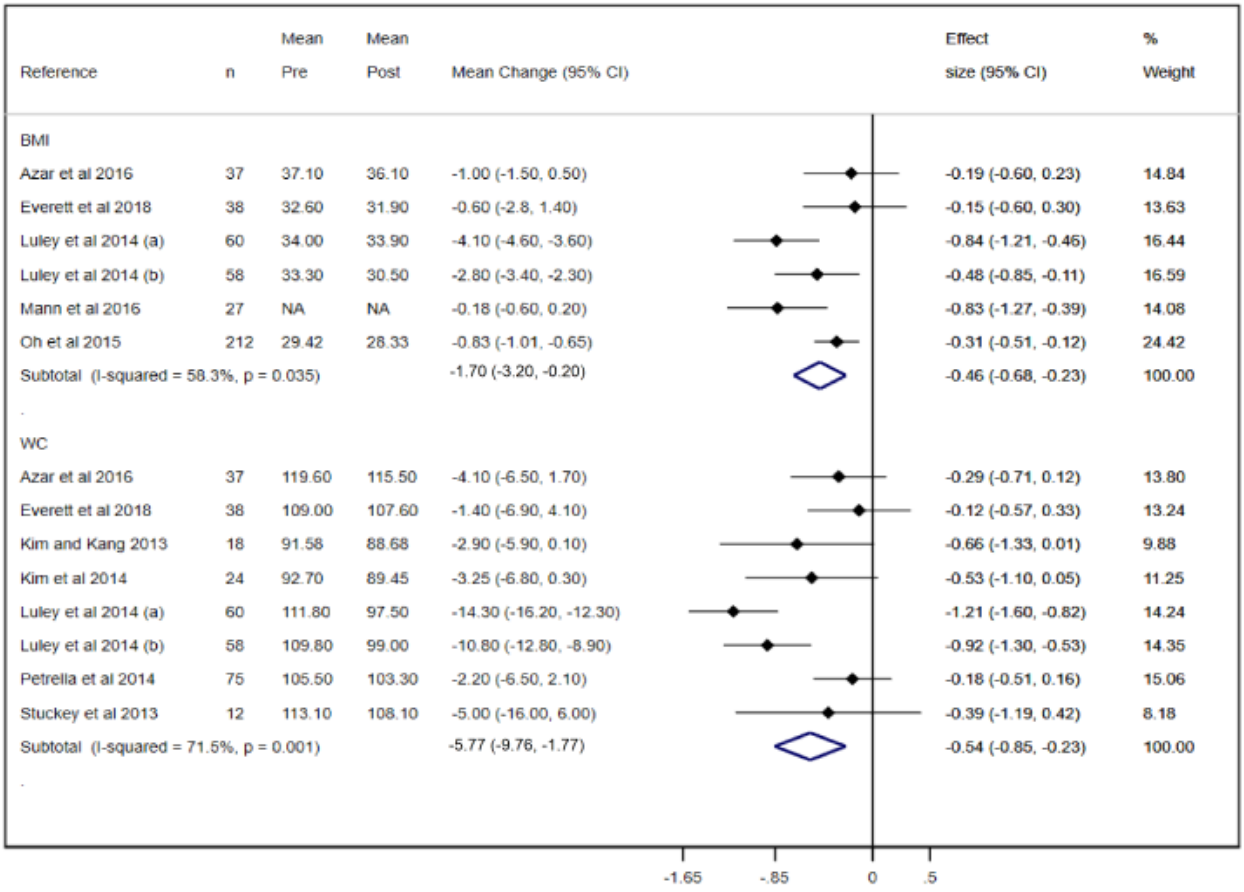
Meta-Analysis

Because of the small number of RCTs, only pooled effect estimates were calculated for mHealth promoting physical activity and lifestyle interventions in pre-post studies. The pre-post pooled mean changes with their heterogeneity statistics for each outcome category are presented below.

Body Composition

The mean changes were −1.70 kg/m^2 (95% CI −3.20 to −0.20; effect size: −0.46) for BMI and −5.77 cm (95% CI −9.76 to −1.77; effect size: −0.54) for WC. All pooled estimates showed moderate to substantial heterogeneity (BMI: I^2=58.3%; WC: I^2=71.5%) (Figure 3).

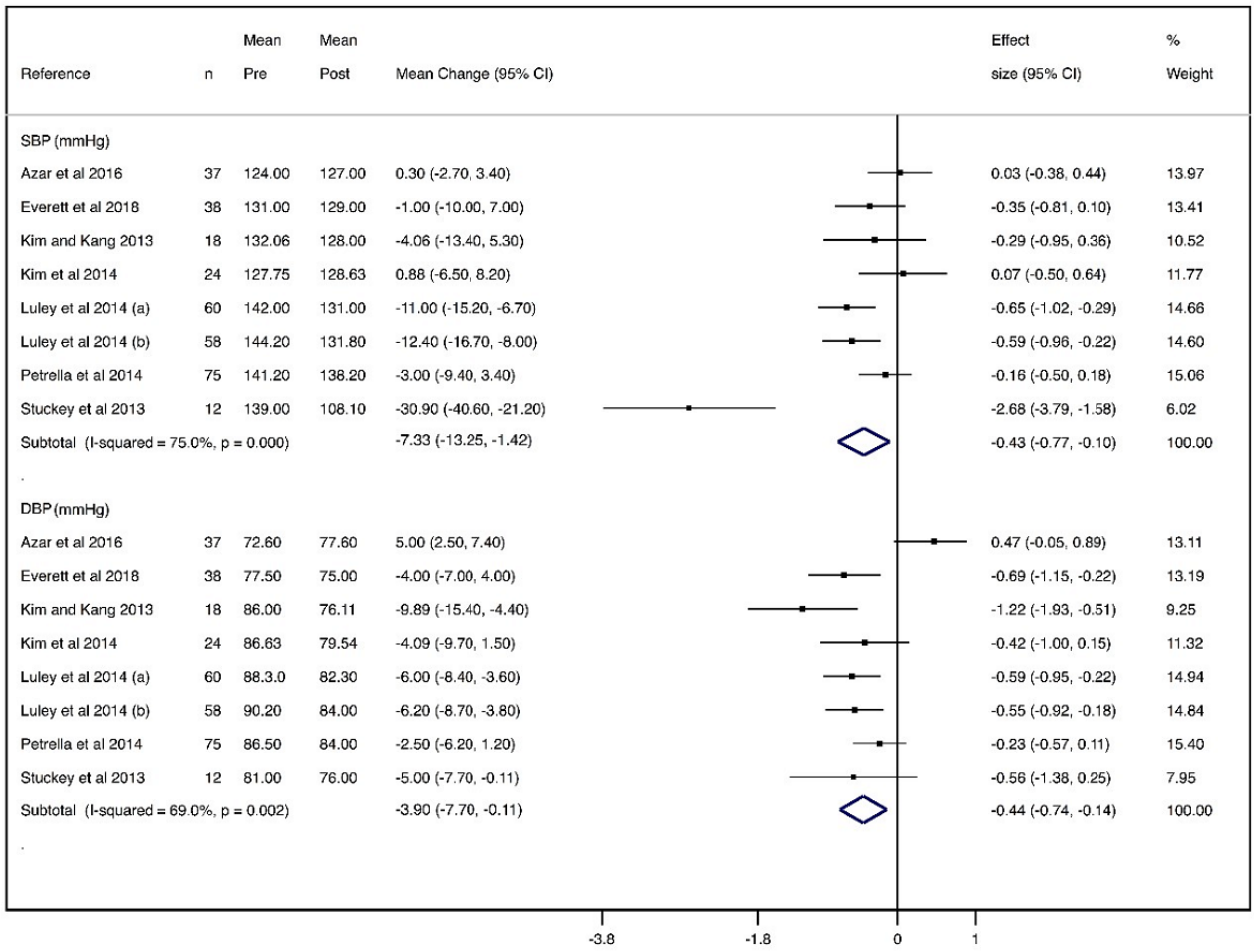
Figure 3. Forest plot of meta-analysis of mean changes and effect sizes for body composition parameters. WC: waist circumference.



Blood Pressure

The mean changes were −7.33 mmHg (95% CI −13.25 to −1.42; effect size: −0.43) for SBP and −3.90 mmHg (95% CI −7.70 to −0.11; effect size: −0.44) for DBP, with substantial heterogeneity for SBP (I^2=75%) and DBP (I^2=69%) (Figure 4).

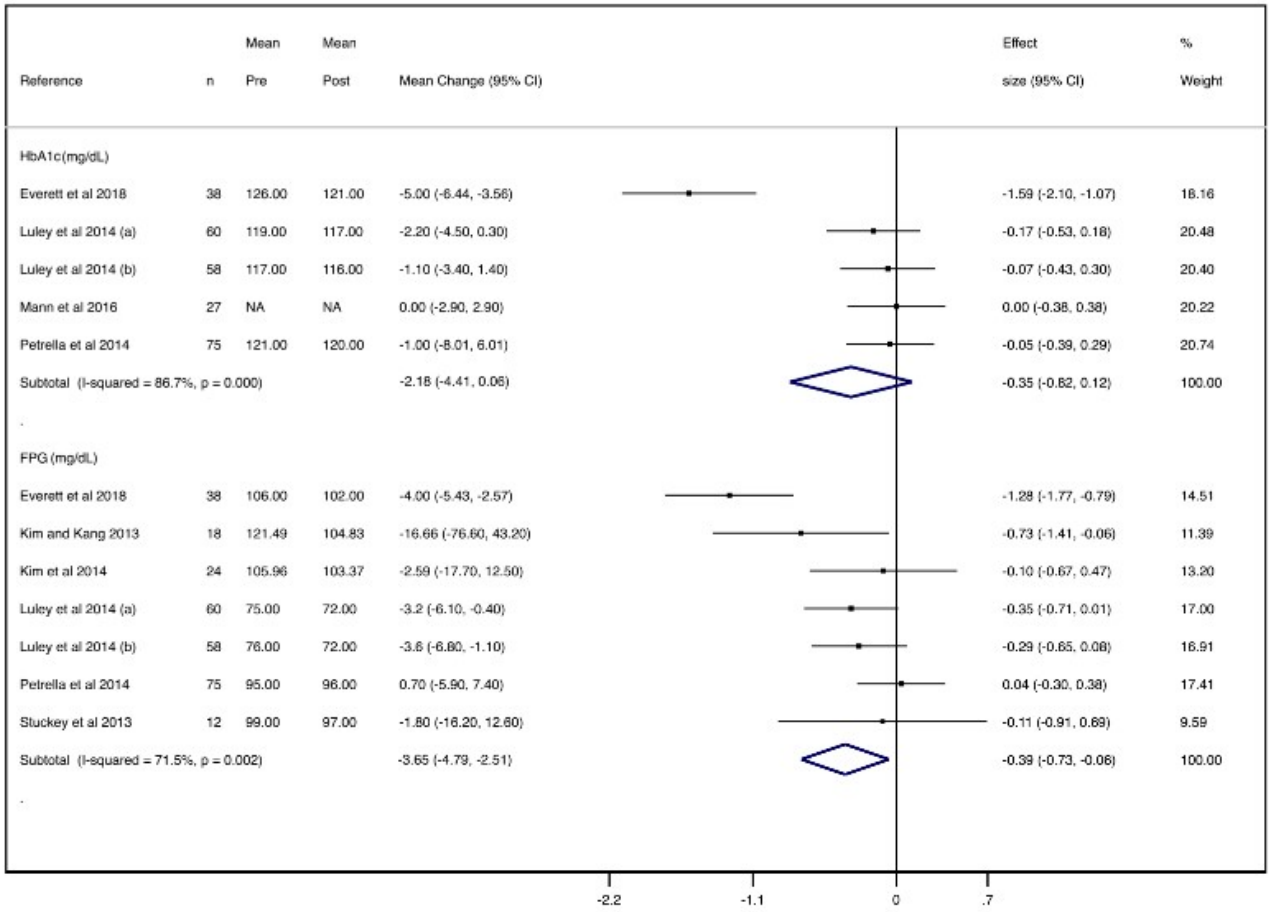
Figure 4. Forest plot of meta-analysis of mean changes and effect sizes for blood pressure parameters. DBP: diastolic blood pressure; SBP: systolic blood pressure.



Glucose Tolerance

The mean changes were −2.18 mg/dL (95% CI −4.41 to 0.06; effect size: −0.35) for HbA_{1c}, with considerable heterogeneity (I^2=86.7%), and −3.65 mg/dL (95% CI −4.79 to −2.51; effect size: −0.39) for FPG, with substantial heterogeneity (I^2=71.5%) (Figure 5).

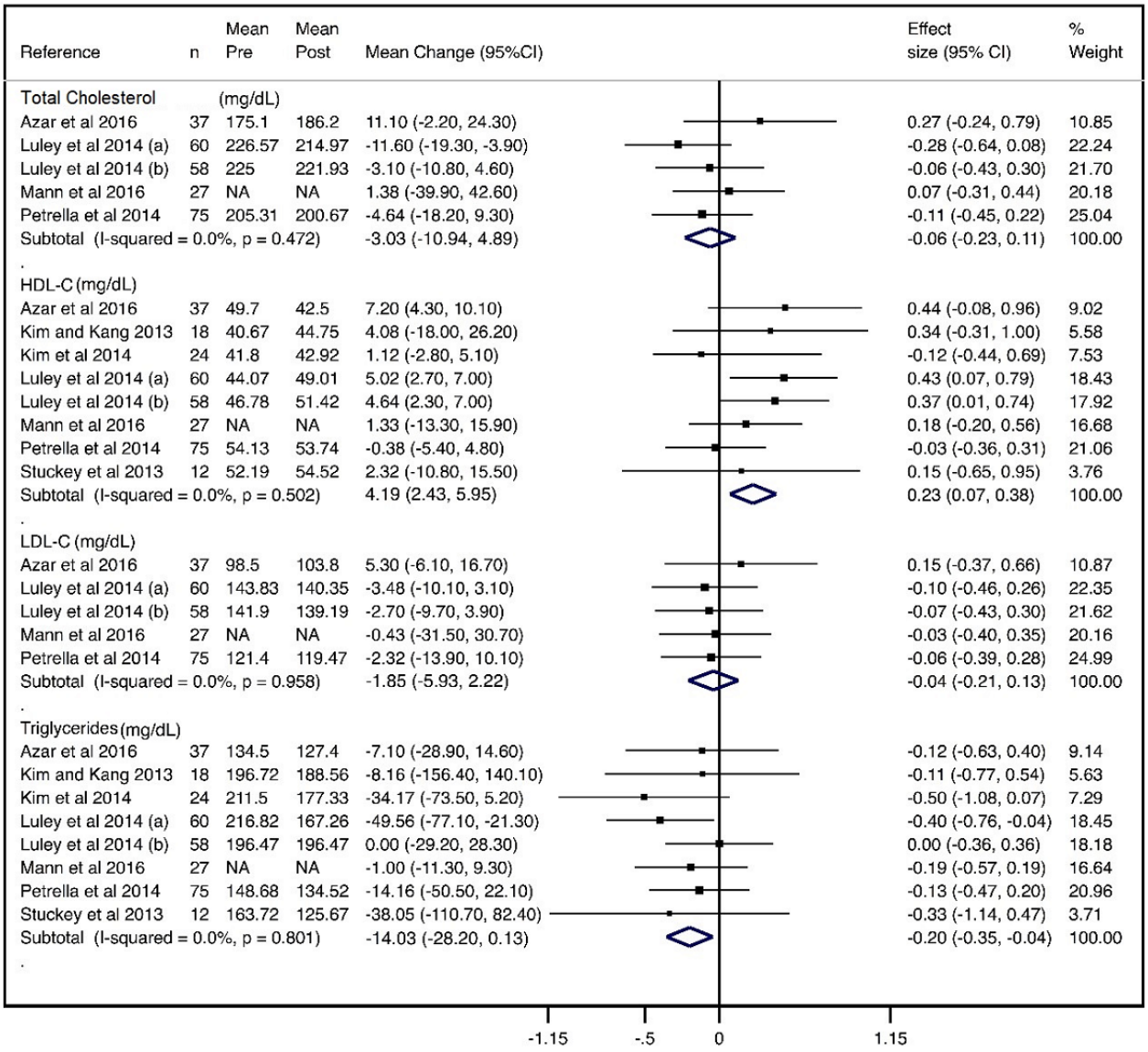
Figure 5. Forest plot of meta-analysis of mean changes and effect sizes for glucose tolerance parameters. FPG: fasting plasma glucose; HbA_{1c}: glycated hemoglobin A_{1c}.



Lipid Profile

The mean changes were −3.03 mg/dL (95% CI −10.94 to 4.89; effect size: −0.06) for total cholesterol, with no heterogeneity (I^2=0.0%), −1.85 mg/dL (95% CI −5.93 to 2.22; effect size: −0.04) for LDL-C, with no heterogeneity (I^2=0.0%), and −14.03 mg/dL (95% CI −28.20 to 0.13; effect size: −0.20) for triglycerides, with no heterogeneity (I^2=0.0%). Pooled mean changes were not relevant for any of the lipid parameters, except for HDL-C, which increased 4.19 mg/dL (95% CI 2.43-5.95; effect size: 0.23), with no heterogeneity (I^2=0.0%) (Figure 6).

Figure 6. Forest plot of meta-analysis of mean changes and effect sizes for lipid profile parameters. HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol.



Sensitivity Analysis

Studies were removed from the analysis one at a time in order to examine their individual impact on the pooled estimates. The pooled mean change of FPG was only significantly modified after removing the study by Everett et al [20] (−3.04 mg/dL; 95% CI −4.94 to −1.15; *P*=.002). None of the remaining studies potentially modified the pooled mean change estimate in magnitude or direction.

Meta-Regression

The random-effects meta-regression models showed that the percentage of females included in the study could influence the pooled estimates of the effect on BMI (*P*=.01) and triglycerides (*P*=.03), and the follow-up period could influence the pooled estimates of the effect on WC (*P*=.005) (Table 4).

Table 4. Meta-regression findings.

Variable	Age			Percentage of women			Follow-up period		
	Value, n	β (95% CI)	*P* value	Value, n	β (95% CI)	*P* value	Value, n	β (95% CI)	*P* value
Body composition									
BMI (kg/m^2)	6	0.04 (−0.37 to 0.45)	.81	6	0.07 (0.03 to 0.12)	.01	6	−0.01 (−0.03 to 0.01)	.21
Waist circumference (cm)	8	0.06 (−0.68 to 0.80)	.86	8	0.01 (−0.16 to 0.18)	.85	8	−0.02 (−0.03 to 0.01)	.005
Blood pressure									
SBP[a] (mmHg)	8	−0.28 (−1.76 to 1.19)	.65	8	−0.12 (−0.44 to 0.21)	.41	8	0.00 (−0.04 to 0.04)	.92
DBP[b] (mmHg)		0.49 (−0.02 to 1.01)	.057	8	0.08 (−0.06 to 0.22)	.23	8	0.00 (−0.02 to 0.02)	.87
Glucose tolerance									
HbA_{1c}[c] (mg/dL)	5	−0.29 (−1.10 to 0.52)	.34	5	0.04 (−0.08 to 0.16)	.37	5	0.01 (−0.04 to 0.07)	.449
FPG[d] (mg/dL)	7	0.33 (−0.12 to 0.78)	.12	7	0.08 (−0.04 to 0.19)	.42	7	0.04 (−0.02 to 0.03)	.75
Lipid profile									
Total cholesterol (mg/dL)	5	1.58 (−1.10 to 4.26)	.16	5	0.28 (−0.44 to 1.00)	.31	5	0.00 (−0.02 to 0.02)	.52
HDL-C[e] (mg/dL)	8	0.24 (−0.08 to 0.56)	.12	8	0.01 (−0.11 to 0.13)	.80	8	0.00 (0.00 to 0.00)	.11
LDL-C[f] (mg/dL)	5	0.73 (−1.21 to 2.67)	.32	5	0.10 (−0.34 to 0.53)	.52	5	0.00 (−0.01 to 0.01)	.75
Triglyceride (mg/dL)	8	0.74 (−2.97 to 4.45)	.64	8	0.49 (0.05 to 0.94)	.03	8	0.00 (−0.01 to 0.01)	.99

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

[c]HbA_{1c}: glycated hemoglobin A_{1c}.

[d]FPG: fasting plasma glucose.

[e]HDL-C: high-density lipoprotein cholesterol.

[f]LDL-C: low density lipoprotein- cholesterol.

Publication Bias

After visually examining the funnel plots and performing Egger tests for every parameter (Table 5), publication bias was only significant for WC (*P*=.04).

Table 5. Egger test findings.

Variable	*P* value
Body composition	
BMI (kg/m^2)	.98
Waist circumference (cm)	.04
Blood pressure	
SBP[a] (mmHg)	.45
DBP[b] (mmHg)	.58
Glucose tolerance	
HbA_{1c}[c] (mg/dL)	.42
FPG[d] (mg/dL)	.53
Lipid profile	
Total cholesterol (mg/dL)	.47
HDL-C[e] (mg/dL)	.31
LDL-C[f] (mg/dL)	.42
Triglyceride (mg/dL)	.24

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

[c]HbA_{1c}: hemoglobin A_{1c}.

[d]FPG: fasting plasma glucose.

[e]HDL-C: high-density lipoprotein cholesterol.

[f]LDL-C: low density lipoprotein- cholesterol.

Discussion

Principal Findings

Traditional approaches, such as physical activity programs, brief recommendation interventions, and pharmacological treatments, have been proven to be effective for controlling MetS [28]. However, they are expensive and time-consuming strategies that require a great commitment by both patients and practitioners. Our systematic review and meta-analysis suggested that physical activity and lifestyle interventions based on mHealth technologies are effective for reducing cardiometabolic risk, since they greatly improve body composition, blood pressure, FPG, and HDL-C levels. However, no relevant changes were observed in HbA_{1c}, total cholesterol, LDL-C, or triglyceride levels.

Our findings are in line with previous evidence on mHealth interventions in chronic disease patients that reported small to moderate positive effects on primary outcomes, such as cholesterol, weight, and blood pressure [29]. These findings show similar effects both when combining mHealth interventions with usual care (consisting of regular hospital visits, regular visits by primary health care providers at home, or visits to the general practitioner) [30-35] and when mHealth interventions are carried out instead of usual care [36-40]. Such results are consistent with our findings despite the different populations targeted; however, our results show much smaller effect sizes for total cholesterol, LDL-C, and triglycerides, which may be explained by the fewer number of included studies reporting those outcomes.

Among the factors involved in the worldwide increase in sedentary behavior, the use of information and communication technologies and particularly the increase in screen time have been described as the main drivers of low daily energy expenditure [41]. Thus, to involve these technologies as vehicles of preventive interventions could be both a risk and an opportunity. Even though we were unable to demonstrate the superiority of mHealth promoting physical activity and lifestyle interventions over usual care (in-person consultations with clinicians) owing to the scarcity of studies comparing data between control and intervention groups, our results showed that mHealth interventions are effective in improving cardiometabolic risk. Our data regarding the effects of interventions based on mHealth technologies are similar to those involving traditional care [42], suggesting that they could represent an alternative treatment strategy because of their acceptability, scalability, cost-effectiveness, customization, and ability to send time-sensitive messages with an “always on” device [43]. Moreover, mHealth physical activity interventions reduce in-person health provider time and increase self-care by enabling patients to manage their progress [23].

However, our results must be interpreted cautiously, since they are threatened by several limitations that should be acknowledged. First, although a systematic search was carried out through the most well-known databases by two different

researchers, some scientific contributions reported as grey literature may have been missed in our systematic search. Second, overall, the risk of bias of the included studies was rated as high. Third, there has been some criticism about using single-group studies for evaluating effectiveness [44], and only pre-post estimates for the intervention group could be used because of the scarcity of RCTs reporting the necessary data for control groups. Fourth, there was heterogeneity of interventions owing to differences in components (ie, self-monitoring, type, and persuasiveness of advice), length, and lack of precision in descriptions of the type and intensity of physical activity. Fifth, it was difficult to elucidate whether the outcome changes were due to physical activity or other lifestyle interventions as they were all designed as multicomponent interventions, and hence, it was impossible to isolate each component effect. Sixth, although our results were calculated as pre-post effect sizes, previous literature has recommended avoiding them in meta-analyses [45]. Seventh, the small sample size of some of the included studies diminished their reliability. Eighth, cardiovascular risk parameters were not the main outcomes of most studies. Lastly, none of the included studies used the mHealth evidence reporting and assessment (mERA) checklist, a tool developed by the WHO mHealth Technical Evidence Review Group in order to improve the completeness of reporting mHealth interventions [46]. Despite all of these limitations, our study, as the only updated synthesis evaluating mHealth technologies promoting physical activity and lifestyle interventions to reduce cardiovascular risk in individuals with MetS, establishes a base for future research providing more consistent evidence of their effectiveness.

Conclusion

Our results show an overall positive effect of physical activity and lifestyle interventions delivered through mobile technologies on MetS indicators, suggesting that they may be effective tools for MetS management. However, further research is needed in order to enable a comparison between the traditional clinical approach and new interventions through mHealth technologies, as these results may be due to the lack of appropriate comparable RCTs because these technologies are novel. Additionally, estimating the independent effect of each component of these interventions would be interesting, and it is important to standardize the implementation of multicomponent interventions in such a way that enough evidence is available for consideration in clinical practice guidelines.

Acknowledgments

This study was funded by Consejería de Educación, Cultura y Deportes – Junta de Comunidades de Castilla-La Mancha and FEDER funds (SBPLY/17/180501/000533).

Conflicts of Interest

None declared.

References

1. Alberti K, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the Metabolic Syndrome. Circulation 2009 Oct 20;120(16):1640-1645. [doi: 10.1161/circulationaha.109.192644]
2. Grundy SM. Metabolic syndrome update. Trends Cardiovasc Med 2016 May;26(4):364-373. [doi: 10.1016/j.tcm.2015.10.004] [Medline: 26654259]
3. Zimmet P, Alberti KG, Shaw J. Global and societal implications of the diabetes epidemic. Nature 2001 Dec 13;414(6865):782-787. [doi: 10.1038/414782a] [Medline: 11742409]
4. Mottillo S, Filion KB, Genest J, Joseph L, Pilote L, Poirier P, et al. The metabolic syndrome and cardiovascular risk a systematic review and meta-analysis. J Am Coll Cardiol 2010 Sep 28;56(14):1113-1132 [FREE Full text] [doi: 10.1016/j.jacc.2010.05.034] [Medline: 20863953]
5. Cornier M, Dabelea D, Hernandez T, Lindstrom R, Steig A, Stob N, et al. The metabolic syndrome. Endocr Rev 2008 Dec;29(7):777-822 [FREE Full text] [doi: 10.1210/er.2008-0024] [Medline: 18971485]
6. Ostman C, Smart NA, Morcos D, Duller A, Ridley W, Jewiss D. The effect of exercise training on clinical outcomes in patients with the metabolic syndrome: a systematic review and meta-analysis. Cardiovasc Diabetol 2017 Aug 30;16(1):110 [FREE Full text] [doi: 10.1186/s12933-017-0590-y] [Medline: 28854979]
7. Reis RS, Salvo D, Ogilvie D, Lambert EV, Goenka S, Brownson RC. Scaling up physical activity interventions worldwide: stepping up to larger and smarter approaches to get people moving. The Lancet 2016 Sep;388(10051):1337-1348. [doi: 10.1016/s0140-6736(16)30728-0]
8. Craig J, Petterson V. Introduction to the Practice of Telemedicine. J Telemed Telecare 2016 Jun 24;11(1):3-9. [doi: 10.1177/1357633x0501100102]
9. Chow CK, Ariyarathna N, Islam SM, Thiagalingam A, Redfern J. mHealth in Cardiovascular Health Care. Heart Lung Circ 2016 Aug;25(8):802-807. [doi: 10.1016/j.hlc.2016.04.009] [Medline: 27262389]
10. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]

11. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Cochrane Collaboration. 2011. URL: https://handbook-5-1.cochrane.org/ [accessed 2019-11-08]
12. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
13. Thomas BH, Ciliska D, Dobbins M, Micucci S. Quality assessment tool for quantitative studies. National Collaborating Centre for Methods and Tools. 2008. URL: https://www.nccmt.ca/knowledge-repositories/search/14 [accessed 2019-11-08]
14. Armijo-Olivo S, Stiles C, Hagen N, Biondo P, Cummings G. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract 2012 Feb;18(1):12-18. [doi: 10.1111/j.1365-2753.2010.01516.x] [Medline: 20698919]
15. DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clinical Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2]
16. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
17. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Statist. Med 1999 Oct 30;18(20):2693-2708. [doi: 10.1002/(sici)1097-0258(19991030)18:20<2693::aid-sim235>3.0.co;2-v]
18. Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. BMJ 2001 Jul 14;323(7304):101-105 [FREE Full text] [doi: 10.1136/bmj.323.7304.101] [Medline: 11451790]
19. Azar KM, Koliwad S, Poon T, Xiao L, Lv N, Griggs R, et al. The Electronic CardioMetabolic Program (eCMP) for Patients With Cardiometabolic Risk: A Randomized Controlled Trial. J Med Internet Res 2016 May 27;18(5):e134 [FREE Full text] [doi: 10.2196/jmir.5143] [Medline: 27234480]
20. Everett E, Kane B, Yoo A, Dobs A, Mathioudakis N. A Novel Approach for Fully Automated, Personalized Health Coaching for Adults with Prediabetes: Pilot Clinical Trial. J Med Internet Res 2018 Feb 27;20(2):e72 [FREE Full text] [doi: 10.2196/jmir.9723] [Medline: 29487046]
21. Kim C, Kang S. Development and a Pilot Test of an Internet-Based Cardiovascular Risk Reduction Program for Korean Male Workers With Metabolic Syndrome. CIN: Computers, Informatics, Nursing 2013;31(4):157-166. [doi: 10.1097/nxn.0b013e3182812829]
22. Kim C, Schlenk EA, Kang S, Park J. Effects of an internet-based lifestyle intervention on cardio-metabolic risks and stress in Korean workers with metabolic syndrome: a controlled trial. Patient Educ Couns 2015 Jan;98(1):111-119. [doi: 10.1016/j.pec.2014.10.013] [Medline: 25468401]
23. Luley C, Blaik A, Götz A, Kicherer F, Kropf S, Isermann B, et al. Weight loss by telemonitoring of nutrition and physical activity in patients with metabolic syndrome for 1 year. J Am Coll Nutr 2014;33(5):363-374. [doi: 10.1080/07315724.2013.875437] [Medline: 25105874]
24. Mann DM, Palmisano J, Lin JJ. A pilot randomized trial of technology-assisted goal setting to improve physical activity among primary care patients with prediabetes. Prev Med Rep 2016 Dec;4:107-112 [FREE Full text] [doi: 10.1016/j.pmedr.2016.05.012] [Medline: 27413670]
25. Oh B, Cho B, Han MK, Choi H, Lee MN, Kang H, et al. The Effectiveness of Mobile Phone-Based Care for Weight Control in Metabolic Syndrome Patients: Randomized Controlled Trial. JMIR Mhealth Uhealth 2015 Aug 20;3(3):e83 [FREE Full text] [doi: 10.2196/mhealth.4222] [Medline: 26293568]
26. Petrella RJ, Stuckey MI, Shapiro S, Gill DP. Mobile health, exercise and metabolic risk: a randomized controlled trial. BMC Public Health 2014 Oct 18;14:1082 [FREE Full text] [doi: 10.1186/1471-2458-14-1082] [Medline: 25326074]
27. Stuckey MI, Kiviniemi AM, Petrella RJ. Diabetes and technology for increased activity study: the effects of exercise and technology on heart rate variability and metabolic syndrome risk factors. Front Endocrinol (Lausanne) 2013;4:121 [FREE Full text] [doi: 10.3389/fendo.2013.00121] [Medline: 24065952]
28. Dunkley A, Charles K, Gray L, Camosso-Stefinovic J, Davies M, Khunti K. Effectiveness of interventions for reducing diabetes and cardiovascular disease risk in people with metabolic syndrome: systematic review and mixed treatment comparison meta-analysis. Diabetes Obes Metab 2012 Jul;14(7):616-625. [doi: 10.1111/j.1463-1326.2012.01571.x] [Medline: 22284386]
29. Eland-de Kok P, van Os-Medendorp H, Vergouwe-Meijer A, Bruijnzeel-Koomen C, Ros W. A systematic review of the effects of e-health on chronically ill patients. J Clin Nurs 2011 Nov;20(21-22):2997-3010. [doi: 10.1111/j.1365-2702.2011.03743.x] [Medline: 21707807]
30. Bergmo T, Wangberg S, Schopf T, Solvoll T. Web-based consultations for parents of children with atopic dermatitis: results of a randomized controlled trial. Acta Paediatr 2009 Feb;98(2):316-320. [doi: 10.1111/j.1651-2227.2008.01033.x] [Medline: 18795905]
31. Bond GE, Burr R, Wolf FM, Price M, McCurry SM, Teri L. The effects of a web-based intervention on the physical outcomes associated with diabetes among adults age 60 and older: a randomized trial. Diabetes Technol Ther 2007 Feb;9(1):52-59 [FREE Full text] [doi: 10.1089/dia.2006.0057] [Medline: 17316098]

32. Cho J, Chang S, Kwon H, Choi Y, Ko S, Moon S, et al. Long-term effect of the Internet-based glucose monitoring system on HbA1c reduction and glucose stability: a 30-month follow-up study for diabetes management with a ubiquitous medical care system. Diabetes Care 2006 Dec;29(12):2625-2631. [doi: 10.2337/dc05-2371] [Medline: 17130195]
33. Finkelstein SM, Speedie SM, Potthoff S. Home telehealth improves clinical outcomes at lower cost for home healthcare. Telemed J E Health 2006 Apr;12(2):128-136. [doi: 10.1089/tmj.2006.12.128] [Medline: 16620167]
34. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007 Jun;9(3):297-306. [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
35. Kim S, Kim H. Effectiveness of mobile and internet intervention in patients with obese type 2 diabetes. Int J Med Inform 2008 Jun;77(6):399-404. [doi: 10.1016/j.ijmedinf.2007.07.006] [Medline: 17881285]
36. Hee-Sung K. Impact of Web-based nurse's education on glycosylated haemoglobin in type 2 diabetic patients. J Clin Nurs 2007 Jul;16(7):1361-1366. [doi: 10.1111/j.1365-2702.2007.01506.x] [Medline: 17584355]
37. Kwon H, Cho J, Kim H, Song B, Ko S, Lee J, et al. Establishment of blood glucose monitoring system using the internet. Diabetes Care 2004 Feb;27(2):478-483. [doi: 10.2337/diacare.27.2.478] [Medline: 14747232]
38. Shea S, Weinstock RS, Starren J, Teresi J, Palmas W, Field L, et al. A Randomized Trial Comparing Telemedicine Case Management with Usual Care in Older, Ethnically Diverse, Medically Underserved Patients with Diabetes Mellitus. Journal of the American Medical Informatics Association 2006 Jan 01;13(1):40-51. [doi: 10.1197/jamia.m1917]
39. Southard BH, Southard DR, Nuckolls J. Clinical trial of an Internet-based case management system for secondary prevention of heart disease. J Cardiopulm Rehabil 2003;23(5):341-348. [doi: 10.1097/00008483-200309000-00003] [Medline: 14512778]
40. Yoon K, Kim H. A short message service by cellular phone in type 2 diabetic patients for 12 months. Diabetes Res Clin Pract 2008 Feb;79(2):256-261. [doi: 10.1016/j.diabres.2007.09.007] [Medline: 17988756]
41. Middelbeek L, Breda J. Obesity and Sedentarism: Reviewing the Current Situation Within the WHO European Region. Curr Obes Rep 2013 Feb 22;2(1):42-49. [doi: 10.1007/s13679-013-0054-y]
42. Maddison R, Rawstorn JC, Shariful Islam SM, Ball K, Tighe S, Gant N, et al. mHealth Interventions for Exercise and Risk Factor Modification in Cardiovascular Disease. Exercise and Sport Sciences Reviews 2019;47(2):86-90. [doi: 10.1249/jes.0000000000000185]
43. Whittaker R, Borland R, Bullen C, Lin RB, McRobbie H, Rodgers A. Mobile phone-based interventions for smoking cessation. Sao Paulo Med. J 2010;128(2):106-107. [doi: 10.1590/s1516-31802010000200014]
44. Ip S, Paulus JK, Balk EM, Dahabreh IJ, Avendano EE, Lau J. Role of Single Group Studies in Agency for Healthcare Research and Quality Comparative Effectiveness Reviews. National Center for Biotechnology Information, U.S. National Library of Medicine. 2013. URL: https://www.ncbi.nlm.nih.gov/books/NBK121314/ [accessed 2019-11-11]
45. Cuijpers P, Weitz E, Cristea IA, Twisk J. Pre-post effect sizes should be avoided in meta-analyses. Epidemiol Psychiatr Sci 2016 Oct 28;26(4):364-368. [doi: 10.1017/s2045796016000809]
46. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]

Abbreviations

DBP: diastolic blood pressure
FPG: fasting plasma glucose
HbA_{1c}: glycated hemoglobin A_{1c}
HDL-C: high-density lipoprotein cholesterol
LDL-C: low-density lipoprotein cholesterol
MetS: metabolic syndrome
mHealth: mobile-based health
RCT: randomized controlled trial
SBP: systolic blood pressure
WC: waist circumference

Edited by G Eysenbach; submitted 13.01.20; peer-reviewed by UB Johansson, J Arsenijevic, L Grepo; comments to author 12.06.20; revised version received 26.06.20; accepted 07.07.20; published 31.08.20.

Please cite as:
Sequi-Dominguez I, Alvarez-Bueno C, Martinez-Vizcaino V, Fernandez-Rodriguez R, del Saz Lara A, Cavero-Redondo I
Effectiveness of Mobile Health Interventions Promoting Physical Activity and Lifestyle Interventions to Reduce Cardiovascular Risk Among Individuals With Metabolic Syndrome: Systematic Review and Meta-Analysis
J Med Internet Res 2020;22(8):e17790
URL: http://www.jmir.org/2020/8/e17790/
doi:10.2196/17790
PMID:32865503

Review

Effects of Internet-Based Cognitive Behavioral Therapy in Routine Care for Adults in Treatment for Depression and Anxiety: Systematic Review and Meta-Analysis

Anne Etzelmueller[1,2,3,4], Dipl-Psych; Christiaan Vis[3,4], MSc; Eirini Karyotaki[3,5], PhD; Harald Baumeister[6], PhD, Dipl-Psych; Nickolai Titov[7], PhD; Matthias Berking[1], PhD; Pim Cuijpers[3,4], PhD; Heleen Riper[3,4,8], PhD; David Daniel Ebert[1,2,3], PhD

[1]Department of Clinical Psychology and Psychotherapy, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany

[2]GET.ON Institute GmbH, Hamburg, Germany

[3]Department of Clinical, Neuro-, & Developmental Psychology, Faculty of Behavioural and Movement Sciences, VU Amsterdam, Amsterdam, Netherlands

[4]Amsterdam Public Health Research Institute, Mental Health, Amsterdam, Netherlands

[5]Department of Global Health and Social Medicine, Harvard Medical School, USA, Boston, MA, United States

[6]Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education, Ulm University, Ulm, Germany

[7]eCentre Clinic Department of Psychology, Macquarie University, Sydney, Australia

[8]Community Mental Health Centre GGZ inGeest, Amsterdam, Netherlands

Corresponding Author:
Anne Etzelmueller, Dipl-Psych
Department of Clinical Psychology and Psychotherapy
Friedrich-Alexander University Erlangen-Nuremberg
Nägelsbachstr 25a
Erlangen, 91052
Germany
Phone: 49 9131 85 67575
Email: anne.etzelmueller@fau.de

Abstract

Background: Although there is evidence for the efficacy of internet-based cognitive behavioral therapy (iCBT), the generalizability of results to routine care is limited.

Objective: This study systematically reviews effectiveness studies of guided iCBT interventions for the treatment of depression or anxiety.

Methods: The acceptability (uptake, participants' characteristics, adherence, and satisfaction), effectiveness, and negative effects (deterioration) of nonrandomized pre-post designs conducted under routine care conditions were synthesized using systematic review and meta-analytic approaches.

Results: A total of 19 studies including 30 groups were included in the analysis. Despite high heterogeneity, individual effect sizes of investigated studies indicate clinically relevant changes, with effect sizes ranging from Hedges' g=0.42-1.88, with a pooled effect of 1.78 for depression and 0.94 for anxiety studies. Uptake, participants' characteristics, adherence, and satisfaction indicate a moderate to high acceptability of the interventions. The average deterioration across studies was 2.9%.

Conclusions: This study provides evidence supporting the acceptability and effectiveness of guided iCBT for the treatment of depression and anxiety in routine care. Given the high heterogeneity between interventions and contexts, health care providers should select interventions that have been proven in randomized controlled clinical trials. The successful application of iCBT may be an effective way of increasing health care in multiple contexts.

(J Med Internet Res 2020;22(8):e18100) doi:10.2196/18100

KEYWORDS

internet-based interventions; depression; anxiety; effectiveness; acceptability; routine care

Introduction

Depressive and anxiety disorders are common mental health problems associated with significant suffering, impairment, and reduction in the quality of life [1,2]. Both disorders lead to considerable socioeconomic costs through decreased work productivity and higher utilization of health care services [3,4].

Despite the proven effectiveness of psychotherapy in the treatment of depression and anxiety [5], the provision of evidence-based treatments depicts a constant challenge given the barriers such as the shortage of treatment, uneven distribution of trained providers, delayed treatment provision, and inadequacy of treatment [6,7]. Furthermore, research on patients' preferences has shown that many do neither make use of psychotherapeutic treatments nor do they receive psychopharmacological treatment [7]. Using the internet to provide psychotherapeutic interventions may increase the coverage of usual care services [8,9] by providing highly accessible and scalable interventions reaching people who cannot be reached otherwise. Recent research suggests that internet-based cognitive behavioral therapy (iCBT) with therapeutic guidance is effective for the prevention [10,11] and treatment [12-15] of common mental disorders. Systematic reviews on studies were also able to show comparable effects to face-to-face treatments in adults [16,17]. In a recent meta-analysis, Romijn et al [13] showed that iCBT interventions for anxiety disorders can also have significant effects obtained in trials implemented in clinical care. They also found that effects were smaller in samples recruited in clinical practice than in samples recruited with an open recruitment method compared with waitlist-control groups [13], which raises the question of the effects of iCBT when implemented in routine practice.

Although randomized controlled trials (RCTs) are considered the gold standard in exploring the efficacy of mental health interventions, the idealized and controlled nature of these trials limits the generalizability of findings to routine care populations [18]. RCTs maximize the internal validity, to ensure that the effect found can be attributed to the investigated intervention [19,20]. Thus, RCT findings are restricted by controlled protocols, explicit eligibility criteria, and patient recruitment and randomization procedures. RCTs provide a highly structured environment, which is considered to possibly have an adherence-fostering effect [21,22]. The efficacy derived from RCTs of internet-based interventions might be overestimated for what can be expected when implementing in routine care, limiting the knowledge base for routine clinical practice [20].

Hence, after establishing the efficacy of an intervention and its subsequent implementation, the so-called *phase IV clinical trials* should follow investigating benefits when implemented as well as potential negative effects implemented [23,24]. Thus, the investigation of the effectiveness of iCBT under routine care conditions is an important part of the evaluation of these services before wide-scale adoption.

Andersson and Hedman [25] reported on the effectiveness of iCBT within four controlled trials and eight open studies for a multitude of mental health problems, indicating that it might be possible to replicate the findings of controlled efficacy trials on guided iCBT in clinical practice. However, in that review, both routine care and RCTs were included, and only eight studies reported effects when the service was delivered under routine conditions. Recently, Andrews et al [15] reported the results of computer-based treatments of depression, panic disorder, generalized anxiety disorder, and social phobia in randomized trials. They also identified eight studies on internet-based treatments in routine clinical practice when delivered outside of a randomized trial reporting an average effect size of g=1.07 across all 4 disorders [15]. However, since then, many more studies have been published. In addition, this review did not specifically try to identify nonrandomized trials, possibly leading to unidentified articles. Additionally, they did not provide disorder-specific results, specific results on guided treatments by mixing guided and unguided treatments, and did not investigate the acceptability and potential negative effects.

The aim of this study was to examine the effects of guided iCBT for the treatment of depression and anxiety under routine care conditions on symptom change, acceptability (uptake, participants' characteristics, adherence, and satisfaction), and predictors of negative effects (deterioration and side effects).

Methods

We report this meta-analysis in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (Multimedia Appendix 1) [26]. This meta-analysis was registered at international prospective register of systematic reviews (PROSPERO; trial registration: CRD42018095704).

We searched PubMed, PsychINFO, and the Cochrane library. We used index terms and text words associated with depression and anxiety, internet interventions, and routine care (for a full search string, the reader is referred to Multimedia Appendix 2). Furthermore, we contacted experts in the field to ask whether they were aware of the studies that we did not identify through our systematic literature searches. Furthermore, we conducted reference tracking on the identified studies and previous meta-analyses in the field [5,14,15,27]. The resulting hits of our literature searches were screened on titles and abstracts by 2 independent reviewers (AE and CV). Studies considered as potentially relevant were screened on full text by the same reviewers independently. In case of disagreement, the opinion of a third senior reviewer (DE) was sought.

Inclusion Criteria

We included studies that (1) examined the effectiveness of a guided or blended iCBT in (2) treating adults with depressive and/or anxiety symptoms (3) under routine care conditions (4) in a pre-post design. We followed the inclusion of adults and older adolescents (aged >16 years) within the treatment provision for adults, as reported in the original studies.

We defined routine care studies as effectiveness studies, which were conducted as nonrandomized clinical trials in settings equal to or representative of routine practice [28]. The definition of routine care differs between countries and health care systems and describes the established way of working at the time of the original study. Depression and anxiety symptoms had to be

established based on cutoff scores on self-report outcome measures, clinical diagnosis, or expert opinion. The definition of anxiety symptoms is based on the Diagnostic and Statistical Manual of Mental Disorders IV classification criteria for anxiety disorders. Furthermore, the interventions were considered as guided when the guidance was related to the therapeutic content [29] and as blended when the internet-based intervention was combined with face-to-face elements in one integrated standardized treatment protocol [27]. Guidance can be delivered via email, a secure message system, telephone, or face-to-face contact and via video or face-to-face contact in blended treatments. Finally, both disorder-specific and transdiagnostic interventions (targeted at both depression and anxiety simultaneously) were included.

Exclusion Criteria

We excluded studies that did not (1) focus primarily on anxiety or depression or (2) provide sufficient data for the calculation of the effect sizes. Studies were also excluded if (1) the service had only been provided as part of a research study, (2) the study could be considered as a feasibility or pilot trial, or (3) patients were randomized at an individual level. However, cluster randomized trials were considered eligible, in which randomization took place not on an individual level but, for example, on a health care institution level. For the definition of feasibility and pilot trials, we followed the NIHR Evaluation, Trials and Studies Coordinating Centre definition of pilot and feasibility trials [30], as recommended by Arain et al [31]. Feasibility trials were defined as "pieces of research done before a main study" (designed around the research question "Can this be done?"), and pilot studies are defined as a version of the main study that is "run in miniature to test whether the components of the main study can all work together" [31]. Additionally, we only included studies published in English, German, or Dutch language.

Data Extraction

We extracted data related to study and iCBT service–related characteristics, acceptability, effects on symptom change, negative effects, and data related to the risk of bias of reported results.

Study characteristics included the year of publication, the country in which the study was conducted, the year of data collection, sample size, eligibility criteria (establishment of depression and/or anxiety diagnosis at baseline [standardized clinical interview, cutoff on standardized questionnaire, and clinical judgment], inclusion of severe cases [yes/no], and exclusion of cases with suicidal ideation [yes/no], and approach to data analysis [ITT/completer]).

iCBT service–related characteristics included intervention name, the symptoms targeted (depression and/or anxiety), if it was a blended treatment (yes/no), evidence base for the used intervention (positive results based on at least one randomized clinical trial [yes/no]), and whether it was a symptom-specific or transdiagnostic treatment. We also included the recruitment pathway (open community, clinical referral, and both), the number of planned intervention modules, guidance focus (content-focused, motivational-focused, and administrative-focused), guidance delivery format (synchronous vs asynchronous, within the treatment platform vs outside, eg, by email), and guidance moment (as a reaction to an action of the participant [eg, after the participant finished a session, as a reaction to a nonresponse] or planned in different intervals [eg, weekly or biweekly]). Furthermore, we included guides' professional training (psychotherapist, psychiatrist, general practitioner [GP], psychologist, psychological registrar, nurse, coach [with lived experience]), training of professionals in iCBT (yes/no), supervision of professionals by a trained clinician (yes/no), the planned and actual intensity of guidance in minutes, and if there was a guidance manual provided (yes/no). Additional information on whether a standardized procedure in the case of symptom deterioration and crisis (yes/no) has been established was included.

Acceptability data were extracted with regard to uptake (the number of people screened for the service, people included, and participants starting the treatment), patient characteristics (age and gender), average symptom severity at baseline, adherence (ie, number of completed modules), mean treatment duration in weeks, and participant satisfaction. Negative effects were extracted with regard to average effects on symptom deterioration, other side effects, and reports of specific subgroups at risk for symptom deterioration.

Two reviewers (AE and CV) extracted the data independently, and data sets were merged. Differences and points of uncertainty were discussed and checked by returning to the original article and in some cases to the authors of the respective article.

Risk of Bias Assessment

Assessing the quality of naturalistic observational studies is challenging as there is no widely accepted tool in doing so [32]. Moreover, established guidelines for the quality assessment of nonrandomized trials are only partially applicable, as they assume comparisons of interventions (Risk Of Bias In Nonrandomized Studies of Interventions-I [33]). Thus, in this study, we selected and adapted criteria from two quality assessment tools [33,34] and adapted them to this study's purposes to evaluate the risk of bias of the included studies. For the present risk of bias assessment, we discussed the aforementioned assessment tools among all coauthors of this manuscript and derived the analysis criteria described in Multimedia Appendix 3 [35]. As a result, we evaluated (1) researcher allegiance (defined as the first or last author of the study also being the first or last author of the intervention development or efficacy paper), (2) confounding introduced by patients' participation in other treatments, (3) confounding introduced by significant confounding variables identified within the individual study (meaning any predictors included such as age, guidance, or recruitment pathway), (4) selection bias introduced by the study population (ie, have the studies only reported on completer data), and (5) selective outcome reporting in comparison with the study protocol or diagnostic measures administered as mentioned in the original studies' methods section. A description of the risk of bias assessment and its operationalization can be found in Multimedia Appendix 3. With regard to *Researcher Allegiance*, we chose the above definition after consideration among the authors and evaluated

a study as at high risk of researcher allegiance when the first or last author of the study was also involved in the development of the treatment manual of the psychotherapy involved or the reporting on the interventions' efficacy. Although the validity of other indicators has been questioned, the involvement of a researcher in developing the treatment under investigation can be considered a valid indicator of potential researcher allegiance [36].

Two reviewers evaluated the quality of the included studies independently (AE and CV). Any disagreement between reviewers was solved by a thorough discussion. If the disagreement could not be resolved, a third senior reviewer was consulted (DE).

Statistical Analysis

Our primary outcome was the reduction of depressive or anxiety symptoms from pre- to posttest assessment. We calculated the difference in depression and anxiety symptoms between pre-post assessment divided by the weighted, pooled standard deviation (Hedges' *g*). We have chosen Hedges' *g* because it allows for small sample size bias correction [37]. As we expected considerable heterogeneity among the studies, we used the random effects model. As a rule of thumb, effect sizes of 0.8 can be viewed as large, 0.5 as moderate, and 0.2 as small [38]. In our main analysis, we included mixed depression and/or anxiety studies into the separate depression and anxiety data sets. Statistical analysis was conducted using the Comprehensive Meta-analysis program (version 2.2.2), and pooled proportions were calculated with R [39] package *meta* [40].

To calculate heterogeneity, we used the I^2-statistic and its 95% CIs as an indicator of heterogeneity in percentages. Heterogeneity was interpreted as low, moderate, and high when 25%, 50%, and 75%, respectively.

We also included the correlations of the used pre-post measures using the mean of 0.76, where none was provided for depression and 0.59 for anxiety following the study by Balk et al [41]. We also conducted sensitivity analyses for correlations set to 0.00, 0.75, and 0.99 to examine the robustness of our findings [42]. We also calculated the prediction interval, which estimates where the true effects are to be expected for 95% of similar studies that might be conducted in the future [43].

As we expected high heterogeneity, we conducted several subgroup analyses to investigate its possible sources. The examined subgroups were related to the method of analysis, time to post assessment, recruitment pathways, disorders, guidance moment (specific timing or as a reaction), guidance modality (email, message, and synchronous), guide profession (with or without specific CBT training), supervision provided (yes/no), guide training provided (yes/no), intervention manual provided (yes/no), approach to data analysis (ITT/completer), and diagnostic method (interview/questionnaire). Subgroup analyses were only carried out with regard to the effects on symptom change. We used the mixed effects model, testing pooled studies within subgroups with random effects models while testing for significant differences between those subgroups with fixed effects models. We only conducted subgroup analysis if the number of studies per category was not less than three. If necessary, we combined predefined subgroups to achieve the necessary group size.

Finally, we conducted meta-regression analyses for the continuous variables, examining the duration of the treatment as a predictor of treatment outcome as well as guidance time, number of contacts, number of sessions completed, and the percentage of treatment completers.

Regarding uptake, we calculated the proportion of (1) included people based on the number of people screened, (2) starters based on the number of people being screened, and (3) starters based on the number of people included. Adherence was analyzed by calculating the percentage of modules completed based on the average number of sessions that were completed by the participants divided by the planned total number of sessions. We also coded the percentage of intervention completers for a 100% completion rate. Additionally, we pooled the age and gender distribution as well as participant satisfaction extracted from original studies. Furthermore, we pooled the percentage of individuals reported to show symptom deterioration (defined as a negative reliable change in the reported outcome), the deterioration rates, reported in the original study.

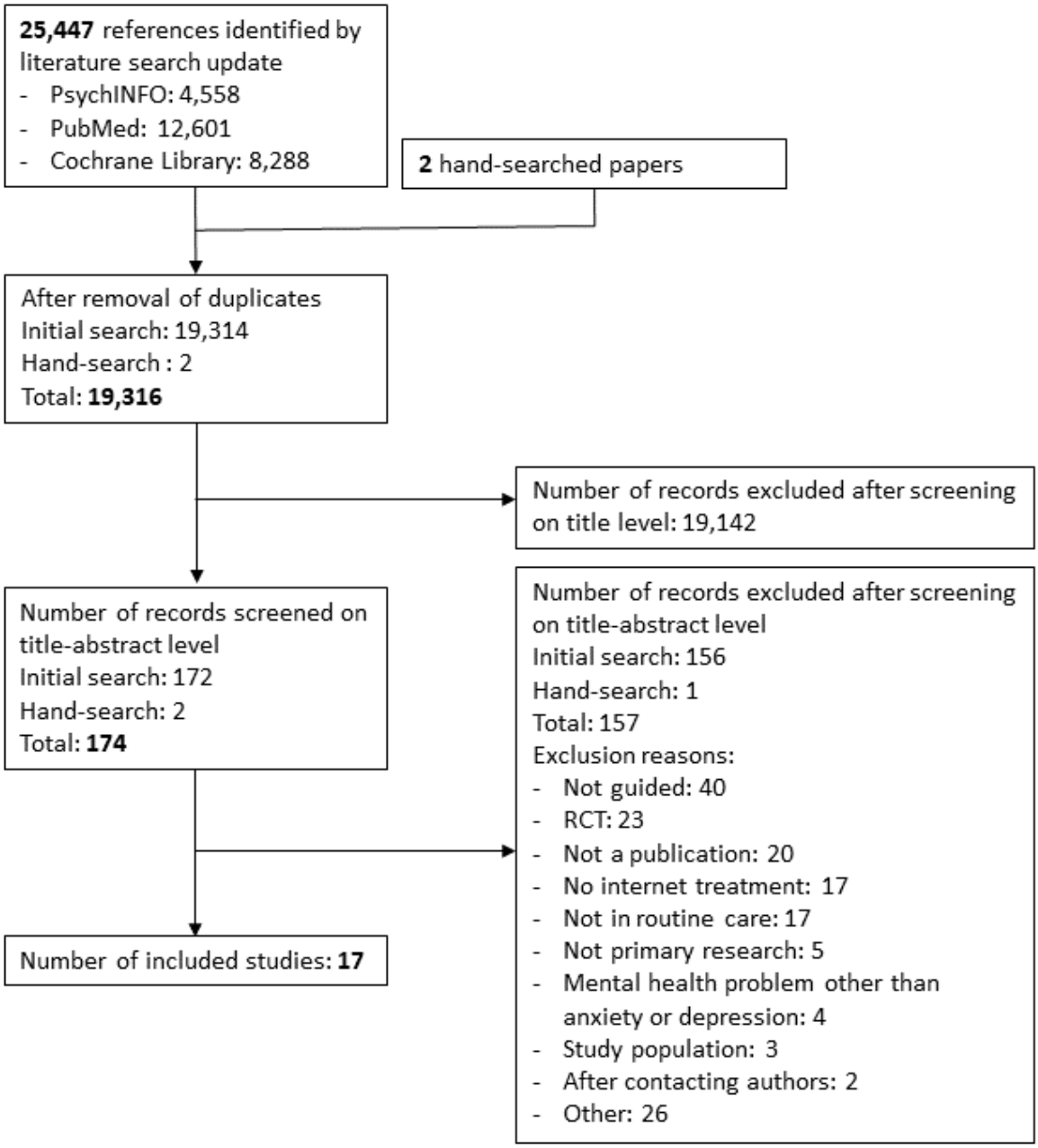
Publication bias was examined by inspecting the funnel plot [44] and conducting the Egger test of the intercept with a one-tailed significance level of α=.05 [45]. In addition, we used Duval and Tweedie's trim and fill procedure [46] to adjust the effect size for missing studies.

Results

Study Selection

A systematic literature search was performed on January 30, 2019. This search resulted in 25,447 citations. After removal of duplicates, 19,316 citations remained for the title and abstract assessment and 174 after the exclusion due to title and abstract. A total of 17 studies fulfilled the eligibility criteria. The full references of the included studies are listed in Multimedia Appendix 4 (Etzelmueller et al, unpublished data, 2020) [47-63]. The study selection process is illustrated in Figure 1.

Figure 1. Study inclusion.



Study Characteristics

Table 1 presents the characteristics of the included studies. Seventeen studies (n=12,096 participants) reporting on the outcomes of the treatment for depression and anxiety were included. Seven of the 17 studies reported multiple groups, of which 5 combined results on multiple treatments in the published study and 2 reported distinct forms of guidance within the same treatment and setting without randomizing patients on an individual level. Of the resulting 30 groups, 8 groups focused on depression, 17 on anxiety, and 5 on both depression and anxiety. We included studies reporting on both depression and anxiety in both the depression and anxiety analyses. Of the included studies, 46.7% (k=14/30, k_{Dep}=4, k_{Anx}=10) administered diagnostic interviews, 36.7% (k=11/30; k_{Dep}=6/13, k_{Anx}=7/22) self-reports, and 16.7% (k=5/30; k_{Dep}=3/13, k_{Anx}=2/22) clinical judgment in their diagnostic process. Of the included studies, 30.0% (k=9/30; k_{Dep}=4/13, k_{Anx}=5/22) administered a cutoff criterion for the self-report questionnaires. Ten studies (33.3%, k=10/30; k_{Dep}=6/13, k_{Anx}=5/22) included patients who did not meet the criteria for a clinical diagnosis of depression or anxiety. A total of 9.1% of anxiety studies (k=2/22) excluded cases with severe symptom severity; all depression studies allowed patients with severe depression severity to be included. Of the included studies, 40.0% (k=12/30; k_{Dep}=4/13, k_{Anx}=9/22) specifically stated that the patients had to be diagnosed with a clinical depression and/or anxiety disorder to follow the iCBT intervention. The rest of the studies did not specify whether the patients had clinical depression and/or anxiety. Of the included studies, 73.0% (k=22/30; k_{Anx}=9/13, k_{Dep}=18/22) stated that suicidal ideation or intent was a reason for excluding the patient from the service.

Table 1. Study characteristics.

Publication and substudy	Year of publication	Data collection[a]	Country	Sample size	Diagnosis conducted	Diagnostic criterion	Inclusion of severe cases	Exclusion: suicidal ideation[b]
Aydos et al (2009) [30]								
N/A[c]	2009	N/A	Australia	17	Interview (MINI[d])	Clinical	Yes	No
Alaoui et al (2015) [56]								
N/A	2015	2009-2013	Sweden	653	Interview (MINI)	Clinical	Yes	Yes
Etzelmueller et al (unpublished data, 2020)								
N/A	N/A	2014-2017	Germany	349	Self-report (Patient health Questionnaire; PHQ8>10)	Clinical and subclinical	No	No
Gellatly et al (2018) [57]								
N/A	2018	2013-2015	United Kingdom	724	Clinical judgment	Caseness[e]	No	No
Hadjistavropoulos et al (2014) [59]								
GAD[f]	2014	2010-2013	Canada	107	Interview (MINI) + GAD7>5	Clinical and subclinical	No	Yes
Depression	2014	2010-2013	Canada	80	Interview (MINI) + PHQ>5	Clinical and subclinical	No	Yes
Panic disorder	2014	2010-2013	Canada	25	Interview (MINI) + Panic Disorder Severity Scale-Self Report; PDSS-SR>8	Clinical and subclinical	No	Yes
Hadjistavropoulos et al (2016) [58]								
Specialized care	2016	2013-2015	Canada	260	Self-report (Anxiety and depression checklist; K10≥17)	Clinical	No	Yes
Nonspecialized care	2016	2013-2015	Canada	198	Self-report (K10≥17)	Clinical	No	Yes
Hedman et al (2013) [60]								
N/A	2013	2007-2012	Sweden	1203	Interview (MINI)	Clinical	No	No
Hedman et al (2014) [48]								
N/A	2014	2007-2013	Sweden	570	Interview (MINI)	Clinical	No	No
Marks et al (2003) [49]								
Phobia/panic	2003	N/A	United Kingdom	27	Clinical judgment (International Statistical Classification of Diseases and Related Health Problems; ICD10)	Clinical	No	Yes
Depression[g]	2003	N/A	United Kingdom	38	Clinical judgment (ICD10)	Clinical	No	Yes
Anxiety/depression	2003	N/A	United Kingdom	33	Clinical judgment (ICD10)	Clinical	No	Yes
OCD[h]	2003	N/A	United Kingdom	9	Clinical judgment (ICD10)	Clinical	No	Yes
Mathiasen et al (2018) [61]								
Depression	2018	2016-2017	Denmark	60	Interview	Clinical	No	Yes
Anxiety	2018	2016-2017	Denmark	143	Interview	Clinical	No	Yes
Morrison et al (2014) [54]								

Publication and substudy	Year of publication	Data collection[a]	Country	Sample size	Diagnosis conducted	Diagnostic criterion	Inclusion of severe cases	Exclusion: suicidal ideation[b]
N/A	2014	2012	United Kingdom	12	Self-report and clinical judgment[i]	Caseness[e]	No	No
Nordgreen et al (2018) [62,63]								
N/A	2018	2014-2016	Norway	124	Interview (MINI)	Clinical	No	Yes
Nordgreen et al (2018[b])								
N/A	2018	N/A	Norway	169	Interview (MINI)	Clinical	No	Yes
Ruwaard et al (2012) [50]								
Depression	2012	2002-2008	The Netherlands	405	Interview (N/A)	Clinical	No	Yes
Panic disorder	2012	2002-2008	The Netherlands	136	Interview (N/A)	Clinical	No	Yes
PTSD[j]	2012	2002-2008	The Netherlands	477	Interview (N/A)	Clinical	No	Yes
Shandley et al (2008) [51]								
general practitioner–guided	2008	N/A	Australia	51	Self-report and interview	Clinical	No	No
Therapist-guided	2008	N/A	Australia	41	Self-report and interview	Clinical	No	No
Titov et al (2017) [53]								
Depression	2017	2013-2016	Australia	5427	Self-report	Principal complaint	No	Yes
Depression[k]	2017	2013-2016	Australia	516	Self-report	Principal complaint	No	Yes
OCD	2017	2013-2016	Australia	69	Self-report	Principal complaint	No	Yes
PTSD	2017	2013-2016	Australia	137	Self-report	Principal complaint	No	Yes
Yu et al (2018) [52]								
N/A	2018	NA	United States	63	Self-report (GAD7≥5)	Clinical	No	Yes

[a]Data collection period.

[b]Exclusion of cases with suicidal ideation.

[c]N/A: not applicable.

[d]MINI: mini-international neuropsychiatric interview.

[e]Caseness for PHQ-9 refers to a person reporting scores of 10 on the PHQ-9.

[f]GAD: Generalized anxiety disorder.

[g]Transdiagnostic treatment for depression.

[h]OCD: obsessive-compulsive disorder.

[i]Participants were initially identified as suitable to receive a low-intensity intervention for depression or low mood through the triage of a patient's self-assessment form by team leaders, all of whom were qualified CBT therapists. Patients then had an initial assessment with a psychological well-being practitioner who considered a person's suitability for MindBalance in reference to the patient's identified difficulties, goals, and the studies' inclusion and exclusion criteria (inclusion: to receive treatment of depression with little or no comorbid anxiety, appropriate for guided self-help in a primary-care setting as determined by current [...] procedures).

[j]PTSD: posttraumatic stress disorder.

[k]Depression treatment for older adults.

iCBT Service–Related Characteristics

Of the studies, 26.3% (*k*=5/19) used transdiagnostic interventions, and all others utilized disorder-specific interventions. We did not identify any blended treatments.

Of the included studies, 31.6% (*k*=6/19) involved clinical referrals in their service pathway, 26.3% (*k*=5/19) did not involve referrals, but only included patients through the general community, whereas 42.1% (*k*=8/19) were recruited in both a community and clinical setting.

On average, iCBT treatments included 8.00 sessions (SD 2.62; *k*=26; depression: *k*=11; mean 8.09, SD 2.84; anxiety: *k*=19; mean 8.00, SD 2.81).

With regard to guidance, 46.7% of the studies (*k*=14/30; k_{Dep}=5/13, k_{Anx}=11/22) stated that guidance focused mainly on motivational and 16.7% (*k*=5/30; k_{De}=4/13, k_{Anx}=4/22) on administrative aspects. All included studies provided feedback on the content of participants who completed the sessions. Of the studies, 73.3% (*k*=22/30; k_{Dep}=10/13, k_{Anx}=14/22) used asynchronous contact methods for communication between participants and guides, 30.0% (*k*=9/30; k_{Dep}=7/13, k_{Anx}=4/22) used build-in message systems, and 16.7% (*k*=5/30; k_{Dep}=1/13, k_{Anx}=5/22) used emails. Of the studies, 23.3% (*k*=7/30; k_{Dep}=3/13, k_{Anx}=7/22) used synchronous contact via telephone contacts, of which one would also use face-to-face contacts. Of the studies, 16.7% (*k*=5/30; k_{Dep}=1/13, k_{Anx}=4/22) stated that they provided feedback as a reaction following a participant's action and 30.0% (*k*=9/30; k_{Dep}=4/13, k_{Anx}=6/22) in specific time intervals, weekly or biweekly.

Of the studies, 23.3% (*k*=7/30; k_{Dep}=4/13, k_{Anx}=6/22) only involved guides not trained in CBT, whereas the other studies included specifically trained professionals, such as psychotherapists, psychiatrists, GPs, or psychologists. Of the studies, 40.0% (*k*=12/30; k_{Dep}=7/13, k_{Anx}=10/22) stated that they provided specific training for the provision of the iCBT intervention to the guides, and 63.3% (*k*=19/30; k_{Dep}=9/13, k_{Anx}=13/22) provided supervision to the guiding participants. A total of 26.7% of the studies (*k*=8/30; k_{Dep}=4/13, k_{Anx}=4/22) reported having provided an iCBT intervention manual. The average reported guidance time was 148.50 min (SD 146.99; *k*=12; 95% CI 92.87-204.12; depression: *k*=4, mean 82.44 min (SD 81.14), 95% CI 45.29-119.60; anxiety: *k*=9; mean 157.60 min, SD 108.10; 95% CI 92.33-222.86). The pooled results are presented in Table 2.

Nine studies (47.4%; *k*=19) reported safety measures in cases of suicidality, suicidal ideation, or severe symptom deterioration. They mention monitoring systems (*k*=2), risk alerts (*k*=1), or reviewing the participants' messages (*k*=2) as ways to identify risk. Procedures were triggered in cases of suicidal ideation or suicidality (*k*=9), inactivity or lack of progress (*k*=2), or an increase in symptoms (*k*=2). Ways to mitigate the risk included contacting the participant via telephone (often in contrast to the usual messaging, *k*=4), structured risk assessments (*k*=1), referred to another service (k=6), and the development of a safety plan together with the participant (*k*=1). Information is depicted in Multimedia Appendix 5 [64-101] and Multimedia Appendix 6 [102-120] on the iCBT service–related characteristics.

Table 2. Pooled results of iCBT service– and acceptability-related outcomes: guidance time, age, gender, completed sessions, completed components, and deterioration rates.

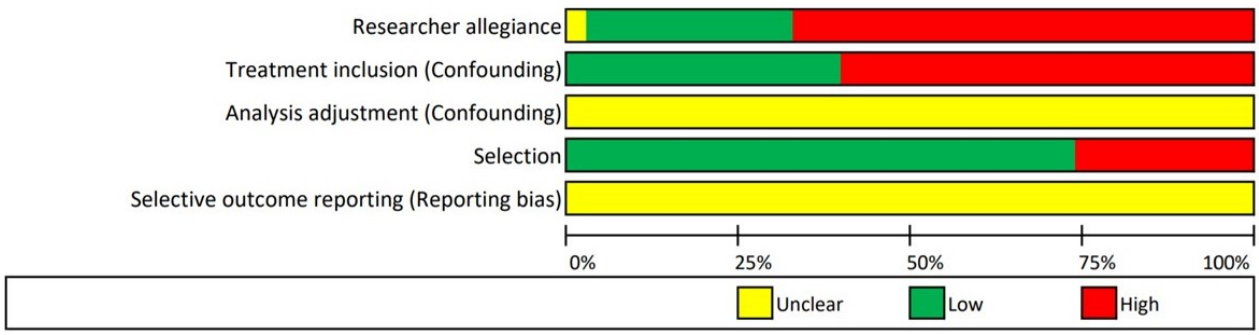
Groups	Number of studies	Pooled mean (SD)	95% CI	Range
Guidance time[a](min)				
All studies	12	148.50 (146.99)	92.9-204.1	43.0-378.6
Depression studies	4	82.44 (290.46)	45.29-119.60	43.0-183.0
Anxiety studies	9	157.60 (108.10)	92.33-222.86	43.0-378.6
Age (years)				
All studies	29	38.3 (3.02)	37.2-39.4	29.8-43.5
Depression studies	12	39.0 (2.12)	37.8-40.2	29.0-41.7
Anxiety studies	21	37.8 (3.04)	36.5-39.2	29.8-43.5
Gender, female (n, %)				
All studies	23	65.4 (20.06)	57.2-72.8	22.2-91.7
Depression studies	11	70.1 (24.37)	55.7-81.4	22.2-91.7
Anxiety studies	17	64.3 (17.25)	56.1-71.6	22.2-86.0
Average percentage of sessions completed				
All studies	14	60.6 (6.49)	57.2-72.8	16.7-90.0
Depression studies	5	62.6 (1.60)	61.2-63.9	16.7-90.0
Anxiety studies	10	57.3 (1.94)	56.1-58.4	16.7-74.3
Average percentage of participant completing all treatment components				
All studies	26	61.0 (14.83)	55.3-66.9	27.3-82.6
Depression studies	12	62.8 (13.61)	55.1-70.0	44.0-82.6
Anxiety studies	18	61.7 (17.75)	53.5-69.3	27.3-82.0
Deterioration (% deterioration in sample)				
All studies	14	2.9 (1.91)	1.9-4.3	1.0-16.6
Depression studies	5	2.5 (0.34)	2.2-2.9	1.0-12.5
Anxiety studies	9	3.1 (2.30)	1.6-5.9	1.0-16.6

[a]Excluded study Ruwaard et al [50] as outlier.

Risk of Bias of the Included Studies

The quality of the included studies varied. Of the studies, 67.0% (*k*=20/30) were rated with a high risk of bias on *Researcher Allegiance*. Of the studies, 63.0% (*k*=19/30) did not exclude patients who were participating in other psychotherapeutic treatments (*Treatment Inclusion Confounding*), and none of the studies reported on the adjustment for confounders in the data analysis. Intention-to-treat data could be extracted from 73.3% of the studies (*k*=22/30), and none of the studies were preceded by a published study protocol. The risk of bias assessment is depicted in Figure 2.

Figure 2. Risk of bias assessment.



iCBT Service Acceptability

Acceptability data on uptake, participant characteristics across studies, adherence, and participant satisfaction were pooled. All acceptability results are depicted in Multimedia Appendices 6 and 7. The pooled results are presented in Table 2.

Uptake

The average proportion of included people based on the number of people screened was 70.2% (*k*=6/30; 95% CI 8.4%-98.4%; range=0.6%-76.0%), the proportion of starters based on the number of people being screened was 48.0% (*k*=10; 95% CI 16.9%-80.8%; range=0.3%- 96.2%), and the proportion of starters based on the number of people included was 73.0% (*k*=7; 95% CI 51.0%-87.6%; *range*=40.6%- 95.9%).

Participant Characteristics

The pooled percentage of female participants was 65.4% (*k*=23, 95% CI 57.2%-72.8%; depression: *k*=11, mean 70.1%, 95% CI 55.7%-81.4%; anxiety: *k*=17, mean 64.3%, 95% CI 56.1%-71.6%). The mean age across studies was 38.30 years (*k*=29, 95% CI 37.22-39.37; depression: *k*=12, mean 38.96, 95% CI 37.77-40.15; anxiety: *k*=21, mean 37.83, 95% CI 36.47-39.20).

Adherence

The average percentage of sessions completed was 61.2% (*k*=14, 95% CI 54.9%-67.5%; depression: *k*=5, mean 62.6%, 95% CI 61.2%-63.9%; anxiety: *k*=10, mean 57.3%, 95% CI 56.1%-58.4%). The percentage of participant completing all treatment components was 61.3% (*k*=26, 95% CI 55.3%-66.9%; depression: *k*=12, mean 62.8%, 95% CI 55.1%-70.0%; anxiety: *k*=18, mean 61.7%, 95% CI 53.5%-69.3%).

Participant Satisfaction

Of the 17 studies, 10 (58.8%) reported participants' satisfaction. Participant satisfaction outcomes were reported inconsistently, using varying measures and different reporting forms. Therefore, these data could not be pooled, but the detailed results and the data extracted on patient satisfaction are depicted in Multimedia Appendix 7 [81,121-123]. Within the studies reporting participants' satisfaction, five studies reported a high and four a very high participants' satisfaction.

Effects of iCBT on Symptom Change

Depression

Effect sizes for changes in depression severity ranged from 0.66 to 1.88 (Hedges' *g*, *k*=13 studies), with 1 study (7.7%) reporting a moderate and 12 (92.3%) a large effect size.

The average pre-post effect size of all depression treatments was *g*=1.18 (95% CI 1.06-1.29), which can be considered a large effect. Heterogeneity was significant and high (I^2=95%; 95% CI 94-97; *P*<.001). The prediction interval is 0.74-1.62, and we can expect that in 95% of all populations, the true effect size will fall within this range.

The details of these results are shown in Figure 3 and Table 3.

Figure 3. Standardized Effects of iCBT treatments for depression in routine care. Full references are available in Multimedia Appendix 4. Combined: multiple measures for the main outcome have been combined in the analysis; Dep.: depression treatment; Mixed: mixed depression and anxiety treatment; NS: nonspecialized care; PHQ 8: Patient health Questionnaire – 8 Item version; PHQ 9: Patient Health Questionnaire; Plus: depression treatment for older adults; Spec.: specialized care.

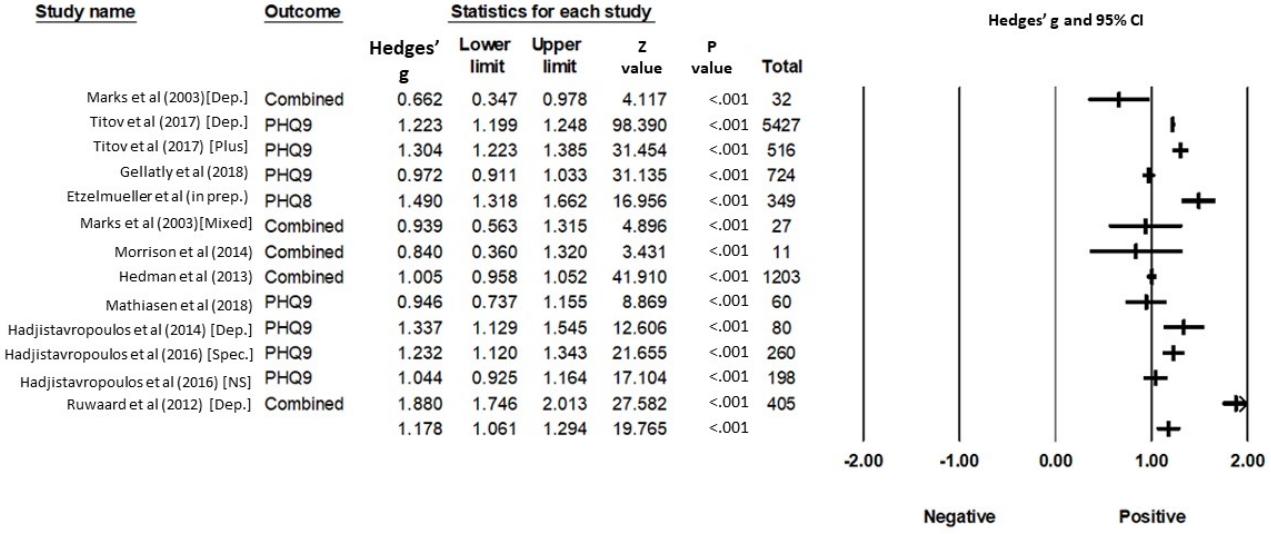


Table 3. Meta-analytic comparison of anxiety and depression interventions.

Characteristics	Effect		Heterogeneity	
	g	95% CI	*P* value	I^2 (95% CI)
Depression				
All studies (n=13)	1.178	1.06-1.29	<.001	95 (94-97)
Pre-post correlation=0.00	1.236	1.10-1.38	<.001	86 (78-91)
Pre-post correlation=0.75	1.155	1.04-1.27	<.001	96 (94-97)
Pre-post correlation=0.99	0.749	0.16-0.88	<.001	100 (100-100)
Outliers excluded[a]	1.176	1.09-1.26	.001	75 (42-86)
Without mixed treatments	1.282	1.26-1.44	<.001	89 (84-92)
Anxiety				
All studies (n=20)	0.94	0.83-1.06	<.001	74 (60-83)
Pre-post correlation=0.00	0.95	0.83-1.07	<.001	93 (91-95)
Pre-post correlation=0.75	0.93	0.82-1.04	<.001	100 (99-100)
Pre-post correlation=0.99	0.70	0.62-0.78	<.001	77 (62-86)
Outliers excluded[b]	0.90	0.81-0.99	<.001	91 (88-95)
Without mixed treatments	0.95	0.81-1.10	<.001	83 (74-89)
Without OCD[c] treatments	0.93	0.81-1.05	<.001	84 (76-90)
Without PTSD[d] treatments	0.88	0.78-0.98	<.001	95 (94-97)
Neither OCD nor PTSD	0.87	0.77-0.98	<.001	86 (78-91)

[a]Three excluded studies [47-49] as well as depression study by Ruwaard et al [50].

[b]Two excluded studies [51,52] as well as posttraumatic stress disorder (PTSD) and panic disorder studies by Ruwaard et al [50] and PTSD study by Titov et al [53].

[c]OCD: obsessive-compulsive disorder.

[d]PTSD: posttraumatic stress disorder.

In this analysis, the pre-post measurement correlation was set to the actual pre-post correlation of the measure (between 0.36 and 0.78). Sensitivity analysis, with correlations set to 0, 0.75, and 0.99, resulted in comparable effect sizes ($g_{Corr=0}$=1.24, $I^2_{Corr=0}$=86, 95% CI 78-91; *P*<.001; $g_{Corr=.75}$=1.16, $I^2_{Corr=.75}$=96, 95% CI 94-97; *P*<.001), with $g_{Corr=0.99}$=0.75 ($I^2_{Corr=.99}$=100, 95% CI 100-100; *P*<.001) resulting in the smallest effect size.

Both the visual inspection of the funnel plot and Egger test (*P*=.90) did not indicate a potential publication bias.

We found five studies to be outliers, defined as not overlapping with the 95% CI of the pooled estimate. Removing these studies [47-49], and the depression group in the study by Ruwaard et al [50], from the analysis did not result in meaningful changes in effect sizes (*g*=1.18, 95% CI 1.09-1.26), but reduced heterogeneity (I^2=75%; 95% CI 42-86; *P*<.001). Removing the mixed anxiety and depression studies did not result in a relevant change in effect size (*g*=1.28; 95% CI 1.13-1.44; I^2=97%; 95% CI 95-98; *P*<.001).

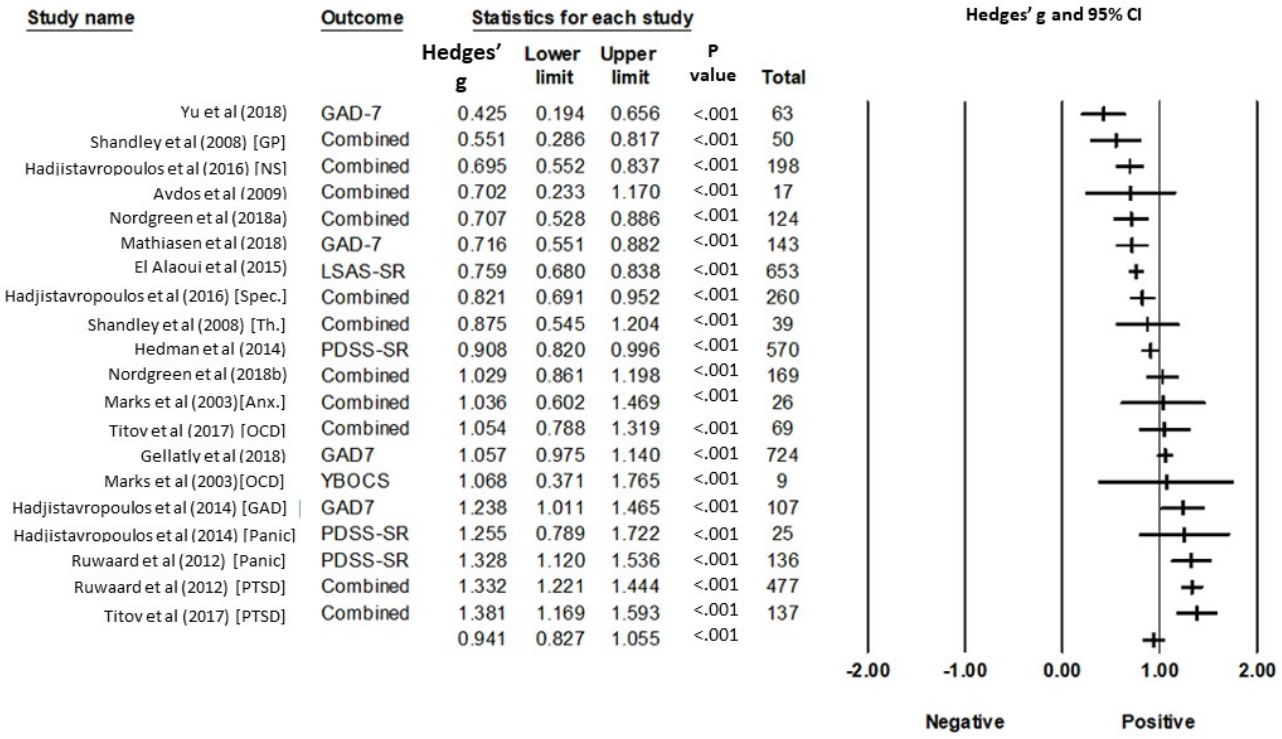
Anxiety

For the included anxiety studies (*k*=20), effect sizes ranged from 0.42 to 1.38 (*Hedges' g*), with 1 study (5.0%) reporting a small, 6 (30.0%) a moderate, and 13 (65.0%) a large effect size.

The average pre-post effect size (*Hedges' g*) of all anxiety interventions, including the interventions that targeted both anxiety and depression, was *g*=0.94 (95% CI 0.83-1.06), which is considered a large effect. Heterogeneity was high (I^2=89, 95% CI 84-92; *P*<.001). The prediction interval is 0.44-1.44, and we can expect that in 95% of all populations, the true effect size will fall within this range. The details of these results are shown in Figure 4 and Table 3.

In the main analysis described above, the pre-post measurement correlation was set to 0.59. Sensitivity analysis with correlations set to 0, 0.75, and 0.99 resulted in comparable effect sizes ($g_{Corr=0}$=0.95, $I^2_{Corr=0}$=74, 95% CI 60-83; *P*<.001; $g_{Corr=.75}$=0.93, $I^2_{Corr=.75}$=93, 95% CI 91-95; *P*<.001), with $g_{Corr=0.99}$=0.70 ($I^2_{Corr=.99}$=99, 95% CI 99-100; *P*<.001) resulting in the smallest effect size.

Figure 4. Standardized Effects of iCBT treatments for anxiety in routine care. Marks (2003) is not providing an anxiety measure for the mixed depression and anxiety treatment; therefore, this study has not been included in the analysis. Full references are available in Multimedia Appendix 4. Combined: multiple measures for the main outcome have been combined in the analysis; GAD: generalized anxiety disorder; GP: general practitioner-guided; LSAS: Liebowitz Social Anxiety Scale; NS: nonspecialized care; OCD: obsessive-compulsive disorder; PDSS-SR: Panic Disorder Severity Scale-Self Report; PTSD: posttraumatic stress disorder; Spec.: specialized care; Th.: therapist-guided; YBOCS: Yale-Brown Obsessive Compulsive Scale.



Both the visual inspection of the funnel plot and Egger test (P=.91) did not indicate a potential publication bias.

We found five studies to be outliers, as their results did not overlap with the 95% CI of the pooled estimate. Removing studies [51,52] as well as PTSD and panic disorder studies by Ruwaard et al [50] and PTSD study by Titov et al [53] from the analysis did not influence the result significantly (g=0.90; 95% CI 0.81-0.99; I^2=77%, 95% CI 62-86; p<.001), but resulted in less, although still high, heterogeneity. Excluding the mixed anxiety and depression studies did not result in a significantly different effect size (g=0.99; 95% CI 0.86-1.12; I^2=92%; 95% CI 88-95; P<.001). Neither removing OCD treatments (g=0.93; 95% CI 0.82-1.05; I^2=90%; 95% CI 86-93; P<.001), PTSD treatments (g=0.88; 95% CI 0.78-0.98; I^2=83%, 95% CI 74-89; P<.001), or both (g=0.87; 95% CI 0.77-0.98; I^2=84%; 95% CI 76-90; P<.001) resulted in significantly different effect sizes, although lowering the heterogeneity.

iCBT Negative Effects

Multimedia Appendix 7 [81,121-123] comprises the results of the negative effects. Less than half of the studies reported deterioration rates (k=7; 41%), with an average deterioration rate of 2.9% (k=14, 95% CI 1.9%-4.3%; depression: k=5, mean 2.5%, 95% CI 2.2%-2.9%; anxiety: k=9, mean 3.1%, 95% CI 1.6%-5.9%; the forest plot can be retrieved from the corresponding author). No study reported other negative effects, and one study mentioned that there were no adverse outcomes. No studies have reported the predictors of deterioration or other negative effects.

Subgroup Analysis for iCBT for the Treatment of Depression, Anxiety, or Mixed Depression and/or Anxiety

Tables 4 and 5 show the results of all examined subgroup analyses. Significant differences between subgroups were found for professional training of coaches, supervision of coaches, and treatment duration for both depression and anxiety studies and for recruitment pathways for depression studies only. Studies evaluating a period of 9 to 13 weeks of treatment duration reported a significant lower effect size (depression: g=1.00, 95% CI 0.95-1.05; I^2=0; 95% CI 0-85; anxiety: g=0.83, 95% CI 0.72-0.95; I^2=59; 95% CI 9-81) compared with studies with less than 9 (depression: g=1.17, 95% CI 1.01-1.32; I^2=95; 95% CI 92-97; anxiety: g=1.16, 95% CI 0.97-1.34; I^2=93; 95% CI 61-91) or more than 13 weeks (depression: g=1.37, 95% CI 1.00 to −1.74; I^2=97; 95% CI 94-98; anxiety: g=0.98, 95% CI 0.78-1.17; I^2=89; 95% CI 78-94). However, the effect sizes within all examined subgroups were high.

Table 4. Subgroup analyses: depression treatments.

Subgroup analysis[a]	Effects			Heterogeneity			Subgroup analysis	
	N[b]	g	95% CI	I^2	*P* value	I^2 95% CI	Q value	*P* value (Q)
Recruitment pathway[c]								
Clinical and community+clinical	8	1.05	0.95-1.14	78	<.001	56-89	7.253	.007
Community	5	1.38	1.16-1.59	96	<.001	94-98	7.253	.007
Specific treatment								
Mixed treatment	7	1.10	0.98-1.22	93	<.001	87-96	0.736	.39
Disorder-specific treatment	6	1.27	0.91-1.62	97	<.001	95-98	0.736	.39
Diagnosis conducted[d]								
Interview	7	1.127	0.89-1.35	97	<.001	95-98	0.946	.33
Questionnaire	5	1.25	1.16-1.34	81	<.001	57-92	0.946	.33
Clinical cutoff/minimal symptom severity								
Yes	4	1.27	1.08-1.45	84	<.001	60-94	0.897	.34
No	7	1.17	0.98-1.35	96	<.001	94-97	0.897	.34
Treatment duration								
<9 weeks	5	1.17	1.01-1.32	95	<.001	92-97	7.485	.02
9-13 weeks	4	1.00	0.95-1.05	0	.85	0-85	7.485	.02
>13 weeks	4	1.37	1.00-1.74	97	<.001	94-98	7.485	.02
Guide cognitive behavioral therapy training (profession)[e]								
Nonprofessional	4	0.92	0.79-1.05	22	.28	0-88	14.151	<.001
Other	9	1.27	1.14-1.40	96	<.001	94-97	14.151	<.001
Guide supervision provided								
No	4	0.91	0.75-1.08	39	.18	0-79	10.339	<.001
Yes	9	1.27	1.13-1.41	96	<.001	94-97	10.339	<.001
Guide training provided								
No	6	0.98	0.94-1.02	7	.37	0-76	21.368	<.001
Yes	7	1.35	1.20-1.51	95	<.001	91-97	21.368	<.001
Intervention manual provided								
No	9	1.039	0.95-1.137	75	<.001	51-87	10.715	<.001
Yes	4	1.467	1.23-1.71	97	<.001	95-98	10.715	<.001
Risk of bias—researcher allegiance								
High	7	1.252	1.08-1.42	97	<.001	95-98	1.347	.25
Low	5	1.119	0.99-1.29	70	.01	23-88	1.347	.25
Risk of bias—confounding (treatment inclusion)								
High	7	1.215	1.06-1.43	96	<.001	94-98	0.985	.32
Low	5	1.105	0.98-1.23	82	<.001	58-92	0.985	.32

[a]Test against "Guidance format: face-to-face vs written guidance," "Guidance modality: Message, Email, Telephone, F2F," and "Guide profession" excluded, as there were too few studies included in analysis.

[b]Number of studies.

[c]Only two studies included via the clinical pathway only. We combined the categories "Both, community and clinical" and "clinical" for this analysis.

[d]Excluding one study [54], as this is the only study using clinical judgment without specifying the use of an interview or questionnaire.

[e]We grouped all studies involving guides not specifically trained in delivering cognitive behavioral therapy in the category "non-professional" and studies involving psychiatrists, psychologists, or psychotherapists in their guidance in the category "other."

Table 5. Subgroup analyses: anxiety treatments.

Characteristics	Effect			Heterogeneity			Subgroup analysis	
	N[a]	g	95% CI	I^2	*P* value	I^2 95% CI	Q value	*P* value (Q)
Recruitment pathway								
Clinical	5	0.77	0.53-1.01	91	<.001	81-95	3.340	.19
Community	7	1.08	0.85-1.31	88	<.001	78-94	3.340	.19
Community+clinical	8	0.90	0.78-1.01	74	<.001	46-87	3.340	.19
Specific disorder								
Panic	6	0.95	0.71-1.13	91	<.001	64-92	0.053	.82
Non panic treatments	14	0.92	0.801-1.09	83	<.001	N/A[b]	0.053	.82
Guidance: modality								
Email	7	1.11	0.95-1.26	56	<.001	0-81	4.744	.09
Message	8	0.88	0.69-1.06	94	<.001	90-96	4.744	.09
Synchronous (Telephone or face-to-face	5	0.86	0.66-1.10	83	<.001	60-92	4.744	.09
Guide cognitive behavioral therapy training (profession)[c]								
Nonprofessional	4	0.87	0.47-1.27	88	<.001	73-95	0.165	.69
Other	16	0.96	0.83-1.09	90	<.001	85-93	0.165	.69
Guidance: moment								
Weekly/biweekly	10	0.66	0.73-1.00	74	<.001	0-85	0.174	.67
Reaction	4	0.83	0.72-0.94	53	<.001	50-86	0.174	.67
Guide supervision provided								
No	6	0.82	0.70-0.94	57	.04	0-83	2.812	.09
Yes	14	0.9811	0.83-1.13	90	<.001	86-94	2.812	.09
Guide training provided								
No	10	0.80	0.67-0.93	83	<.001	70-90	5.779	.02
Yes	10	1.07	0.89-1.26	90	<.001	85-94	5.779	.02
Intervention manual provided								
No	16	0.88	0.75-0.94	81	<.001	70-88	37.209	<.001
Yes	4	1.30	1.19-1.41	29	.24	0-74	37.209	<.001
Approach to data analysis								
Completer	4	1.05	0.98-1.12	0	<.001	0-77	2.796	.096
ITT	16	0.92	0.78-1.06	91	<.001	86-94	2.796	.096
Diagnostic method								
Interview	15	0.97	0.84-1.06	88	<.001	83-92	0.388	.53
Questionnaire	5	0.87	0.6-1.14	91	<.001	82-95	0.388	.53
Treatment duration								
<9 weeks	5	1.16	0.97-1.34	83	<.001	61-91	8.686	.01
9-13 weeks	8	0.83	0.72-0.95	59	.02	9-81	8.686	.01
>13 weeks	6	0.98	0.78-1.17	89	<.001	78-94	8.686	.01
Risk of bias—researcher allegiance								
High	N/A	0.99	0.83-1.14	90	<.001	0-60	1.613	.20
Low	N/A	0.82	0.63-1.02	39	<.001	66-92	1.613	.20
Risk of bias—confounding (treatment inclusion)								

Characteristics	Effect			Heterogeneity			Subgroup analysis	
	N[a]	g	95% CI	I^2	*P* value	I^2 95% CI	Q value	*P* value (Q)
High	N/A	1.03	0.89-1.18	89	<.001	83-93	4.852	.03
Low	N/A	0.82	0.70-0.93	69	<.001	37-84	4.852	.03

[a]Number of studies.

[b]N/A: not applicable.

[c]We grouped all studies involving guides not specifically trained in delivering cognitive behavioral therapy in the category "non-professional," and studies involving psychiatrists, psychologists, or psychotherapists in their guidance in the category "other."

Depression studies that recruited in community settings only reported significantly higher effect sizes (g=1.37, 95% CI 1.16-1.59; I^2=96; 95% CI 94-98), compared with studies that recruited in clinical or clinical and community settings (g=1.05, 95% CI 0.95-1.14; I^2=78; 95% CI 56-89). Across all recruitment pathways, effect sizes were large, but heterogeneity remained high. We did not find this difference in anxiety studies.

Studies only involving guides, not trained in CBT, showed a significantly lower effect size in depression studies ($g_{\text{Non-professional, Depression}}$=0.92, 95% CI 0.79-1.05; I^2=22; 95% CI 0-88) than all other studies, including specifically trained professionals ($g_{\text{Other,Depression}}$=1.27, 95% CI 1.14-1.40; I^2=96; 95% CI 94-97). We did not find this effect in anxiety studies ($g_{\text{Non-professional,Anxiety}}$=0.87, 95% CI 0.17-1.27; I^2=88; 95% CI 73-95; $g_{\text{Other,Anxiety}}$=0.96, 95% CI 0.83-1.09; I^2=90; 95% CI 85-93).

Depression studies reporting to having provided supervision to their coaches, trained their professionals, and provided an intervention manual reported a significantly higher effect size ($g_{\text{Supervision}}$=1.27, 95% CI 1.13-1.41; I^2=96; 95% CI 94-97; g_{Training}=1.35, 95% CI 1.20-1.51; I^2=95; 95% CI 91-97; g_{Manual}=1.47, 95% CI 1.23-1.71; I^2=97; 95% CI 95-98) compared with studies not reporting to provide these ($g_{\text{NoSupervision}}$=0.91, 95% CI 0.75-1.08; I^2=39; 95% CI 0-79; $g_{\text{NoTraining}}$=0.98, 95% CI 0.94-1.02; I^2=7; 95% CI 0-76; g_{NoManual}=1.04, 95% CI 0.95-1.13; I^2=75; 95% CI 51-87). For anxiety studies, we found similar effects for the reporting of training and providing an intervention manual (g_{Training}=1.07, 95% CI 0.89-1.26; I^2=90; 95% CI 85-94; g_{Manual}=1.30, 95% CI 1.19-1.41; I^2=29; 95% CI 0-74 compared with $g_{\text{NoTraining}}$=0.80, 95% CI 0.67-0.93; I^2=83; 95% CI 70-88; g_{NoManual}=0.88, 95% CI 0.75-0.94; I^2=81; 95% CI 70-88), but not for supervision.

There were no differences between subgroups regarding all other examined subgroups, both for depression and anxiety studies.

Subgroup analyses comparing studies rated with high versus low risk indicated that R*esearcher Allegiance* did not have a significant influence on the estimated effect sizes for neither anxiety nor depression studies. The heterogeneity within the studies reporting a low risk of bias on *Researcher Allegiance* did reveal an I^2 of 39 compared with an I^2 of 90 for studies reporting a high risk of bias. Moreover, anxiety studies rated as at high risk of *Treatment Inclusion Confounding* had higher estimated effect sizes. This was not replicated in subgroup analyses of interventions targeting depression. Anxiety studies at high risk of *Selection Bias* reported significantly lower effect sizes. Similar outcomes were not replicated in the depression trials.

Meta-Regression Analysis for iCBT for the Treatment of Anxiety, Depression, or Mixed Depression and Anxiety

Meta-regression analyses indicated that longer treatment duration in depression studies was positively associated with a higher effect (P=.02; β=0.03, R^2=0.00). This effect was not found in anxiety studies (P=.94). None of the examined variables, that is, guidance time, number of contacts, number of sessions completed, or the percentage of treatment completers, were significantly associated with the observed effect sizes, neither in depression nor anxiety studies.

Discussion

This study aims to examine the acceptability, effects on symptom change, and negative effects of guided iCBT interventions in treating depression and anxiety in routine care. Regarding the uptake of the service, on average, 70.2% of people screened were not offered inclusion, and of those included, 73.0% started the intervention. The vast majority of participants reached were female, with an average age of 38.3 years, and 61.3% of participants completed the interventions as planned. Reported participant satisfaction was high, although inconsistently reported results did not allow us to pool effects. The average professional guidance time per participant was 133.49 min over the treatment duration. With regard to the effects on symptom change, the results indicated large average reductions for both depression (g=1.18; 95% CI 1.06-1.29) and anxiety (g=0.94; 95% CI 0.83-1.062). However, the heterogeneity between studies was high. Nevertheless, all examined effect sizes were at least moderate, indicating the intervention's potential when delivered under routine care conditions with effects ranging from moderate to large. The average deterioration rates were 3.2% for depression and 3.1% for anxiety. Subgroup analyses indicated a range of iCBT service–related characteristics to be associated with the observed treatment effects.

Regarding uptake, we found that many participants who were in contact with the iCBT service did not start the intervention. Pretreatment dropout is hard to assess, and, accordingly, reasons

for not starting an iCBT intervention after inclusion have not been discussed in the original publications.

The average age of participants found in this study (mean 38.30) appears to be slightly lower than that reported in RCTs on guided iCBT interventions for the treatment of depression (mean 42.5 [124]) but comparable with reports on the mean age of participants within guided iCBT interventions for the treatment of anxiety [125]. The percentage of females in the routine care study population was higher for depression studies compared with guided iCBT for the treatment of depression [124] and similar to reports on participants in guided iCBT interventions for the treatment of anxiety [125] in experimental settings. As similar distributions between female and male users are reported in face-to-face mental health service utilization [126], this effect might be explained by gender differences in help-seeking behavior than being related to iCBT service–related factors [127] as well as by gender differences in the prevalence of depression and anxiety disorder [128,129]. Future studies should focus on ways to attract men to use iCBT interventions.

The pooled reported percentage of sessions completed, that is, 62.6% in depression and 57.3% in anxiety studies, was lower than that described in meta-analyses on adherence in RCTs on iCBT interventions. Comparing the adherence to iCBT and face-to-face CBT, van Ballegooijen et al [130] reported that on average, participants completed 80.8% of treatment sessions in the iCBT and 83.9% in the face-to-face intervention [130]. Similarly, the percentage of participants completing the treatment as planned was lower (62.8% for depression and 61.7% for anxiety studies) than reported elsewhere [130,131]. These differences might be due to the assumed adherence-fostering effect of randomized controlled settings versus routine care [132]. However, completion rates were reported inconsistently across studies, applying different criteria such as study or treatment completers, including several definitions of treatment completions. To facilitate comparability, literature on iCBT completion should settle on one reporting standard. Further investigation of factors promoting the acceptance of iCBT interventions, also when reporting on effectiveness results in routine care, may lead to a deeper understanding that might foster intervention development and upscaling.

Results on the effectiveness of iCBT ($g_{Depression}$=1.18, 95% CI 1.06-1.29 and $g_{Anxiety}$=0.94, 95% CI 0.83-1.062) confirm findings of recently published systematic reviews and meta-analyses on RCTs of iCBT for depression and anxiety. Königbauer et al [12] found medium to large pre-post within-group effects ranging between −0.64 and −2.24 for interventions treating clinical depression [12]. To our knowledge, no recent meta-analysis has reported on pre-post effect sizes of studies targeting guided iCBT interventions for the treatment of anxiety. On an individual study level, pre-post effects in randomized trials ranged from 0.54 to 2.40 (please see Multimedia Appendix 8 for references) [133-162] compared with 0.66 to 1.88 in depression and 0.42 to 1.38 (*Hedges' g*) in anxiety within this analysis.

With regard to randomized pragmatic trials conducted under routine care conditions, Andrews et al [15] examined a sample of 64 papers reporting results of RCTs on the effectiveness of iCBT for the treatment of depression, panic disorder, generalized anxiety disorder, and social phobia in comparison with control groups in routine practice. This review study reported effect sizes for depression, panic disorder, generalized anxiety disorder, and social phobia ranging from g=0.67 to 1.31 [15]. The same study identified eight papers investigating the effectiveness of iCBT, reporting an average effect size of g=1.07 across the treatment of depression, panic disorder, generalized anxiety disorder, and social phobia [15]. The between-group effects were moderate to large (g=0.72; 95% CI 0.60-0.83; P<.001; of I^2=53, 95% CI 31-66) in the most recent meta-analysis of iCBT treatments for anxiety compared with control conditions in reducing symptoms of anxiety in an adult population [13]. Additionally, the results of this study are in line with meta-analytic findings on face-to-face CBT treatments implemented in routine care with pre-post effect size found in randomized trials ranging from d=0.69 to 2.28 for depression [28] and g=0.73 to 2.59 for anxiety treatments [163].

The results of deterioration rates (3.2% in depression and 3.1% in anxiety studies) were slightly lower, but within the 95% CI of findings based on RCTs for internet-based guided self-help interventions (3.36%) for depression [164] and anxiety (5.8% [165]), and also comparable with deterioration rates in face-to-face psychotherapy for depression [166]. Criteria defining deterioration varied between studies, and unfortunately, neither were reports on other negative effects included in most primary studies nor reported any study predictors of deterioration. This seems of utmost importance to identify those individuals that should potentially be referred to other mental health services. Their investigation is of specific importance within naturalistic study designs and under routine care conditions [164,165,167].

Most evaluated iCBT services for depression (69.2%) excluded severe cases and individuals with suicidal ideation (k=9/13) at baseline. However, a large-scale study showed that iCBT services can also result in positive effects on suicidality, reducing the prevalence of suicidal ideation from 50% at baseline to 27% after treatment [168]. In addition, a recent individual patient data meta-analysis on RCTs indicated that guided iCBT also resulted in clinically meaningful results in individuals with severe depression symptomatology [124]. Given that many individuals applying to iCBT services either do not have access to other immediate care or are not willing to utilize alternative treatment services, future studies should explore the balance between potential risk and benefits of opening up those services to populations showing elevated suicidal ideation. In such cases, it seems of utmost importance to monitor potential upcoming crises using standard operating procedures involving trained clinicians and to evaluate treatment success at the end of the service. In case of nonresponse, individuals should be motivated and guided to utilize other mental health care services, if available. Such standardized crisis procedures were only reported to be employed by less than half of the studies included in this review. iCBT services in routine care might profit from clear pathways of referral to other services in cases of nonresponse and symptom deterioration. Furthermore, future research should facilitate our understanding

of the effects of routine outcome monitoring in routinely applied iCBT [169], as this monitoring could help evaluate participants' progress throughout the course of treatment, using standardized outcome measures to elicit clients as part of a measurement-based care delivery approach in routine mental and behavioral health care [170,171].

The finding that treatment outcomes of depression interventions were greater when recruitment was carried out using an open recruitment strategy in a community setting compared with when recruited in a clinical setting is in line with the findings of Romijn et al [13] with regard to randomized pragmatic studies on anxiety disorder treatments. However, in our study, this interaction was only found for depression and could not be confirmed for anxiety disorders. One potential explanation for the difference in effects might be differences in the characteristics of the included patients. There is evidence that iCBT recruiting via open recruitment strategies, such as through web-based channels, might only reach a specific population that is different from those seeking help in a clinical setting [19]. It is often argued that internet interventions might reach individuals that would otherwise not seek treatment or only at a later time point. Given that, for example, the chronicity of depression is associated with worse treatment outcomes [172], the difference in effect might be explained by reaching a population with lower chronicity. However, such an assumption needs to be confirmed in future studies.

Further subgroup analyses indicated that iCBT services for the treatment of depression utilize trained professionals (psychotherapists and psychiatrists) to result in larger pre-post changes compared with iCBT services that used only nonprofessionals not trained in CBT (psychologists without specialized CBT training, nurses, GPs, counselors, coaches, and lived experience coordinators). However, we did not find this effect in the anxiety studies. Moreover, effects in the subgroup of depression studies involving nonprofessionals were large, indicating the potential to deliver iCBT services, for example, in contexts when there might be a shortage of trained clinicians. In cases where nonprofessionals deliver guidance in iCBT services, supervision by trained clinicians, including the availability of professionals for crisis intervention, seems warranted. Further subgroup analyses also indicated that providing supervision to coaches is also associated with higher average treatment effects for depression, but not for anxiety studies. Furthermore, training the professional and providing an intervention manual is positively related to the interventions' effectiveness. This result must be interpreted with caution as we coded all studies not mentioning supervision, training, or manual provision in their publication as not providing these components. Furthermore, these components do not inform us about actual treatment fidelity. Further research should focus on the effects on treatment outcomes of providing supervision, training, and intervention manuals to professionals working with iCBT interventions in routine care as well as the assessment of treatment fidelity.

Moreover, we did not find a difference in effects on mean symptom change between iCBT services who applied diagnostic interviews for patient allocation versus those that used self-reports only. This is in line with meta-analytic findings from RCTs on guided digital interventions for depression [124] and with studies directly comparing the effectiveness of iCBT services when treatment allocation was based on an automatic web-based assessment versus clinician assessment [173]. This indicates that such services can be used in contexts when implementing services with initial clinician assessment is not possible, without affecting average treatment success. However, it must be noted that although results might not indicate differences on the group level, it might be the case that using web-based assessments only, without a clinical assessment, will overlook relevant diagnostic information that requires immediate attention, such as suicidal risk or an underlying treatment need for comorbid disorders such as PTSD on an individual level.

The strengths of this meta-analysis include the exclusive focus on evaluating iCBT interventions for their acceptability and clinical outcomes under real-world conditions. Unlike previous systematic reviews that mixed efficacy with effectiveness trials, in this review, we focused only on studies conducted in regular care settings. This is important as we strive to report routine care results free from biases possibly being introduced within efficacy studies such as stricter application of protocolized procedures, eligibility criteria, and randomization [19-22]. Moreover, we presented an overview of implementation indicators existing in the included studies that can be used to gain a better understanding of how iCBT can be adopted by regular care services. Nevertheless, the findings of this study should be interpreted with caution due to several limitations.

First, the heterogeneity in our sample was high and significant, illustrating a great variation in the results of the included studies. Thus, we cannot draw firm conclusions regarding the average effect of iCBT in routine care. Moreover, within-group effect sizes do not depict an optimal estimator for the treatment effect because they are not independent of each other and do not account for recovery occurring independent of the treatment, thereby leading to an overestimation of the treatment effect [42]. However, in comparison with and on the basis of the reported efficacy of iCBT interventions established in RCTs, they depict the best available indicator of the effects of iCBT solutions in a routine care environment. Furthermore, we found that treatment duration had a significant influence on treatment effects. This result also supports the hypothesis that findings on pre-post changes in symptom severity might have been influenced by spontaneous or unexplained recovery, which is a common factor in depression [174]. However, our main results are in line with within-group effect sizes found in RCTs, where spontaneous recovery also occurs, and we, therefore, conclude that our effects can be considered substantial. Although heterogeneity was not explained by any other of the examined subgroups, several assumptions can be made regarding its sources. One other explanation for the high heterogeneity might be the influence of contextual factors of observational studies, such as sampling methods, participant characteristics, within-group effect sizes, and differences between the studies in reporting outcomes. It can be hypothesized that a greater harmonization regarding the conduct and reporting of effectiveness studies in routine care could lead to greater comparability of the studies' results. Another reason for the observed heterogeneity might be the different contexts of regular

care facilities across different countries. There is great variability in the degree of e-mental health penetration in different countries. For instance, Australia is considered one of the frontrunners in the e-mental health field, whereas Norway adopted these interventions very recently [175]. Thus, professionals might differ in the way they interact with e-mental health around the world. Finally, the interventions might differ in the way that they have been developed. These results also imply the importance of establishing a firm evidence base for individual iCBT interventions before their larger upscale.

Second, firm conclusions on treatment effects might be biased by studies that also included participants who could also participate in other psychotherapeutic treatments. Meanwhile, the data do not allow conclusions on the percentage of participants receiving additional treatment and represents the routine practice. Additionally, no study has reported adjusting for confounders such as baseline symptom severity, treatment fidelity (provision and use), or changes in the treatment over the course of the studies, which should be considered in future reports on the effects of iCBT in routine care.

Future studies should add to the body of literature on iCBT interventions examined under routine care conditions. Additionally, these studies should not solely focus on the effectiveness of the interventions, but if possible, it would be helpful if they also reported on specific service-, implementation-, and context-related outcomes. One way of achieving this might be through taxonomy and guidelines for the reporting of iCBT effectiveness, implementation, and context outcomes in routine care. In contrast to standards of reporting RCTs, no such international standards exist when it comes to reporting nonrandomized intervention studies. Proctor et al [176] suggested a list of outcomes for implementation-related research, and Hermes et al [177] recently made suggestions on how to build upon these ideas to establish a measurement system for the implementation of behavioral intervention technologies. Moreover, such research should always be discussed and evaluated in the light of the quality criteria established to help all involved stakeholders, patients, practitioners, and decision makers at the local and policy level to identify not only effective but also safe interventions [178].

In conclusion, this study provides further evidence supporting the acceptability and effectiveness of guided iCBT for the treatment of depression and anxiety when implemented in routine care, whereas results on negative effects are less clear. Guided iCBT may be an effective way of overcoming barriers to treatment provision. It may substantially increase the coverage of usual care services and offer an innovative treatment format for the treatment of depression and anxiety.

Acknowledgments

The authors wish to acknowledge Silja Schenk's assistance in managing the literature search in an early stage of the project. The authors also want to thank Dr Robin Kok for providing his knowledge on his website.

Authors' Contributions

AE and DE initiated and conceptualized the study and drafted the manuscript with initial feedback from all authors. AE, CV, and DE acquired and managed the data. AE, CV, and EK conducted the data analysis, with critical feedback from DE and PC. All authors revised and edited the manuscript, provided final approval for this version, and agreed to be accountable for this work.

Conflicts of Interest

Associate Prof DE reports to have received consultancy fees or served in the scientific advisory board for several companies such as Minddistrict, Sanofi, Lantern, and Schön Kliniken; German health insurance companies (BARMER and Techniker Krankenkasse); and chambers of psychotherapists. Dr DE is one of the stakeholders of the Institute for Health Training Online (GET.ON), which aims to implement scientific findings related to digital health interventions into routine care. AE is employed by the Institute for Health Training Online (GET.ON) as a research coordinator. Prof. NT is funded by the Australian Government to develop and provide a free national online and telephone-delivered treatment service. Prof. HB served in the e-mental health–associated scientific advisory boards, e-mental health interest groups, and task forces. All other authors do not report any conflicts of interest.

Multimedia Appendix 1
Prisma Checklist.
[DOCX File , 18 KB - jmir_v22i8e18100_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 13 KB - jmir_v22i8e18100_app2.docx]

Multimedia Appendix 3
Risk of Bias Assessment Definition.
[DOCX File , 16 KB - jmir_v22i8e18100_app3.docx]

Multimedia Appendix 4
References of included studies.
[DOCX File , 16 KB - jmir_v22i8e18100_app4.docx]

Multimedia Appendix 5
E and F - iCBT service-related characteristics.
[DOCX File , 67 KB - jmir_v22i8e18100_app5.docx]

Multimedia Appendix 6
Uptake and participant characteristics.
[DOCX File , 53 KB - jmir_v22i8e18100_app6.docx]

Multimedia Appendix 7
Acceptability - Participant satisfaction and negative effects.
[DOCX File , 46 KB - jmir_v22i8e18100_app7.docx]

Multimedia Appendix 8
References for studies targeting guided iCBT interventions for the treatment of anxiety reporting on pre-post effect sizes.
[DOCX File , 17 KB - jmir_v22i8e18100_app8.docx]

References

1. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health 2013;34:119-138 [FREE Full text] [doi: 10.1146/annurev-publhealth-031912-114409] [Medline: 23514317]
2. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the national comorbidity survey replication (NCS-R). J Am Med Assoc 2003 Jun 18;289(23):3095-3105. [doi: 10.1001/jama.289.23.3095] [Medline: 12813115]
3. Donohue JM, Pincus HA. Reducing the societal burden of depression: a review of economic costs, quality of care and effects of treatment. Pharmacoeconomics 2007;25(1):7-24. [doi: 10.2165/00019053-200725010-00003] [Medline: 17192115]
4. Whiteford H, Degenhardt L, Rehm J, Baxter A, Ferrari A, Erskine H, et al. Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. Lancet 2013 Nov 9;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]
5. Cuijpers P, Karyotaki E, de Wit L, Ebert DD. The effects of fifteen evidence-supported therapies for adult depression: a meta-analytic review. Psychother Res 2020 Mar;30(3):279-293. [doi: 10.1080/10503307.2019.1649732] [Medline: 31394976]
6. Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jönsson B, et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 2011 Sep;21(9):655-679. [doi: 10.1016/j.euroneuro.2011.07.018] [Medline: 21896369]
7. Mack S, Jacobi F, Gerschler A, Strehle J, Höfler M, Busch MA, et al. Self-reported utilization of mental health services in the adult German population--evidence for unmet needs? Results of the DEGS1-mental health module (DEGS1-MH). Int J Methods Psychiatr Res 2014 Sep;23(3):289-303 [FREE Full text] [doi: 10.1002/mpr.1438] [Medline: 24687693]
8. Ebert DD, Nobis S, Lehr D, Baumeister H, Riper H, Auerbach RP, et al. The 6-month effectiveness of Internet-based guided self-help for depression in adults with type 1 and 2 diabetes mellitus. Diabet Med 2017 Jan;34(1):99-107. [doi: 10.1111/dme.13173] [Medline: 27334444]
9. Ebert D, van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health. Eur Psychol 2018 May;23(2):167-187. [doi: 10.1027/1016-9040/a000318]
10. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
11. Sander L, Rausch L, Baumeister H. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research and effectiveness of internet-based interventions for the prevention of mental disorders: a systematic review and meta-analysis. JMIR Ment Health 2016 Aug 17;3(3):e38 [FREE Full text] [doi: 10.2196/mental.6061] [Medline: 27535468]
12. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. J Affect Disord 2017 Dec 1;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
13. Romijn G, Batelaan N, Kok R, Koning J, van Balkom A, Titov N, et al. Internet-delivered cognitive behavioral therapy for anxiety disorders in open community versus clinical service recruitment: meta-analysis. J Med Internet Res 2019 Apr 17;21(4):e11706 [FREE Full text] [doi: 10.2196/11706] [Medline: 30994462]

14. Andersson G, Carlbring P, Titov N, Lindefors N. Internet interventions for adults with anxiety and mood disorders: a narrative umbrella review of recent meta-analyses. Can J Psychiatry 2019 Jul;64(7):465-470 [FREE Full text] [doi: 10.1177/0706743719839381] [Medline: 31096757]
15. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
16. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
17. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa T. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry 2019 Jul 1;76(7):700-707 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.0268] [Medline: 30994877]
18. Wells K. Treatment research at the crossroads: the scientific interface of clinical trials and effectiveness research. Am J Psychiatry 1999 Jan;156(1):5-10. [doi: 10.1176/ajp.156.1.5] [Medline: 9892291]
19. Rothwell PM. External validity of randomised controlled trials: 'to whom do the results of this trial apply?'. Lancet 2005;365(9453):82-93. [doi: 10.1016/S0140-6736(04)17670-8] [Medline: 15639683]
20. Singal AG, Higgins PD, Waljee AK. A primer on effectiveness and efficacy trials. Clin Transl Gastroenterol 2014 Jan 2;5:e45 [FREE Full text] [doi: 10.1038/ctg.2013.13] [Medline: 24384867]
21. Ebert DD, Baumeister H. Internet-based self-help interventions for depression in routine care. JAMA Psychiatry 2017 Aug 1;74(8):852-853. [doi: 10.1001/jamapsychiatry.2017.1394] [Medline: 28636717]
22. van der Lem R, van der Wee NJ, van Veen T, Zitman FG. Efficacy versus effectiveness: a direct comparison of the outcome of treatment for mild to moderate depression in randomized controlled trials and daily practice. Psychother Psychosom 2012;81(4):226-234. [doi: 10.1159/000330890] [Medline: 22584117]
23. Vlahakes GJ. The value of phase 4 clinical testing. N Engl J Med 2006 Jan 26;354(4):413-415 [FREE Full text] [doi: 10.1056/nejme058272]
24. Hill T. Conducting phase IV clinical studies: a moral imperative? Ecancermedicalscience 2012;6:276. [doi: 10.3332/ecancer.2012.276]
25. Andersson G, Hedman E. Effectiveness of guided internet-based cognitive behavior therapy in regular clinical settings. Verhaltenstherapie 2013;23(3):140-148. [doi: 10.1159/000354779] [Medline: 2013612983]
26. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
27. Erbe D, Eichert H, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
28. Hans E, Hiller W. Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: a meta-analysis of nonrandomized effectiveness studies. J Consult Clin Psychol 2013 Feb;81(1):75-88. [doi: 10.1037/a0031080] [Medline: 23379264]
29. Zarski A, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to internet-based mobile-supported stress management: a pooled analysis of individual participant data from three randomized controlled trials. J Med Internet Res 2016 Jun 29;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
30. Glossary. National Institute for Health Research. URL: https://www.nihr.ac.uk/glossary?letter=F&postcategory=-1 [accessed 2020-06-24]
31. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. BMC Med Res Methodol 2010 Jul 16;10:67 [FREE Full text] [doi: 10.1186/1471-2288-10-67] [Medline: 20637084]
32. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. New York, USA: Wiley; 2006.
33. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Br Med J 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
34. Higgins JP, Altman DG. Assessing risk of bias in included studies. In: Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: John Wiley & Sons; 2008.
35. Leykin Y, DeRubeis RJ. Allegiance in psychotherapy outcome research: separating association from bias. Clin Psychol (New York) 2009;16(1):54-65. [doi: 10.1111/j.1468-2850.2009.01143.x]
36. Munder T, Brütsch O, Leonhart R, Gerger H, Barth J. Researcher allegiance in psychotherapy outcome research: an overview of reviews. Clin Psychol Rev 2013 Jun;33(4):501-511. [doi: 10.1016/j.cpr.2013.02.002] [Medline: 23500154]
37. Hedges L, Olkin I. Statistical Methods Meta-Analysis. New York, USA: Academic Press; 2014.
38. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, USA: Behavioral Sciences; 1988.
39. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Core Team; 2019.

40. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019 Nov;22(4):153-160. [doi: 10.1136/ebmental-2019-300117] [Medline: 31563865]
41. Balk EM, Earley A, Patel K, Trikalinos TA, Dahabreh IJ. Empirical assessment of within-arm correlation imputation in trials of continuous outcomes. Methods Res Reports 2012;12:EHC141. [Medline: 23326900]
42. Cuijpers P, Weitz E, Cristea IA, Twisk J. Pre-post effect sizes should be avoided in meta-analyses. Epidemiol Psychiatr Sci 2017 Aug;26(4):364-368 [FREE Full text] [doi: 10.1017/S2045796016000809] [Medline: 27790968]
43. Borenstein M, Higgins J, Hedges LV, Rothstein HR. Basics of meta-analysis: I is not an absolute measure of heterogeneity. Res Synth Methods 2017 Mar;8(1):5-18. [doi: 10.1002/jrsm.1230] [Medline: 28058794]
44. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. Br Med J 2011 Jul 22;343:d4002. [doi: 10.1136/bmj.d4002] [Medline: 21784880]
45. Sterne J, Egger M. Regression methods to detect publication and other bias in meta-analysis. In: Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. New York, USA: John Wiley & Sons; 2006.
46. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
47. Gellatly J, Chisnall L, Seccombe N, Ragan K, Lidbetter N, Cavanagh K. @Home etherapy service for people with common mental health problems: an evaluation. Behav Cogn Psychother 2018 Jan;46(1):115-120. [doi: 10.1017/S1352465817000297] [Medline: 28506333]
48. Hedman E, Ljótsson B, Kaldo V, Hesser H, El Alaoui S, Kraepelien M, et al. Effectiveness of internet-based cognitive behaviour therapy for depression in routine psychiatric care. J Affect Disord 2014 Feb;155:49-58. [doi: 10.1016/j.jad.2013.10.023] [Medline: 24238951]
49. Marks IM, Mataix-Cols D, Kenwright M, Cameron R, Hirsch S, Gega L. Pragmatic evaluation of computer-aided self-help for anxiety and depression. Br J Psychiatry 2003 Jul;183:57-65. [doi: 10.1192/bjp.183.1.57] [Medline: 12835245]
50. Ruwaard J, Lange A, Schrieken B, Dolan CV, Emmelkamp P. The effectiveness of online cognitive behavioral treatment in routine clinical practice. PLoS One 2012;7(7):e40089 [FREE Full text] [doi: 10.1371/journal.pone.0040089] [Medline: 22792217]
51. Shandley K, Austin DW, Klein B, Pier C, Schattner P, Pierce D, et al. Therapist-assisted, Internet-based treatment for panic disorder: can general practitioners achieve comparable patient outcomes to psychologists? J Med Internet Res 2008 May 19;10(2):e14 [FREE Full text] [doi: 10.2196/jmir.1033] [Medline: 18487138]
52. Yu JS, Szigethy E, Wallace M, Solano F, Oser M. Implementation of a guided, digital cognitive behavioral program for anxiety in primary care: preliminary findings of engagement and effectiveness. Telemed J E Health 2018 Nov;24(11):870-878. [doi: 10.1089/tmj.2017.0280] [Medline: 29480752]
53. Titov N, Dear B, Staples L, Bennett-Levy J, Klein B, Rapee R, et al. The first 30 months of the mindspot clinic: evaluation of a national e-mental health service against project objectives. Aust N Z J Psychiatry 2017 Dec;51(12):1227-1239. [doi: 10.1177/0004867416671598] [Medline: 27733709]
54. Morrison C, Walker G, Ruggeri K, Hacker Hughes J. An implementation pilot of the MindBalance web-based intervention for depression in three IAPT services. Cogn Behav Therap 2014 Nov 14;7:e15. [doi: 10.1017/s1754470x14000221]
55. Aydos L, Titov N, Andrews G. Shyness 5: the clinical effectiveness of internet-based clinician-assisted treatment of social phobia. Australas Psychiatry 2009 Dec;17(6):488-492. [doi: 10.1080/10398560903284943] [Medline: 20001373]
56. El Alaoui S, Hedman E, Kaldo V, Hesser H, Kraepelien M, Andersson E, et al. Effectiveness of internet-based cognitive-behavior therapy for social anxiety disorder in clinical psychiatry. J Consult Clin Psychol 2015 Oct;83(5):902-914. [doi: 10.1037/a0039198] [Medline: 26009780]
57. Gellatly J, Chisnall L, Seccombe N, Ragan K, Lidbetter N, Cavanagh K. Home etherapy service for people with common mental health problems: an evaluation. Behav Cogn Psychother 2018 Jan;46(1):115-120. [doi: 10.1017/S1352465817000297] [Medline: 28506333]
58. Hadjistavropoulos H, Nugent M, Alberts N, Staples L, Dear B, Titov N. Transdiagnostic internet-delivered cognitive behaviour therapy in Canada: an open trial comparing results of a specialized online clinic and nonspecialized community clinics. J Anxiety Disord 2016 Aug;42:19-29 [FREE Full text] [doi: 10.1016/j.janxdis.2016.05.006] [Medline: 27244278]
59. Hadjistavropoulos H, Pugh N, Nugent M, Hesser H, Andersson G, Ivanov M, et al. Therapist-assisted internet-delivered cognitive behavior therapy for depression and anxiety: translating evidence into clinical practice. J Anxiety Disord 2014 Dec;28(8):884-893 [FREE Full text] [doi: 10.1016/j.janxdis.2014.09.018] [Medline: 25445078]
60. Hedman E, Ljótsson B, Rück C, Bergström J, Andersson G, Kaldo V, et al. Effectiveness of internet-based cognitive behaviour therapy for panic disorder in routine psychiatric care. Acta Psychiatr Scand 2013 Dec;128(6):457-467. [doi: 10.1111/acps.12079] [Medline: 23406572]
61. Mathiasen K, Riper H, Andersen TE, Roessler KK. Guided internet-based cognitive behavioral therapy for adult depression and anxiety in routine secondary care: observational study. J Med Internet Res 2018 Nov 28;20(11):e10927 [FREE Full text] [doi: 10.2196/10927] [Medline: 30487118]

62. Nordgreen T, Gjestad R, Andersson G, Carlbring P, Havik O. The effectiveness of guided internet-based cognitive behavioral therapy for social anxiety disorder in a routine care setting. Internet Interv 2018 Sep;13:24-29 [FREE Full text] [doi: 10.1016/j.invent.2018.05.003] [Medline: 30206515]
63. Nordgreen T, Gjestad R, Andersson G, Carlbring P, Havik OE. The implementation of guided Internet-based cognitive behaviour therapy for panic disorder in a routine-care setting: effectiveness and implementation efforts. Cogn Behav Ther 2018 Jan;47(1):62-75. [doi: 10.1080/16506073.2017.1348389] [Medline: 28714775]
64. Titov N, Andrews G, Choi I, Schwencke G, Mahoney A. Shyness 3: randomized controlled trial of guided versus unguided internet-based CBT for social phobia. Aust N Z J Psychiatry 2008 Dec;42(12):1030-1040. [doi: 10.1080/00048670802512107] [Medline: 19016091]
65. Titov N, Andrews G, Schwencke G. Shyness 2: treating social phobia online: replication and extension. Aust N Z J Psychiatry 2008 Jul;42(7):595-605. [doi: 10.1080/00048670802119820] [Medline: 18612863]
66. Titov N, Andrews G, Schwencke G, Drobny J, Einstein D. Shyness 1: distance treatment of social phobia over the Internet. Aust N Z J Psychiatry 2008 Jul;42(7):585-594. [doi: 10.1080/00048670802119762] [Medline: 18612862]
67. Furmark T, Carlbring P, Hedman E, Sonnenstein A, Clevberger P, Bohman B, et al. Guided and unguided self-help for social anxiety disorder: randomised controlled trial. Br J Psychiatry 2009 Nov;195(5):440-447. [doi: 10.1192/bjp.bp.108.060996] [Medline: 19880935]
68. Andersson G, Carlbring P, Holmström A, Sparthan E, Furmark T, Nilsson-Ihrfelt E, et al. Internet-based self-help with therapist feedback and in vivo group exposure for social phobia: a randomized controlled trial. J Consult Clin Psychol 2006 Aug;74(4):677-686. [doi: 10.1037/0022-006X.74.4.677] [Medline: 16881775]
69. Buntrock C, Berking M, Smit F, Lehr D, Nobis S, Riper H, et al. Preventing depression in adults with subthreshold depression: health-economic evaluation alongside a pragmatic randomized controlled trial of a web-based intervention. J Med Internet Res 2017 Jan 4;19(1):e5 [FREE Full text] [doi: 10.2196/jmir.6587] [Medline: 28052841]
70. Buntrock C, Ebert DD, Lehr D, Smit F, Riper H, Berking M, et al. Effect of a web-based guided self-help intervention for prevention of major depression in adults with subthreshold depression. Obstet Gynecol Surv 2016 May 3;71(9):526-527. [doi: 10.1097/01.ogx.0000490201.67580.e4]
71. Nobis S, Lehr D, Ebert DD, Baumeister H, Snoek F, Riper H, et al. Efficacy of a web-based intervention with mobile phone support in treating depressive symptoms in adults with type 1 and type 2 diabetes: a randomized controlled trial. Diabetes Care 2015 May;38(5):776-783. [doi: 10.2337/dc14-1728] [Medline: 25710923]
72. Nobis S, Ebert DD, Lehr D, Smit F, Buntrock C, Berking M, et al. Web-based intervention for depressive symptoms in adults with types 1 and 2 diabetes mellitus: a health economic evaluation. Br J Psychiatry 2018 Apr;212(4):199-206. [doi: 10.1192/bjp.2018.10] [Medline: 30071909]
73. Schlicker S, Weisel KK, Buntrock C, Berking M, Nobis S, Lehr D, et al. Do nonsuicidal severely depressed individuals with diabetes profit from internet-based guided self-help? Secondary analyses of a pragmatic randomized trial. J Diabetes Res 2019;2019:2634094 [FREE Full text] [doi: 10.1155/2019/2634094] [Medline: 31218230]
74. Ebert DD, Lehr D, Heber E, Riper H, Cuijpers P, Berking M. Internet- and mobile-based stress management for employees with adherence-focused guidance: efficacy and mechanism of change. Scand J Work Environ Health 2016 Sep 1;42(5):382-394 [FREE Full text] [doi: 10.5271/sjweh.3573] [Medline: 27249161]
75. McClay CA, Collins K, Matthews L, Haig C, McConnachie A, Morrison J, et al. A community-based pilot randomised controlled study of life skills classes for individuals with low mood and depression. BMC Psychiatry 2015 Feb 6;15:17 [FREE Full text] [doi: 10.1186/s12888-015-0384-2] [Medline: 25884922]
76. Richards D, Timulak L, O'Brien E, Hayes C, Vigano N, Sharry J, et al. A randomized controlled trial of an internet-delivered treatment: its potential as a low-intensity community intervention for adults with symptoms of depression. Behav Res Ther 2015 Dec;75:20-31. [doi: 10.1016/j.brat.2015.10.005] [Medline: 26523885]
77. Littlewood E, Duarte A, Hewitt C, Knowles S, Palmer S, Walker S, REEACT Team. A randomised controlled trial of computerised cognitive behaviour therapy for the treatment of depression in primary care: the randomised evaluation of the effectiveness and acceptability of computerised therapy (REEACT) trial. Health Technol Assess 2015 Dec;19(101):viii, xxi-viii, 171 [FREE Full text] [doi: 10.3310/hta191010] [Medline: 26685904]
78. Proudfoot J, Goldberg D, Mann A, Everitt B, Marks I, Gray J. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. Psychol Med 2003 Feb;33(2):217-227. [doi: 10.1017/s0033291702007225] [Medline: 12622301]
79. Dear BF, Staples LG, Terides MD, Karin E, Zou J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for generalized anxiety disorder and comorbid disorders: a randomized controlled trial. J Anxiety Disord 2015 Dec;36:63-77 [FREE Full text] [doi: 10.1016/j.janxdis.2015.09.003] [Medline: 26460536]
80. Titov N, Dear BF, Johnston L, Lorian C, Zou J, Wootton B, et al. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. PLoS One 2013;8(7):e62873 [FREE Full text] [doi: 10.1371/journal.pone.0062873] [Medline: 23843932]

81. Titov N, Dear B, Staples L, Terides M, Karin E, Sheehan J, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided treatment for major depressive disorder and comorbid anxiety disorders: a randomized controlled trial. J Anxiety Disord 2015 Oct;35:88-102 [FREE Full text] [doi: 10.1016/j.janxdis.2015.08.002] [Medline: 26422822]
82. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. Br J Psychiatry 2005 Nov;187:456-461. [doi: 10.1192/bjp.187.5.456] [Medline: 16260822]
83. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: a randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
84. Carlbring P, Bohman S, Brunt S, Buhrman M, Westling BE, Ekselius L, et al. Remote treatment of panic disorder: a randomized trial of internet-based cognitive behavior therapy supplemented with telephone calls. Am J Psychiatry 2006 Dec;163(12):2119-2125. [doi: 10.1176/ajp.2006.163.12.2119] [Medline: 17151163]
85. Carlbring P, Westling BE, Ljungstrand P, Ekselius L, Andersson G. Treatment of panic disorder via the internet: a randomized trial of a self-help program. Behav Ther 2001;32(4):751-764. [doi: 10.1016/S0005-7894(01)80019-8]
86. Marks IM, Kenwright M, McDonough M, Whittaker M, Mataix-Cols D. Saving clinicians' time by delegating routine aspects of therapy to a computer: a randomized controlled trial in phobia/panic disorder. Psychol Med 2004 Jan;34(1):9-17. [doi: 10.1017/s003329170300878x] [Medline: 14971623]
87. Mathiasen K, Riper H, Ehlers LH, Valentin JB, Rosenberg NK. Internet-based CBT for social phobia and panic disorder in a specialised anxiety clinic in routine care: results of a pilot randomised controlled trial. Internet Interv 2016 May;4:92-98 [FREE Full text] [doi: 10.1016/j.invent.2016.03.001] [Medline: 30135794]
88. Schneider AJ, Mataix-Cols D, Marks IM, Bachofen M. Internet-guided self-help with or without exposure therapy for phobic and panic disorders. Psychother Psychosom 2005;74(3):154-164. [doi: 10.1159/000084000] [Medline: 15832066]
89. Greist JH, Marks IM, Baer L, Kobak KA, Wenzel KW, Hirsch MJ, et al. Behavior therapy for obsessive-compulsive disorder guided by a computer or by a clinician compared with relaxation as a control. J Clin Psychiatry 2002 Feb;63(2):138-145. [doi: 10.4088/jcp.v63n0209] [Medline: 11874215]
90. Nordgreen T, Haug T, Öst LG, Andersson G, Carlbring P, Kvale G, et al. Stepped care versus direct face-to-face cognitive behavior therapy for social anxiety disorder and panic disorder: a randomized effectiveness trial. Behav Ther 2016 Mar;47(2):166-183. [doi: 10.1016/j.beth.2015.10.004] [Medline: 26956650]
91. Nordgreen T, Standal B, Mannes H, Haug T, Sivertsen B, Carlbring P, et al. Guided self-help via internet for panic disorder: dissemination across countries. Comput Hum Behav 2010 Jul;26(4):592-596. [doi: 10.1016/j.chb.2009.12.011]
92. Knaevelsrud C, Maercker A. Internet-based treatment for PTSD reduces distress and facilitates the development of a strong therapeutic alliance: a randomized controlled clinical trial. BMC Psychiatry 2007 Apr 19;7:13 [FREE Full text] [doi: 10.1186/1471-244X-7-13] [Medline: 17442125]
93. Lange A, Rietdijk D, Hudcovicova M, van de Ven JP, Schrieken B, Emmelkamp PM. Interapy: a controlled randomized trial of the standardized treatment of posttraumatic stress through the internet. J Consult Clin Psychol 2003 Oct;71(5):901-909. [doi: 10.1037/0022-006X.71.5.901] [Medline: 14516238]
94. Ruwaard J, Schrieken B, Schrijver M, Broeksteeg J, Dekker J, Vermeulen H, et al. Standardized web-based cognitive behavioural therapy of mild to moderate depression: a randomized controlled trial with a long-term follow-up. Cogn Behav Ther 2009;38(4):206-221. [doi: 10.1080/16506070802408086] [Medline: 19221919]
95. Ruwaard J, Broeksteeg J, Schrieken B, Emmelkamp P, Lange A. Web-based therapist-assisted cognitive behavioral treatment of panic symptoms: a randomized controlled trial with a three-year follow-up. J Anxiety Disord 2010 May;24(4):387-396. [doi: 10.1016/j.janxdis.2010.01.010] [Medline: 20227241]
96. Klein B, Richards JC, Austin DW. Efficacy of internet therapy for panic disorder. J Behav Ther Exp Psychiatry 2006 Sep;37(3):213-238. [doi: 10.1016/j.jbtep.2005.07.001] [Medline: 16126161]
97. Richards JC, Klein B, Austin DW. Internet cognitive behavioural therapy for panic disorder: Does the inclusion of stress management information improve end-state functioning? Clin Psychol 2006;10(1):2-15. [doi: 10.1080/13284200500378795]
98. Bergström J, Andersson G, Ljótsson B, Rück C, Andréewitch S, Karlsson A, et al. Internet-versus group-administered cognitive behaviour therapy for panic disorder in a psychiatric setting: a randomised trial. BMC Psychiatry 2010 Jul 2;10:54 [FREE Full text] [doi: 10.1186/1471-244X-10-54] [Medline: 20598127]
99. Titov N, Fogliati VJ, Staples LG, Gandy M, Johnston K, Wootton B, et al. Treating anxiety and depression in older adults: randomised controlled trial comparing guided. Self-guided internet-delivered cognitive-behavioural therapy. BJPsych Open 2016 Jan;2(1):50-58 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002139] [Medline: 27703754]
100. Wootton BM, Dear BF, Johnston L, Terides MD, Titov N. Remote treatment of obsessive-compulsive disorder: a randomized controlled trial. J Obsess-Compuls Rel 2013 Oct;2(4):375-384. [doi: 10.1016/j.jocrd.2013.07.002]
101. Spence J, Titov N, Johnston L, Jones MP, Dear BF, Solley K. Internet-based trauma-focused cognitive behavioural therapy for PTSD with and without exposure components: a randomised controlled trial. J Affect Disord 2014 Jun;162:73-80. [doi: 10.1016/j.jad.2014.03.009] [Medline: 24767009]
102. Mattick RP, Clarke J. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behav Res Ther 1998 Apr;36(4):455-470. [doi: 10.1016/s0005-7967(97)10031-6] [Medline: 9670605]

103. Baker SL, Heinrichs N, Kim H, Hofmann SG. The liebowitz social anxiety scale as a self-report instrument: a preliminary psychometric analysis. Behav Res Ther 2002 Jun;40(6):701-715. [doi: 10.1016/s0005-7967(01)00060-2] [Medline: 12051488]
104. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psych Ann 2002 Sep 1;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
105. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
106. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
107. Houck PR, Spiegel DA, Shear MK, Rucci P. Reliability of the self-report version of the panic disorder severity scale. Depress Anxiety 2002;15(4):183-185. [doi: 10.1002/da.10049] [Medline: 12112724]
108. Sheehan KH, Sheehan DV. Assessing treatment effects in clinical trials with the discan metric of the Sheehan disability scale. Int Clin Psychopharmacol 2008 Mar;23(2):70-83. [doi: 10.1097/YIC.0b013e3282f2b4d6] [Medline: 18301121]
109. Svanborg P, Asberg M. A new self-rating scale for depression and anxiety states based on the comprehensive psychopathological rating scale. Acta Psychiatr Scand 1994 Jan;89(1):21-28. [doi: 10.1111/j.1600-0447.1994.tb01480.x] [Medline: 8140903]
110. Marks I, Mathews A. Brief standard self-rating for phobic patients. Behav Res Ther 1979;17(3):263-267. [doi: 10.1016/0005-7967(79)90041-x] [Medline: 526242]
111. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun;4:561-571. [doi: 10.1001/archpsyc.1961.01710120031004] [Medline: 13688369]
112. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23:56-62 [FREE Full text] [doi: 10.1136/jnnp.23.1.56] [Medline: 14399272]
113. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988 Dec;56(6):893-897. [doi: 10.1037//0022-006x.56.6.893] [Medline: 3204199]
114. Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, et al. The Yale-Brown obsessive compulsive scale. I. Development, use, and reliability. Arch Gen Psychiatry 1989 Nov;46(11):1006-1011. [doi: 10.1001/archpsyc.1989.01810110048007] [Medline: 2684084]
115. Chambless DL, Caputo GC, Bright P, Gallagher R. Assessment of fear of fear in agoraphobics: the body sensations questionnaire and the agoraphobic cognitions questionnaire. J Consult Clin Psychol 1984 Dec;52(6):1090-1097. [doi: 10.1037/0022-006X.52.6.1090]
116. Chambless DL, Caputo G, Jasin SE, Gracely EJ, Williams C. The mobility inventory for agoraphobia. Behav Res Ther 1985;23(1):35-44. [doi: 10.1016/0005-7967(85)90140-8]
117. Nieuwenhuijsen K, de Boer AG, Verbeek JH, Blonk RW, van Dijk FJ. The depression anxiety stress Sscales (DASS): detecting anxiety disorder and depression in employees absent from work because of mental health problems. Occup Environ Med 2003 Jun;60(Suppl 1):i77-i82 [FREE Full text] [doi: 10.1136/oem.60.suppl_1.i77] [Medline: 12782751]
118. Horowitz M, Wilner N, Alvarez W. Impact of event scale: a measure of subjective stress. Psychosom Med 1979 May;41(3):209-218. [doi: 10.1097/00006842-197905000-00004] [Medline: 472086]
119. Taylor S, Cox BJ. Anxiety sensitivity: multiple dimensions and hierarchic structure. Behav Res Ther 1998 Jan;36(1):37-51. [doi: 10.1016/s0005-7967(97)00071-5] [Medline: 9613015]
120. McDonald SD, Calhoun PS. The diagnostic accuracy of the PTSD checklist: a critical review. Clin Psychol Rev 2010 Dec;30(8):976-987. [doi: 10.1016/j.cpr.2010.06.012] [Medline: 20705376]
121. Borkovec TD, Nau SD. Credibility of analogue therapy rationales. J Behav Ther Exp Psychiatry 1972 Dec;3(4):257-260. [doi: 10.1016/0005-7916(72)90045-6]
122. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000 Jun;31(2):73-86. [doi: 10.1016/s0005-7916(00)00012-4] [Medline: 11132119]
123. Nguyen TD, Attkisson C, Stegner BL. Assessment of patient satisfaction: development and refinement of a service evaluation questionnaire. Eval Program Plann 1983;6(3-4):299-313. [doi: 10.1016/0149-7189(83)90010-1] [Medline: 10267258]
124. Karyotaki E, Ebert DD, Donkin L, Riper H, Twisk J, Burger S, et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. Clin Psychol Rev 2018 Jul;63:80-92. [doi: 10.1016/j.cpr.2018.06.007] [Medline: 29940401]
125. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2015 Mar 05(3):CD011565. [doi: 10.1002/14651858.CD011565] [Medline: 25742186]
126. Seidler Z, Dawes A, Rice S, Oliffe J, Dhillon H. The role of masculinity in men's help-seeking for depression: a systematic review. Clin Psychol Rev 2016 Nov;49:106-118. [doi: 10.1016/j.cpr.2016.09.002] [Medline: 27664823]
127. Thompson A, Anisimowicz Y, Miedema B, Hogg W, Wodchis W, Aubrey-Bassler K. The influence of gender and other patient characteristics on health care-seeking behaviour: a QUALICOPC study. BMC Fam Pract 2016 Mar 31;17:38 [FREE Full text] [doi: 10.1186/s12875-016-0440-0] [Medline: 27036116]

128. Seedat S, Scott KM, Angermeyer MC, Berglund P, Bromet EJ, Brugha TS, et al. Cross-national associations between gender and mental disorders in the World Health Organization World Mental Health Surveys. Arch Gen Psychiatry 2009 Jul;66(7):785-795 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2009.36] [Medline: 19581570]
129. Kuehner C. Gender differences in unipolar depression: an update of epidemiological findings and possible explanations. Acta Psychiatr Scand 2003 Sep;108(3):163-174. [doi: 10.1034/j.1600-0447.2003.00204.x] [Medline: 12890270]
130. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit J, et al. Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
131. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
132. Ebert DD, Baumeister H. Internet- und mobil-basierte Interventionen in der Psychotherapiein Überblick. Psychotherapeutenjournal 2016;1:31. [doi: 10.1515/pubhef-2015-0063]
133. Allen AR, Newby JM, Mackenzie A, Smith J, Boulton M, Loughnan SA, et al. Internet cognitive-behavioural treatment for panic disorder: randomised controlled trial and evidence of effectiveness in primary care. BJPsych Open 2016 Mar;2(2):154-162 [FREE Full text] [doi: 10.1192/bjpo.bp.115.001826] [Medline: 27703768]
134. Andersson G, Waara J, Jonsson U, Malmaeus F, Carlbring P, Ost LG. Internet-based exposure treatment versus one-session exposure treatment of snake phobia: a randomized controlled trial. Cogn Behav Ther 2013;42(4):284-291. [doi: 10.1080/16506073.2013.844202] [Medline: 24245707]
135. Andersson G, Waara J, Jonsson U, Malmaeus F, Carlbring P, Ost LG. Internet-based self-help versus one-session exposure in the treatment of spider phobia: a randomized controlled trial. Cogn Behav Ther 2009;38(2):114-120. [doi: 10.1080/16506070902931326] [Medline: 20183690]
136. Berger T, Boettcher J, Caspar F. Internet-based guided self-help for several anxiety disorders: a randomized controlled trial comparing a tailored with a standardized disorder-specific approach. Psychotherapy (Chic) 2014 Jun;51(2):207-219. [doi: 10.1037/a0032527] [Medline: 24041199]
137. Berger T, Caspar F, Richardson R, Kneubühler B, Sutter D, Andersson G. Internet-based treatment of social phobia: a randomized controlled trial comparing unguided with two types of guided self-help. Behav Res Ther 2011 Mar;49(3):158-169. [doi: 10.1016/j.brat.2010.12.007] [Medline: 21255767]
138. Berger T, Hohl E, Caspar F. Internet-based treatment for social phobia: a randomized controlled trial. J Clin Psychol 2009 Oct;65(10):1021-1035. [doi: 10.1002/jclp.20603] [Medline: 19437505]
139. Bergström J, Andersson G, Ljótsson B, Rück C, Andréewitch S, Karlsson A, et al. Internet-versus group-administered cognitive behaviour therapy for panic disorder in a psychiatric setting: a randomised trial. BMC Psychiatry 2010 Jul 2;10:54 [FREE Full text] [doi: 10.1186/1471-244X-10-54] [Medline: 20598127]
140. Carlbring P, Bohman S, Brunt S, Buhrman M, Westling BE, Ekselius L, et al. Remote treatment of panic disorder: a randomized trial of internet-based cognitive behavior therapy supplemented with telephone calls. Am J Psychiatry 2006 Dec;163(12):2119-2125. [doi: 10.1176/ajp.2006.163.12.2119] [Medline: 17151163]
141. Carlbring P, Gunnarsdóttir M, Hedensjö L, Andersson G, Ekselius L, Furmark T. Treatment of social phobia: randomised trial of internet-delivered cognitive-behavioural therapy with telephone support. Br J Psychiatry 2007 Feb;190:123-128. [doi: 10.1192/bjp.bp.105.020107] [Medline: 17267928]
142. Carlbring P, Maurin L, Törngren C, Linna E, Eriksson T, Sparthan E, et al. Individually-tailored, internet-based treatment for anxiety disorders: a randomized controlled trial. Behav Res Ther 2011 Jan;49(1):18-24. [doi: 10.1016/j.brat.2010.10.002] [Medline: 21047620]
143. Furmark T, Carlbring P, Hedman E, Sonnenstein A, Clevberger P, Bohman B, et al. Guided and unguided self-help for social anxiety disorder: randomised controlled trial. Br J Psychiatry 2009 Nov;195(5):440-447. [doi: 10.1192/bjp.bp.108.060996] [Medline: 19880935]
144. Hedman E, Andersson G, Ljótsson B, Andersson E, Rück C, Mörtberg E, et al. Internet-based cognitive behavior therapy vs. cognitive behavioral group therapy for social anxiety disorder: a randomized controlled non-inferiority trial. PLoS One 2011 Mar 25;6(3):e18001 [FREE Full text] [doi: 10.1371/journal.pone.0018001] [Medline: 21483704]
145. Hedman E, El Alaoui S, Lindefors N, Andersson E, Rück C, Ghaderi A, et al. Clinical effectiveness and cost-effectiveness of Internet- vs. group-based cognitive behavior therapy for social anxiety disorder: 4-year follow-up of a randomized trial. Behav Res Ther 2014 Aug;59:20-29. [doi: 10.1016/j.brat.2014.05.010] [Medline: 24949908]
146. Ivarsson D, Blom M, Hesser H, Carlbring P, Enderby P, Nordberg R, et al. Guided internet-delivered cognitive behavior therapy for post-traumatic stress disorder: a randomized controlled trial. Internet Interv 2014 Mar;1(1):33-40. [doi: 10.1016/j.invent.2014.03.002]
147. Johnston L, Titov N, Andrews G, Spence J, Dear BF. A RCT of a transdiagnostic internet-delivered treatment for three anxiety disorders: examination of support roles and disorder-specific outcomes. PLoS One 2011;6(11):e28079 [FREE Full text] [doi: 10.1371/journal.pone.0028079] [Medline: 22132216]
148. Jones SL, Hadjistavropoulos HD, Soucy JN. A randomized controlled trial of guided internet-delivered cognitive behaviour therapy for older adults with generalized anxiety. J Anxiety Disord 2016 Jan;37:1-9. [doi: 10.1016/j.janxdis.2015.10.006] [Medline: 26561733]

149. Kok RN, van Straten A, Beekman AT, Cuijpers P. Short-term effectiveness of web-based guided self-help for phobic outpatients: randomized controlled trial. J Med Internet Res 2014 Sep 29;16(9):e226 [FREE Full text] [doi: 10.2196/jmir.3429] [Medline: 25266929]
150. Nordgren LB, Hedman E, Etienne J, Bodin J, Kadowaki A, Eriksson S, et al. Effectiveness and cost-effectiveness of individually tailored internet-delivered cognitive behavior therapy for anxiety disorders in a primary care population: a randomized controlled trial. Behav Res Ther 2014 Aug;59:1-11 [FREE Full text] [doi: 10.1016/j.brat.2014.05.007] [Medline: 24933451]
151. Oromendia P, Orrego J, Bonillo A, Molinuevo B. Internet-based self-help treatment for panic disorder: a randomized controlled trial comparing mandatory versus optional complementary psychological support. Cogn Behav Ther 2016 Jun;45(4):270-286. [doi: 10.1080/16506073.2016.1163615] [Medline: 27007256]
152. Paxling B, Almlöv J, Dahlin M, Carlbring P, Breitholtz E, Eriksson T, et al. Guided internet-delivered cognitive behavior therapy for generalized anxiety disorder: a randomized controlled trial. Cogn Behav Ther 2011;40(3):159-173. [doi: 10.1080/16506073.2011.576699] [Medline: 21770848]
153. Robinson E, Titov N, Andrews G, McIntyre K, Schwencke G, Solley K. Internet treatment for generalized anxiety disorder: a randomized controlled trial comparing clinician vs technician assistance. PLoS One 2010 Jun 3;5(6):e10942 [FREE Full text] [doi: 10.1371/journal.pone.0010942] [Medline: 20532167]
154. Silfvernagel L, Carlbring P, Kabo J, Edström S, Eriksson J, Månson L, et al. Individually tailored internet-based treatment for young adults and adults with panic attacks: randomized controlled trial. J Med Internet Res 2012 Jun 26;14(3):e65 [FREE Full text] [doi: 10.2196/jmir.1853] [Medline: 22732098]
155. Spence J, Titov N, Dear BF, Johnston L, Solley K, Lorian C, et al. Randomized controlled trial of Internet-delivered cognitive behavioral therapy for posttraumatic stress disorder. Depress Anxiety 2011 Jul;28(7):541-550. [doi: 10.1002/da.20835] [Medline: 21721073]
156. Titov N, Andrews G, Choi I, Schwencke G, Mahoney A. Shyness 3: randomized controlled trial of guided versus unguided internet-based CBT for social phobia. Aust N Z J Psychiatry 2008 Dec;42(12):1030-1040. [doi: 10.1080/00048670802512107] [Medline: 19016091]
157. Titov N, Andrews G, Johnston L, Robinson E, Spence J. Transdiagnostic internet treatment for anxiety disorders: a randomized controlled trial. Behav Res Ther 2010 Sep;48(9):890-899. [doi: 10.1016/j.brat.2010.05.014] [Medline: 20561606]
158. Titov N, Dear BF, Schwencke G, Andrews G, Johnston L, Craske MG, et al. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav Res Ther 2011 Aug;49(8):441-452. [doi: 10.1016/j.brat.2011.03.007] [Medline: 21679925]
159. Tulbure BT, Szentagotai A, David O, tefan S, Månsson KN, David D, et al. Internet-delivered cognitive-behavioral therapy for social anxiety disorder in Romania: a randomized controlled trial. PLoS One 2015;10(5):e0123997 [FREE Full text] [doi: 10.1371/journal.pone.0123997] [Medline: 25938241]
160. van Ballegooijen W, Riper H, Klein B, Ebert DD, Kramer J, Meulenbeek P, et al. An internet-based guided self-help intervention for panic symptoms: randomized controlled trial. J Med Internet Res 2013 Jul 29;15(7):e154 [FREE Full text] [doi: 10.2196/jmir.2362] [Medline: 23896222]
161. Wims E, Titov N, Andrews G, Choi I. Clinician-assisted Internet-based treatment is effective for panic: a randomized controlled trial. Aust N Z J Psychiatry 2010 Jul;44(7):599-607. [doi: 10.3109/00048671003614171] [Medline: 20560847]
162. Wootton BM, Dear BF, Johnston L, Terides MD, Titov N. Remote treatment of obsessive-compulsive disorder: a randomized controlled trial. J Obsess-Compuls Rel 2013 Oct;2(4):375-384. [doi: 10.1016/j.jocrd.2013.07.002]
163. Stewart RE, Chambless DL. Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: a meta-analysis of effectiveness studies. J Consult Clin Psychol 2009 Aug;77(4):595-606. [doi: 10.1037/a0016032] [Medline: 19634954]
164. Ebert DD, Donkin L, Andersson G, Andrews G, Berger T, Carlbring P, et al. Does internet-based guided-self-help for depression cause harm? An individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials. Psychol Med 2016 Oct;46(13):2679-2693 [FREE Full text] [doi: 10.1017/S0033291716001562] [Medline: 27649340]
165. Rozental A, Magnusson K, Boettcher J, Andersson G, Carlbring P. For better or worse: an individual patient data meta-analysis of deterioration among participants receiving internet-based cognitive behavior therapy. J Consult Clin Psychol 2017 Feb;85(2):160-177. [doi: 10.1037/ccp0000158] [Medline: 27775414]
166. Cuijpers P, Reijnders M, Karyotaki E, de Wit L, Ebert DD. Negative effects of psychotherapies for adult depression: a meta-analysis of deterioration rates. J Affect Disord 2018 Oct 15;239:138-145. [doi: 10.1016/j.jad.2018.05.050] [Medline: 30005327]
167. Rozental A, Castonguay L, Dimidjian S, Lambert M, Shafran R, Andersson G, et al. Negative effects in psychotherapy: commentary and recommendations for future research and clinical practice. BJPsych Open 2018 Jul;4(4):307-312 [FREE Full text] [doi: 10.1192/bjo.2018.42] [Medline: 30083384]
168. Mewton L, Andrews G. Cognitive behaviour therapy via the internet for depression: a useful strategy to reduce suicidal ideation. J Affect Disord 2015 Jan 1;170:78-84. [doi: 10.1016/j.jad.2014.08.038] [Medline: 25233243]

169. Kendrick T, El-Gohary M, Stuart B, Gilbody S, Churchill R, Aiken L, et al. Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. Cochrane Database Syst Rev 2016 Jul 13;7:CD011119 [FREE Full text] [doi: 10.1002/14651858.CD011119.pub2] [Medline: 27409972]
170. Lambert MJ, Harmon KL. The merits of implementing routine outcome monitoring in clinical practice. Clin Psychol Sci Pract 2018 Dec 12;25(4):e12268. [doi: 10.1111/cpsp.12268]
171. Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. Psychother Res 2015;25(1):6-19. [doi: 10.1080/10503307.2013.817696] [Medline: 23885809]
172. Cuijpers P, Andersson G, Donker T, van Straten A. Psychological treatment of depression: results of a series of meta-analyses. Nord J Psychiatry 2011 Dec;65(6):354-364. [doi: 10.3109/08039488.2011.596570] [Medline: 21770842]
173. Mason E, Andrews G. The use of automated assessments in internet-based CBT: the computer will be with you shortly. Internet Interv 2014 Oct;1(4):216-224. [doi: 10.1016/j.invent.2014.10.003]
174. Cuijpers P. The challenges of improving treatments for depression. J Am Med Assoc 2018 Dec 25;320(24):2529-2530. [doi: 10.1001/jama.2018.17824] [Medline: 30500053]
175. Titov N, Dear B, Nielssen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: a descriptive analysis of successful clinics in five countries. Internet Interv 2018 Sep;13:108-115 [FREE Full text] [doi: 10.1016/j.invent.2018.07.006] [Medline: 30206525]
176. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
177. Hermes ED, Lyon AR, Schueller SM, Glass JE. Measuring the implementation of behavioral intervention technologies: recharacterization of established outcomes. J Med Internet Res 2019 Jan 25;21(1):e11752 [FREE Full text] [doi: 10.2196/11752] [Medline: 30681966]
178. Klein JP, Knaevelsrud C, Bohus M, Ebert DD, Gerlinger G, Günther K, et al. [Finding the needles in the haystack: quality criteria for the use of internet-based self-management interventions in the prevention and treatment of mental disorders]. Fortschr Neurol Psychiatr 2019 Mar;87(3):187-191. [doi: 10.1055/a-0849-9902] [Medline: 30891719]

Abbreviations

GP: general practitioner
iCBT: internet-based cognitive behavioral therapy
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 03.02.20; peer-reviewed by A Rozental, E Szigethy, K Magnusson, N Lindefors; comments to author 26.02.20; revised version received 06.05.20; accepted 03.06.20; published 31.08.20.

Please cite as:
Etzelmueller A, Vis C, Karyotaki E, Baumeister H, Titov N, Berking M, Cuijpers P, Riper H, Ebert DD
Effects of Internet-Based Cognitive Behavioral Therapy in Routine Care for Adults in Treatment for Depression and Anxiety: Systematic Review and Meta-Analysis
J Med Internet Res 2020;22(8):e18100
URL: http://www.jmir.org/2020/8/e18100/
doi:10.2196/18100
PMID:32865497

Proposal

Public Health in the Information Age: Recognizing the Infosphere as a Social Determinant of Health

Jessica Morley[1], BA; Josh Cowls[1,2], BA, MSc; Mariarosaria Taddeo[1,2], BA, MA, PhD; Luciano Floridi[1,2], BA, MPhil, PhD

[1]Oxford Internet Institute, University of Oxford, Oxford, United Kingdom

[2]Alan Turing Institute, London, United Kingdom

Corresponding Author:
Jessica Morley, BA
Oxford Internet Institute
University of Oxford
1 St Giles',
Oxford, OX1 3JS
United Kingdom
Phone: 44 01865 287210
Email: jessica.morley@phc.ox.ac.uk

Abstract

Since 2016, social media companies and news providers have come under pressure to tackle the spread of political mis- and disinformation (MDI) online. However, despite evidence that online health MDI (on the web, on social media, and within mobile apps) also has negative real-world effects, there has been a lack of comparable action by either online service providers or state-sponsored public health bodies. We argue that this is problematic and seek to answer three questions: why has so little been done to control the flow of, and exposure to, health MDI online; how might more robust action be justified; and what specific, newly justified actions are needed to curb the flow of, and exposure to, online health MDI? In answering these questions, we show that four ethical concerns—related to paternalism, autonomy, freedom of speech, and pluralism—are partly responsible for the lack of intervention. We then suggest that these concerns can be overcome by relying on four arguments: (1) education is necessary but insufficient to curb the circulation of health MDI, (2) there is precedent for state control of internet content in other domains, (3) network dynamics adversely affect the spread of accurate health information, and (4) justice is best served by protecting those susceptible to inaccurate health information. These arguments provide a strong case for classifying the quality of the infosphere as a social determinant of health, thus making its protection a public health responsibility. In addition, they offer a strong justification for working to overcome the ethical concerns associated with state-led intervention in the infosphere to protect public health.

(J Med Internet Res 2020;22(8):e19311) doi:10.2196/19311

KEYWORDS

COVID-19; public health; misinformation; disinformation; infodemic; infodemiology; infosphere; social determinants of health; information ethics

Introduction

The internet's capacity to generate and spread misinformation had already been identified and described 24 years ago [1]. However, it was the result of both the US presidential election and the UK's referendum on European Union membership in 2016 that woke up civil society to the real-world effects of the online spread of false or inaccurate information (also known as *misinformation*) or deliberately misleading information (also known as *disinformation*; on the distinction between misinformation and disinformation [MDI] see [2] and more recently, in relation to health, [3]). As ever more political news and reporting has moved online—where network effects and a lack of gatekeepers mean that half-truths and mistruths can spread at greater speed and scale—the task of rapidly debunking false claims has been taken up by a growing army of fact-checking organizations. Additionally, social media companies have come under pressure to be more transparent about who has purchased specific political adverts and to provide consumers with "explanations" of the reasoning behind targeted adverts they see online [4]. These may only be small (and perhaps relatively ineffective) measures, but at least they show that there is growing consensus on the obligation of online

service providers (OSPs) to take positive action to protect citizens and the democratic process [5]. Unfortunately, there have been almost no coordinated actions taken to tackle the equivalent issues associated with the online propagation of *misinformation* and *disinformation* as they relate to health. Throughout this paper, we use "health information" in a broad sense including medical information (eg, symptoms and treatments of specific diseases or injuries) and wellness information (eg, diet or fitness advice).

This lack of action is increasingly concerning, given the real impact of "the medical misinformation mess" [6]. "One day, being an inforg [informational organisms] will be so natural that any disruption in our normal flow of information will make us sick. Even literally." That time has now come [7]. Yet, those actively seeking health advice and those browsing the web, social media, or even app stores for other reasons are faced with an almost constant barrage of medical news stories, social media posts, spurious website results, direct-to-consumer drug and medical adverts, and hospital and digital-health service marketing messages. Almost all of these are entirely inaccurate [6]. For instance, studies of vaccine-related internet content have shown consistently that most of this content is misleading and that false messages are more likely to be liked and shared than those that are accurate [8]. As a consequence of the lack of intervention in this state of affairs, hesitancy around getting vaccinated is now a major global health concern. At the same time, myriad online communities promoting self-harm, anorexia, and homeopathy now exist; unevidenced and unregulated apps are freely available for download; and the reckless promotion of fad diets and unproven wellness trends by celebrities on unregulated social media platforms is leading to the spread of various dangerous behaviors [9,10]. In short, as [11] states, "A child who needlessly experiences disabilities caused by measles, an adult who dies after stopping a statin despite having high-risk coronary heart disease, and a patient with cancer who ceases chemotherapy in favour of a bogus alternative are all victims of misinformation that is being promulgated on social media and other internet platforms."

Events concerning the coronavirus disease (COVID-19) pandemic have only exacerbated the problem, to the extent that on February 15, 2020, the World Health Organization (WHO) Director-General Tedros Adhanom Ghebreyesus stated that: "We're not just fighting an epidemic; we're fighting an infodemic" [12]. Note that the word "infodemic" is not new. It was introduced in 2003 to refer to the spread of MID about severe acute respiratory syndrome [13].

Just like traditional political institutions, public health bodies underestimated the capacity of the web and social media to exert serious and potentially dangerous influence over health-related behavior [14]. This raises the following crucial questions:

- Why has so little been done to date to control the flow of, and exposure to, health MID online?
- In what circumstances is more robust action justified?
- What specific newly justified actions are needed to curb the flow of, and exposure to, online health-related misinformation and disinformation (OHMDI)?

In the following pages, we shall answer these questions in turn. Specifically, the section Missed Opportunities considers the early days of health information on the web, looking at why proactive attempts to govern better online health information were unsuccessful. The section Protecting the Individual Over the Group highlights how the criticism of public health interventions as paternalistic and antithetical to principles of bioethics has prevented further such attempts. The section Justifying Intervention discusses why these arguments are flawed in the specific context of OHMDI. The section the Infosphere as Social Determinant of Health provides a framework within which public health bodies may be able to act to tackle the OHMDI problem. The section Prevention, Protection, and Promotion: an Action Ontology provides a mode-of-action ontology for public health bodies. The section Actions and Agents considers the specific actions within this ontology that different actors in the system could take to tackle the problem. The last section concludes the article.

Missed Opportunities

The quality of health information on the internet first became a major cause of concern for health care professionals, information specialists, and consumers of health care in the mid-1990s [15], when the web came to be portrayed by the medical community as a dangerous space that lacked the gatekeeping functions necessary to protect naïve health consumers [16]. The initial response to this mounting concern over the quality of online health information was a push from the medical community for greater regulation. However, when the rapid proliferation of content made it apparent that regulation would be unable to keep up, the focus of the community shifted toward market-based levers. This resulted in a proliferation of largely unsuccessful kitemarking, filtering, and accreditation schemes [17]. Textbox 1 summarizes four important examples of initiatives developed during this period.

Textbox 1. Examples of schemes concerning regulation of online health information.

Schemes

- *The eHealth Code of Ethics*, produced by the Internet Healthcare Coalition in response to issues raised by the Food and Drug Administration in 1996 [18]
- *OMNI*, a search engine programmed to look solely for high-quality health care information designed to return only "validated" results [18]
- *The European Commission's code of good practice for health websites*, published in June 2002 and based on the principles of transparency and honesty; authority, privacy and confidentiality; currency, accountability, and accessibility [19]
- The creation of the *Health on the Net Foundation* in May 1996, following the 1995 international meeting *Use of the Internet and World-Wide Web for Telematics in Healthcare*, which was charged with enabling the appropriate and efficient use of reliable health information online

Such attention from academia and supranational policy makers prompted the WHO to act, submitting a proposal to the Internet Corporation for Assigned Names and Numbers (ICANN) for the creation of a sponsored top-level .health domain. This proposal suggested that the WHO would, through consultation, develop a set of quality and ethical criteria that would-be .health sites would have to meet; it would ensure compliance by random checks conducted on approved sites and have an annual reregistration process [20]. The intention was not to police or regulate all health information on the web but to offer a reliable go-to domain to support users who wanted to narrow their search field to include credible sources only [21].

This idea was broadly supported by ICANN's chairman at the time, Vinton Cerf, who stated that "we feel it would [have been] a great benefit to consumers for guaranteeing the quality of health and medical information on the web" and encouraged the WHO to pursue the idea further [22]. However, this support was not strong enough in the face of opposition from various stakeholders. They successfully argued that the web could not be policed, users were already sophisticated enough to recognize quackery [21], and no one body should assume the right to veto many thousands of websites [22].

These arguments against the need for a sponsored .health domain have been so influential that the WHO has been discouraged from bidding for the domain name again [23] and almost all other quality measures have also failed to gain traction (excepting the HONcode certification scheme). ICANN opened a new large-scale program to create multiple general top-level domains (gTLDs). In June 2011, the domains .health, .care, .diet, .doctor, .healthcare, .help, .hospital, and .med all went to the highest private bidder [24]. Mackey and Nayer [24] describe this process and fallout in detail. Textbox 2 shows the current owners of each health-related domain. There was no requirement for the domain purchasers to meet any specific criteria. For example, it is currently possible to make unrestricted purchases of potentially dangerous domain names such as suicidetips.health.

Such a hands-off approach to the governance of health-related domains suggests that the global community has reached the conclusion that the right strategy for improving the quality of health information on the internet is not content moderation, monitoring, or certification of reliable sources; instead, the focus should be on educating content producers and consumers [25]. This is the argument provided by Eysenbach [26] in a *JMIR* editorial when discussing the difference between moderating *content* and *source* quality (original emphasis):

> *No single body (let alone the domain registrar) should determine what is "correct" health information. It cannot be the goal to "censor" content or the messages on .health websites. It will always remain up to the website owners to ensure "message credibility," and will always remain the responsibility of users to learn how to distinguish quality sites.*

This argument of scale being a barrier to intervention has only become stronger with the advent of other sources of unregulated online health information, such as social media and mobile apps. The following section considers the ethical arguments that have pushed internet governance, medical device regulators, and public health bodies in this direction.

Textbox 2. Operators of health domains as of January 2020, according to the Internet Assigned Numbers Authority.

.health

DotHealth LLC

.care

Binky Moon LLC (subsidiary of Donuts Inc)

.diet

Binky Moon LLC (subsidiary of Donuts Inc)

.doctor

Binky Moon LLC (subsidiary of Donuts Inc)

.healthcare

Binky Moon LLC (subsidiary of Donuts Inc)

.help

Unriregistry Corp

.hospital

Binky Moon LLC (subsidiary of Donuts Inc)

.med

Medistry LLC

Protecting the Individual Over the Group

To understand the reasons why suggestions that online health information (encompassing all sources such as the web, social media, and mobile apps) should be subject to more stringent controls have typically failed, it is important to recall that public health is focused on the population at large and is primarily *preventive*, while clinical medicine is focused on the individual and is primarily *reactive* [27]. Thus, public health bodies seek to understand the societal conditions under which people can lead healthier lives to minimize those threats to health that can be averted or lessened *only* through collective actions aimed at the community [28]. These actions include, for example, disease surveillance, epidemiological modeling, national screening and vaccination programs, water sanitization efforts, and quarantining [28]. In short, public health officials take the position that society as a whole bears responsibility for the prevention of ill health. This recognizes that sometimes individuals acting alone are powerless to make the necessary changes and that only by acting together through public institutions can they protect the health of the communities to which they belong. This argument provides the philosophical justification for state interventions that override individual freedoms for the sake of promoting public health [29]. The same argument motivated the oft-quoted conclusion of the 1905 Supreme Court ruling *Jacobson vs. Massachusetts*:

> *The liberty secured by the Constitution of the US to every person within its jurisdiction does not impart an absolute right in each person to be, at all times and in all circumstances, wholly freed from restraint. There are manifold restraints to which every person is necessarily subject for the common good.*

Although public health interventions that are justified on this basis are intended to be a form of collective problem solving, they are often perceived as being overly authoritarian [30]. As a result, although science has advanced to improve the efficacy of public health programs, public acceptance of them has declined [31] as the philosophy underpinning public policy has shifted from consequentialism, contractarianism, and communitarianism toward liberalism and principlism. Principlism is an approach to ethical evaluations and decision making that relies on the application of moral principles, rather than high-level normative theories such as virtue ethics, deontologism, or consequentialism [32]. It is popular in professional contexts but liberalism has prevailed in public health contexts because these alternative philosophies, all of which are central to public health, have been criticized for undermining the rights of the individual, as they rely on the idea that the end justifies the means, and for embracing an inherently paternalistic approach. Indeed, as reported by Buchanan [29], for many, the central concern of public health ethics is when it is justifiable to override individual freedom for the sake of public health. This is even a critical concern of the landmark Lalonde [33] report, which first recognized the role of factors other than the quality of the health care system in managing public health. It states: "The ultimate philosophical issue...is whether and to what extent the government can get into the business of modifying human behavior, even if it does so to improve health" [33]. Thus, the question for several philosophers working in this area, such as Dworkin in [34-36], is not whether public health interventions are paternalistic, but when are paternalistic interventions justified [29,37]?

These questions have been complicated as principlism, as a basis for bioethics in the clinical domain, has expanded into the public health domain, with a growing emphasis being placed on the principle of "autonomy." As a result, public health policy has been pushed away from the idea that some aspects of health are outside of individual control, with the focus instead on encouraging citizens to take individual (and sole) responsibility for their health [16]. For example, Public Health England's *Change for Life* program is based entirely around individuals taking action to improve their own health by eating better and taking part in regular exercise, but it does not include initiatives to improve access to resources that would enable these behavioral changes.

In this context, it has become ethically and politically difficult to argue in favor of tougher online health information controls. A website, social media post, or mobile app can be written by one person and read, commented, shared, downloaded, and edited by thousands. Intervening by, for example, automatically removing or flagging MDI would be perceived as a paternalistic (or even censorious) restriction on individual autonomy, particularly when the current overarching health policy paradigm is heavily infused with the (misguided) belief that information automatically leads to individual empowerment [38]. Furthermore, regulation of online health information is likely to be accused of being in conflict with the right to freedom of speech [17] and so harmful to the development of a pluralistic society. This is because pluralism, tolerance, and broadmindedness must go together, according to the European Court of Human Rights, and there must be room for individuals to express controversial opinions, including those about health [39]. In addition, in 2012, Internet Freedom was declared a human right by the United Nations (UN) Human Rights Council, which called on states to ensure that the rights to freedom of expression and information, as presented in Article 19 of the Universal Declaration of Human Rights, would be upheld online as well as offline [5].

As a consequence, although platforms including Twitter and Instagram do block specific hashtags (for example, #proana, an abbreviation of "pro-anorexia," is not searchable on Instagram), public health bodies have thus far managed to justify only noncoercive state-level interventions focused on educating citizens. For example, throughout 2019, Public Health England, National Health Service (NHS) England, and the Department of Health and Social Care ran the #ValueofVaccines campaign with the intention of maintaining parental confidence in vaccines and shifting conversations away from antivaccination [40]. More coercive forms of information control are perceived to be neither necessary nor proportionate.

This is undeniably a formidable set of arguments to tackle. However, in the context of online health information, there are a number of pertinent and convincing objections. These are set out in the following section.

Justifying Intervention

Overview

The shift toward liberalism and principlism as the philosophical grounding of public policy, including public health policy, has hindered those who have previously argued in favor of tougher regulation of online health information. Even the call by the 66th World Health Assembly in 2013 for all health-related gTLDs to be used to promote public health and for member states to work with ICANN's Government Advisory Committee to ensure proper governance for .health [24] was unsuccessful in overcoming the arguments that such intervention would be unnecessarily coercive. It *is*, however, possible to make the case for intervening against this hostile backdrop by focusing on the following four arguments: (1) education is necessary but insufficient, (2) precedent, (3) network dynamics, and (4) justice.

Education Is Necessary but Insufficient

Education has always been, and will always be, a vital and ongoing part of public health campaigns. It will undoubtedly play a key role in improving the extent to which individual citizens are *resilient* and *resistant* to health MDI. However, there is mounting evidence that education alone is likely to be an insufficient "solution." Pluviano et al [41], for example, conducted an experiment where beliefs in the idea of vaccinations being linked to dangerous side effects and intentions to vaccinate a future child were measured before and after an educational intervention. They found that, at best, the educational interventions were ineffective and, at worst, had the unintended opposite effect of reinforcing inaccurate beliefs and reducing intentions to vaccinate. This may well be because the web and social media create fertile conditions for the spread of postmodernist beliefs that question the legitimacy of science and authority, and reject the idea of a single "truth" [42]. This would not only help explain why those that spread and accept MDI are unlikely to be persuaded by evidence, facts, and reasoning but also indicate that relying on education alone is becoming less effective over time due to the nature of the environment it is trying to control.

Precedent

Most of the arguments concerning whether it is acceptable for states to engage in what some may perceive as censorship to protect the public from potentially harmful information are not unique to health information. It is relevant, therefore, that there is precedent for taking a stronger approach to content moderation in areas other than health. For example, Zittrain and Palfrey [43], Brown and Cowls [44], and Macdonald et al [45] discuss various states that have successfully defied an early wave of "cyberlibertarianism" to block content in the name of national security or moral protection. Internet filtering that targets the websites of insurgents, extremists, and terrorists generally garners wide public support, as does the filtering of content that is perceived to be antithetical to accepted societal norms, such as pornographic content or hate speech [43]. Thus, all leading social media companies stipulate in their terms of service that terrorist content is forbidden and have, since 2016, collectively maintained a shared industry database of hashes (unique digital fingerprints) that identifies content produced in support of terrorism so that it can be removed as quickly as possible or, ideally, prevented from being posted at all. These interventions are far from unproblematic, but they clearly indicate that, in specific cases, for particular purposes, and within appropriate constraints, it *is* reasonable to make an ethically sound case for restricting the rights of individuals to post and access any sort of information unrestrictedly [44].

Network Dynamics

When considering the dynamics of information spreading and persuasion online, and especially on social media, it is questionable whether taking a relatively *laissez-faire* approach to online health information is actually an effective approach to protecting autonomy and pluralism. Let us begin by considering how information sources are selected online.

Trust in a source of health information is determined by a complex set of interacting factors [46], but a particularly influential factor is *perceived credibility* [47]. In the offline world, this is largely controlled by the gatekeeping function performed by clinicians. In an online world, however, this gatekeeping function is removed. This means that a far greater burden is placed on individual internet users to make their own judgements about credibility and to determine which sources they trust [48]. Several studies, including a particularly compelling one by Chen et al [49], have shown that individuals with lower eHealth literacy lack the skills necessary to determine the credibility of the source accurately, thus placing their trust in inappropriate sources of information, like social media posts. One potential reason for this is that social endorsement (ie, likes and shares) acts as a signal of perceived trustworthiness to those with lower eHealth literacy [48,50-52]. Thus, if social endorsement is having at least a minor impact on the extent to which individuals trust health information that they read online, it is questionable whether they are *genuinely* making an autonomous decision about which information to treat as credible and act upon.

Building on the previous phenomenon, the algorithms driving both search engine results and social media feeds prioritize posts or websites that lead to greater engagement. Human nature means that often these are posts and sites that are more consistent with already held beliefs, emotive or controversial [53]. OHMDI is considerably more likely to meet these "criteria" than scientific evidence-based medical information, meaning that OHMDI is far more likely to benefit from algorithmic amplification than content produced by reputable health sources. Agents who deliberately try to manipulate or confuse debates about health care are well aware of this phenomenon and exploit it to their advantage [54].

The combination of these examples of network dynamics amplifying OHMDI provide a robust rebuttal against the argument that pluralism is a universal good. Although providing all views with an equal platform might be justifiable or even desirable in some contexts, the benefits of this—for example, ensuring that all perspectives are heard or providing individuals with the opportunity to develop their own opinions—are less likely to apply in health care. Unlike in politics, the widespread practice of evidence-based medicine means that there is often a high degree of consensus on most common medical questions

(although this can change with time, and may be less true for more emergent fields). In some cases, the removal of gatekeepers and the presumption that all "voices" have an equal right to be heard has the potential to do considerable damage in the health care context, sometimes more than in the political context. This makes prioritizing diversity of opinions less justifiable and desirable, and instead creates the foundation for arguing that some "opinions" are more important than others and should (ethically) have a greater (rather than equal) opportunity to be heard. All beliefs are born equal, but some grow to become knowledge, whereas others remain mere opinions.

Justice

The threat posed by OHMDI is more closely reminiscent of the threat posed by infectious diseases than of the threat posed by individual "unhealthy" behaviors. OHMDI acts as a pathogen and spreads like a virus through the internet and social media, exposing all those who are susceptible, not just those who have autonomously decided to seek out "alternative" information [55]. We saw that this has recently led to speaking of an "*info*demic" in the context of the COVID-19 *epi*demic (emphasis added). Protecting those who are more susceptible to OHMDI (often those who have defining characteristics also associated with poorer health outcomes) is more about meeting the other aim of public health interventions—reducing health inequalities—than it is about paternalistically deciding what "is best" for society. In other words, it is a matter of justice.

As the Marmot Review *Fair Society, Healthy Lives* made clear, health inequalities do not arise by chance. They cannot be attributed simply to genetic makeup, "bad" unhealthy behavior, or difficulties in accessing medical care. Instead, social and economic differences in health status are caused by social and economic inequalities in society. Health inequalities that are preventable by reasonable means are unfair and putting them right is a matter of social justice [56]. From this perspective, group-level interventions can still be respectful and protective of individual autonomy in a Kantian-Rawlsian sense (ie, as an integration of freedom and responsibility). Autonomous agents can accept moral constraints (provided they are transparent [29]) out of respect for others' freedom, autonomy and dignity, and hence fairness and equal opportunity for all members of society [37].

Taken together, these four arguments—(1) that education is necessary but insufficient, (2) that there is precedent for acceptable state-led control of the internet content in other domains, (3) that network dynamics adversely affect the spread of accurate health information, and (4) that justice is best served by protecting those susceptible to OHMDI—justify working to overcome the ethical concerns associated with state-led intervention in the infosphere in the name of public health. Infosphere can be interpreted both minimally, as the whole informational environment, and maximally, as a synonym with reality. Infosphere here refers to the whole information environment, the whole system of information objects including all agents and patients, messages, their attributes, and mutual relations [57] (see also [58]). Yet, this justification in itself may not be enough to warrant attention from public health bodies if they do not see the infosphere as falling within their sphere of influence.

The Infosphere as Social Determinant of Health

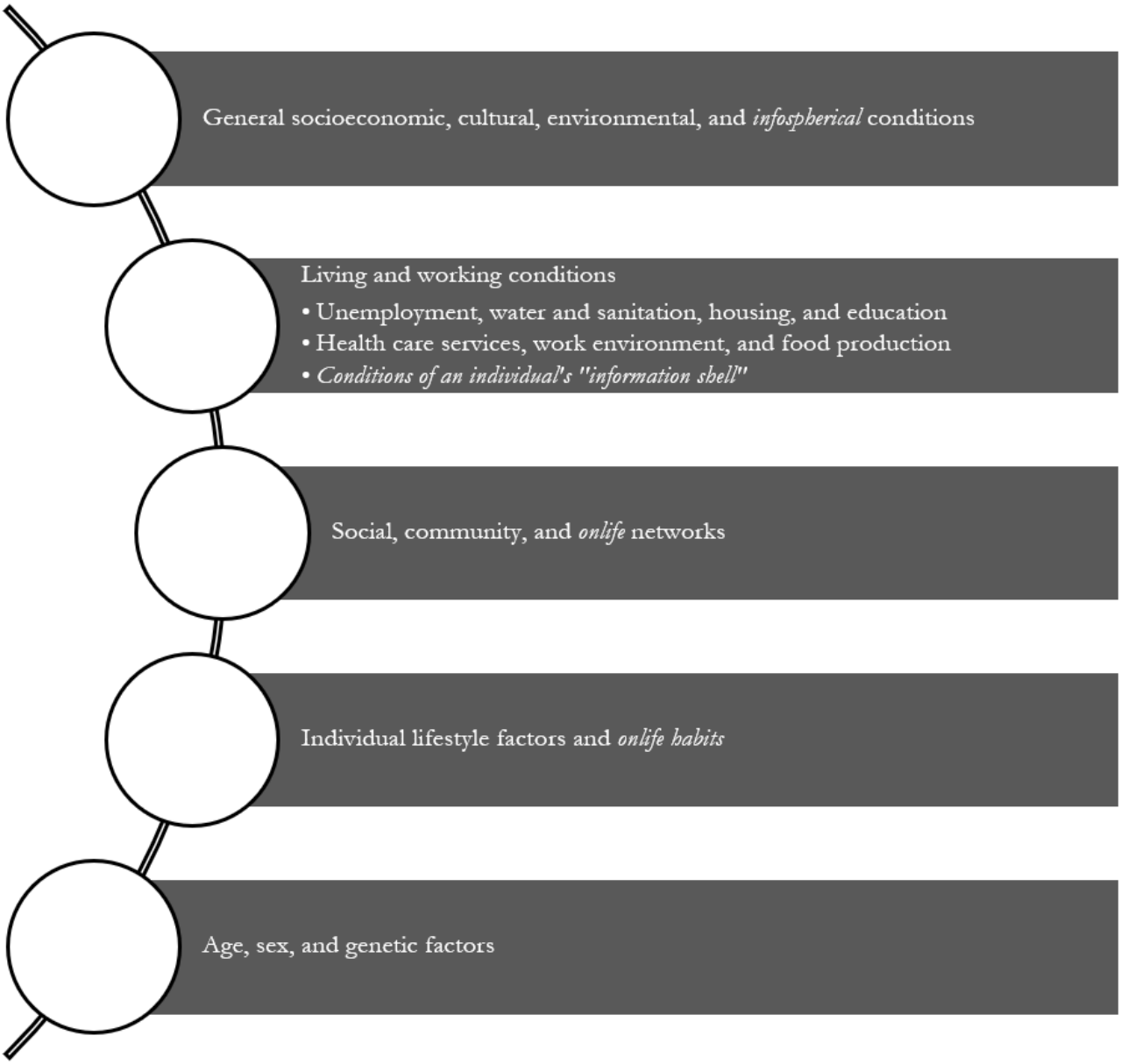
As discussed in the previous section, the controversy surrounding the debate over whether it is ethically acceptable or even preferable to limit individual rights to protect the group is certainly one factor hindering state-led intervention on OHMDI. However, public health policy makers are not typically afraid of controversy. Often, public health interventions, for example the UK's sugar tax, require public health bodies to confront powerful private corporations and frequently face public backlash. Both the moves to ban smoking in public spaces and to make wearing a seatbelt a legal requirement in the United Kingdom were socially controversial at the time, and yet, both are now widely accepted social norms [59]. It is possible, therefore, that this controversy is not the primary source of public health bodies' hesitancy. Instead, it may be that public health bodies have not intervened because they do not see online information control as being within their remit. Therefore, before we move on, we must consider the boundaries of public health bodies' remit.

To do this, we can turn to the 1978 Declaration of Alma-Ata, which is generally considered to be a major milestone in the field of public health. It described public health as "a social goal whose realisation requires the action of many social and economic sectors in addition to the health sector." Over time, these social and economic circumstances that together influence health came to be known as the *social determinants of health* (SDOH). The SDOH are alternatively referred to as the wider determinants of health. This approach to public health became a global commitment that themed the World Health Conference on October 21, 2011, when the WHO adopted the Rio Political Declaration on SDOH [60]. As a result, contemporary public health policies and strategies are determined by analyzing variations in SDOH and how these lead to inequalities in health care and threats to public health [61]. The problem for public health bodies operating today is that, although the SDOH approach provides global public health bodies with a much broader and more flexible remit than they had previously, the social determinants themselves have not been updated for the information age. Instead, almost all SDOH theories today are still based on the Dahlgren and Whitehead [61] model of health determinants developed for the WHO in 1991.

This model covers the *biosphere* and the *social sphere* but not the *infosphere.* This lack of attention to the infosphere is understandable, if one considers that this model was developed before the World Wide Web and declining costs of personal computers had enabled the global expansion of internet access [62], when social media and mobile apps did not exist, and when the dominant rhetoric was that online as a space had no bearing on physical realities. However, today the boundaries between online and offline are considerably less distinct and people living in the information age do live *onlife* [37,63]. It seems time that public health bodies accepted that the infosphere (encompassing all sources of online information) has a definite

determining influence on the public's physical health. Here we use the word physical to mean observable or demonstrable rather than "bodily" health. Undoubtedly, OHMDI will negatively affect individual's mental health as well as their bodily health and both are equally important. Acknowledging this influence would make it possible to consider the infosphere as an SDOH and adapt the Dahlgren and Whitehead [61] model accordingly, as shown in Figure 1 adapted from [57,61].

Figure 1. Onlife determinants of health adapted from Dahlgren and Whitehead (1991) [61] with italics highlighting new elements. The shell is a person's personal world of information. It is constantly evolving through time and has a significant influence on a person's behavior [57].



In this model, the determinants at the top (general socioeconomic, cultural, environmental, and now *informational* conditions) are those over which public health bodies have the greatest degree of influence. In contrast, the determinants at the bottom (age, sex, and genetic factors) are those over which public health bodies have little to no influence. Thus, the model describes the remit of public health bodies and anticipates the range of activities these bodies might decide are necessary to improve the public's health.

We can now conclude that not only is it possible to overcome the ethical concerns regarding individual autonomy vs group-level protection to justify government-led control of online health information, but also doing so definitely *is* within the remit of public health bodies. Having reached this conclusion, we must now move to consider *what* public health bodies can actually do to promote online health information.

Prevention, Protection, and Promotion: an Action Ontology

In the previous sections, we had to distinguish between public health and clinical health before we could assess the ethical arguments used against state-led intervention in the infosphere and then identify the remit of public health before we could determine whether the conditions of the infosphere fell within it. Similarly, we now must consider *what* type of actions public health bodies can and do typically take before we can consider

how these types of actions might be adapted for the infosphere context and the onlife experience.

Typically, public health bodies that operate at both a national or international scale conduct monitoring activities that enable them to identify public health *threats* such as air quality or pathogens that have the potential to cause harm. Depending on the threat level, responses can include issuing advice on *how* the public can keep themselves well or putting in place emergency measures such as the closing of airports to stop the spread of infectious disease in keeping with the 2005 International Health Regulations [64]. In short, almost all public health activities fall into one of the following three categories [65]:

1. Prevention: reducing the incidence of ill health by supporting healthier lifestyles
2. Protection: surveillance and monitoring of infectious disease, emergency responses, and vaccinations
3. Promotion: health education and commissioning services to meet specific health needs, for example, occupational health programs that promote self-care

Public Health England's 2019 prevention guidance, for example, aims to address SDOH by breaking them down into the protective and risk factors [66] listed in Textbox 3.

Textbox 3. Positive and negative influences on a person's health across the life course [66]. These are illustrative examples only, there are many other factors that could be listed in both categories.

Protective factors

- Having a healthy and balanced diet
- An environment that enables physical activity
- Good educational attainment
- Being in stable employment with a good income
- Living in good quality housing
- Having networks of support including family and friends

Risk factors

- Smoking
- Adverse childhood experiences
- Crime and violence
- Drug and alcohol misuse
- Poor educational attainment
- Poor mental health

To make these types of activities relevant to the current discussion, we must take the previous step of arguing that, if the quality of the infosphere is an SDOH, then, poor information quality (OHMDI) is also a public health threat within the infosphere [8,67], just as poor air or water quality are public health threats within the biosphere. Thus, we must frame it as the potential source of an infodemic or a digital pandemic [14,68].

In the early 1990s, when (as discussed) the potential for this threat was first foreseen, there was insufficient evidence to support those claiming that online health information of poor quality could cause genuine harm to people's health. This made it difficult to classify OHMDI officially as a threat to public health and use this as a means of demanding a public response [69]. This is no longer the case. Since the mid-2000s, the evidence demonstrating that the content people access online can affect significantly their health behaviors [70] has been growing. Researchers have connected proanorexia content with the rise in eating disorder incidence [71], antivaccination messages with a loss of trust in public vaccination programs [72], celebrity endorsements with mass uptake of fad diets and increased reliance on homeopathy and naturopathy over clinical intervention [9], and participation in specific chatrooms with suicidal ideation [73]. Furthermore, research has also shown that the creation of this potentially harmful content is not always unintentional. In 2007, the Wikipedia community identified a pharmaceutical company that was editing articles on Wikipedia and deleting side effects of specific medications [74]. Broniatowski et al [54] found that the Russian Internet Research Agency was using the hashtag #VaccinateUS to promote political discord. Therefore, it is clear that there is now an evidenced need to treat OHMDI as a public health threat and demand a robust and coordinated response [75], particularly as repeated exposure to potentially harmful information increases the risk that it poses [76].

Having completed this preliminary step, we can now return to the "prevent, protect, and promote" activities of public health bodies and clarify that:

1. Actions that lead to the automatic blocking of content classed as posing the highest risk to public health are *preventative*.
2. Actions that lead to the monitoring of content on social media or the wider web and the subsequent removal of potentially harmful information are *protective*.

3. Actions that improve access to and visibility of high-quality information are *promotional*.

Having established this ontology for the mode-of-action, we can now move on to examining the specific types of actions within each of these categories that different agents within the infosphere-as-SDOH may take.

Actions and Agents

As the ultimate guardian of the public's health, state-led public health bodies should take on overarching responsibility for the infosphere, ensuring that it flourishes and *protects* and *promotes* public health by *preventing* the appearance of threats in the form of OHMDI. The internet and the OHMDI circulating on it should be seen as the primary locus of this responsibility. In this respect, it should be noted that of the three categories for internet control identified by Eriksson and Giacomello [77], two (access to the internet and functionality of the internet) are outside of public health's scope. However, it *is* possible and within scope for public health bodies to intervene in the third category of internet control: activity on the internet.

To make this clearer, it is possible to use the level of abstraction method of analysis. A LoA can be imagined as an interface that enables one to observe some aspects of a system analyzed while making other aspects opaque or indeed invisible. For example, one may analyze a house at the LoA of a buyer, of an architect, of a city planner, of a plumber, and so on. LoAs are common in computer science, where systems are described at different LoAs (computational, hardware, user-centered, etc). LoAs can be combined in more complex sets and can be, but are not necessarily always, hierarchical [78].

Public health bodies can regulate activity on the internet, by taking responsibility for interventions at the LoA *for* the web (LoA_{FOR}) while enabling (and regulating) OSPs to take responsibility for interventions at the LoA *in* the web [5]. Responsibilities also present themselves at the LoA *on* the web, but as these responsibilities primarily concern themselves with controlling access to the metadata about user activities online [5], which public health bodies already do, for example, in cases of digital epidemiologic surveillance, we do not discuss these here. At the LoA_{FOR}, public health bodies should develop programs of work focused on the following four areas identified by Chou et al [79]:

1. Defining the prevalence and trends of health MDI and identifying content for removal (protective monitoring)
2. Understanding what health MDI is shared and how it spreads so that it is possible to intervene earlier (preventive action)
3. Evaluating the reach and influence of high-risk health MDI (protective monitoring)
4. Developing and testing promotional responses

Collectively, these programs of work would enable public health bodies to monitor the most prevalent content being shared online, identify weaknesses in any current strategies, and detect new sources and causes of MDI before they result in significant harm [80]. To ensure the effectiveness of successful responses developed in (4) and based on the knowledge generated in (1), (2), and (3), public health bodies should leverage existing legal frameworks such as customer protection acts and laws governing health advertising [8] to determine when content is permissible (eg, the conditions under which celebrities or "influencers" are permitted to endorse and promote health- and wellness-related products such as "Skinny Tea") and impose fines and other sanctions when these conditions are not met or when an organization, state, or person is found to be deliberately misrepresenting the scientific consensus with the intention of causing harm [81].

When these existing legal frameworks are found lacking, public health bodies should consider new, primary, or secondary legislation to ensure the protection of the public's health [82]. They should also consider subsidies and tax breaks for OSPs that reflect social responsibility for public health in their terms of service *and* enforce these terms [81].

Finally, as Mackey and Nayyar [24] argued, global public health stakeholders should come together to rectify the mistakes of the past and recognize (as ICANN's Independent Objector did) that the right to health is a fundamental human right, and this should include access to accurate health information. This means that .health and other health-related gTLDs should be protected by ensuring that the appearance of OHMDI is prevented and high-quality information is promoted. Collective action by then WHO, health-related UN organizations, and national governments should be demanded to secure a safe space for the health-related internet that abides by ethical principles, practices, and rules that honor public health interests [24] and ensure that information located within this space is authentic, truthful, accurate, clear, impartial, and evidence-based as much as possible [83].

Public health bodies can take the responsibility for creating the frameworks within which they and partnered private companies can intervene in the infosphere in the name of public health. Importantly, this must be done in a way that mitigates potential ethical risks related to information and data, such as privacy infringement, as much as is possible by encouraging public health bodies to consider how their interventions will affect the rights of both users and the environment [5]. At the same time, public heath bodies should regularly check that they are compliant with the International Health Regulations (2005) [64], that is, responding to a pressing public or social need pursuing a legitimate aim, being proportionate to the legitimate aim, and being no more restrictive than is required to achieve the purpose sought by restricting the right. They should also ensure that their actions are underpinned by the foundational values of public health ethics: transparency, confidentiality, and community consent [82].

By acting in this way, public health bodies can minimize both the harms of poor infosphere conditions and the ethical risks associated with public health policy. However, this does not mean that OSPs are discharged of all responsibility. OSPs should take responsibility for what it is *in* the infosphere and regulate it. They should discharge this responsibility, as stressed by Perakslis and Califf [11], by identifying, detecting, responding to, and recovering from OHMDI, and protecting accurate

information (the five core cybersecurity functions listed by The National Institute of Standards and Technology Cybersecurity Framework [11]) from damage, destruction, misuse, and corruption.

There is already evidence that some OSPs are taking tentative steps in this direction. In November 2016, Facebook banned misinformation from advertisements on the site, including those promoting antivaccination messages. Pinterest has banned all antivaccination content outright. YouTube has removed advertising revenue and monetization from antivaccine channels and videos [81]. Twitter (at least in the United Kingdom) signposts users to the NHS website first when they search for derivates of "vaccine" on the website or app. The current COVID-19 pandemic has also sparked some specific action: Apple is rejecting all coronavirus-related apps that are not from governments or official health organizations; Google Play is blocking all searches for coronavirus; the UK government has set up a Rapid Response Unit to directly respond to false coronavirus narratives by, for example, issuing a rebuttal on social media or asking platforms to remove harmful content [84,85]; and, still in the United Kingdom, the Culture Secretary, Oliver Dowden, asked platforms to be more aggressive in contrasting conspiracy theories linking the coronavirus pandemic to 5G networks. Although positive steps, these measures are sporadic, *ad hoc*, unsystematic, and often far too narrow to make a considerable impact, especially when the far more extensive measures taken to tackle copyright-related, pornographic, or terrorist content are considered. Health-related information is hardly ever less, and often considerably more, important. Only by taking proactive, coordinated measures will OSPs, public health bodies, and app-store providers be able to stay one-step ahead of the rapidly evolving conditions of the infosphere and play their role in protecting public health [8].

Conclusion

The WHO and the United Nations International Children's Fund may have become aware of the effects of deliberate disinformation campaigns before these spread online, with there being a notable coordinated campaign deployed to discourage women from receiving the neonatal tetanus vaccine throughout the late 1980s [86]. However, in the more than 3 decades since, frustratingly little has been done by public health bodies to tackle the internet's ever-growing role in the "medical misinformation mess" [6]. As we have shown, this is also because, in the past, there has not been a strong enough case for prioritizing societal interests over individual rights [69] in this context. In the face of the "rising tide of medical misinformation" [75] and the adverse effect it is having on global health, a different approach should be adopted.

The problems with online health information, its quality, impact, and control, that we have discussed here are complex and multifaceted [87]. However, we have argued that the nature of the infosphere, its governance (or lack thereof), structures, affordances, and content is certainly contributing to new public health harms. We need to change the strategy. Before the situation becomes completely untenable and unmanageable, a robust and coordinated response from public health bodies, private corporations, and individuals is reasonable, ethically justified, and pragmatically necessary. It is also technically feasible. Although OSPs do bear responsibility as to the content circulating in the infosphere, it is up to public health bodies to shape, foster, and implement the necessary policies and actions to curtail the spreading of OHMDI.

If this joint response can be coordinated effectively, and the infosphere is appropriately recognized as a social determinant of (public) health and, therefore, a public good [75,88], then the twin goals of protecting the public's health and reducing health inequalities can be supported. Identifying and implementing the most appropriate and efficacious interventions that fall within this framework may not be easy, but we should not let the scale of the challenge become a deterrent. Decisive action is needed, and it is needed as soon as possible.

Conflicts of Interest

None declared.

References

1. Floridi. Brave.Net.World: the Internet as a disinformation superhighway? The Electronic Library 1996 Jun;14(6):509-514 [FREE Full text] [doi: 10.1108/eb045517]
2. Floridi L. Information: A Very Short Introduction. Oxford, England: Oxford University Press; Mar 26, 2010.
3. Frish Y, Greenbaum D. Is social media a cesspool of misinformation? Clearing a path for patient-friendly safe spaces online. Am J Bioeth 2017 Mar;17(3):19-21. [doi: 10.1080/15265161.2016.1274795] [Medline: 28207361]
4. Boerman S, Kruikemeier S, Zuiderveen Borgesius FJ. Online behavioral advertising: a literature review and research agenda. J Advertising 2017 Jun 29;46(3):363-376. [doi: 10.1080/00913367.2017.1339368]
5. Taddeo M, Floridi L. The debate on the moral responsibilities of online service providers. Sci Eng Ethics 2016 Dec;22(6):1575-1603. [doi: 10.1007/s11948-015-9734-1] [Medline: 26613596]
6. Ioannidis J, Stuart ME, Brownlee S, Strite SA. How to survive the medical misinformation mess. Eur J Clin Invest 2017 Nov;47(11):795-802. [doi: 10.1111/eci.12834] [Medline: 28881000]
7. Floridi L. A look into the future impact of ICT on our lives. Inf Soc 2007 Feb 24;23(1):59-64. [doi: 10.1080/01972240601059094]

8. McKee M, Middleton J. Information wars: tackling the threat from disinformation on vaccines. BMJ 2019 May 13;365:l2144. [doi: 10.1136/bmj.l2144] [Medline: 31085543]
9. Collier R. Containing health myths in the age of viral misinformation. CMAJ 2018 May 14;190(19):E578 [FREE Full text] [doi: 10.1503/cmaj.180543] [Medline: 29759962]
10. Ghenai A, Mejova Y. Fake cures: user-centric modeling of health misinformation in social media. 2018 Nov Presented at: ACM Conference on Computer Supported Cooperative Work and Social Computing; 2018; Jersey City, New Jersey p. 1-20 URL: https://arxiv.org/abs/1809.00557 [doi: 10.1145/3274327]
11. Perakslis E, Califf RM. Employ cybersecurity techniques against the threat of medical misinformation. JAMA 2019 Jul 16;322(3):207-208. [doi: 10.1001/jama.2019.6857] [Medline: 31199429]
12. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
13. Rothkopf D. When the buzz bites back. The Washington Post. 2003 May 11. URL: http://www1.udel.edu/globalagenda/2004/student/readings/infodemic.html [accessed 2020-07-20]
14. Seymour B, Getman R, Saraf A, Zhang LH, Kalenderian E. When advocacy obscures accuracy online: digital pandemics of public health misinformation through an antifluoride case study. Am J Public Health 2015 Mar;105(3):517-523. [doi: 10.2105/ajph.2014.302437]
15. Gagliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. BMJ 2002 Mar 09;324(7337):569-573 [FREE Full text] [doi: 10.1136/bmj.324.7337.569] [Medline: 11884320]
16. Lewis T. Seeking health information on the internet: lifestyle choice or bad attack of cyberchondria? Media Cult Soc 2016 Jun 30;28(4):521-539. [doi: 10.1177/0163443706065027]
17. Terry N. Education and debate: regulating health information: a US perspective. BMJ 2002 Mar 09;324(7337):602-606 [FREE Full text] [doi: 10.1136/bmj.324.7337.602] [Medline: 11884330]
18. Theodosiou L, Green J. Emerging challenges in using health information from the internet. Adv Psychiatr Treat 2018 Jan 02;9(5):387-396. [doi: 10.1192/apt.9.5.387]
19. Watson R. European Commission to publish a code of practice for websites. BMJ 2002 Mar 09;324(7337):567 [FREE Full text] [doi: 10.1136/bmj.324.7337.567] [Medline: 11884316]
20. Morris K. WHO sets dot-health in its sights. Lancet 2000 Nov;356(9242):1664. [doi: 10.1016/s0140-6736(05)70377-9]
21. Brown P. WHO calls for a health domain name to help consumers. BMJ 2002 Mar 09;324(7337):566 [FREE Full text] [doi: 10.1136/bmj.324.7337.566/a] [Medline: 11884314]
22. Illman J. WHO s plan to police health websites rejected. BMJ 2000 Nov 25;321(7272):1308-1308. [doi: 10.1136/bmj.321.7272.1308/b]
23. Solomonides A. ICANN, health information and the "Dot Health" top level domain. 2014 Presented at: IEEE 27th International Symposium on Computer-Based Medical Systems; 2014; New York, USA. [doi: 10.1109/cbms.2014.148]
24. Mackey T, Liang BA, Kohler JC, Attaran A. Health domains for sale: the need for global health Internet governance. J Med Internet Res 2014 Mar 05;16(3):e62-e26 [FREE Full text] [doi: 10.2196/jmir.3276] [Medline: 24598602]
25. Purcell G, Wilson P, Delamothe T. The quality of health information on the internet. BMJ 2002 Mar 09;324(7337):557-558 [FREE Full text] [doi: 10.1136/bmj.324.7337.557] [Medline: 11884303]
26. Eysenbach G. The new health-related top-level domains are coming: will cureforcancer.health go to the highest bidder? J Med Internet Res 2014 Mar 05;16(3):e73 [FREE Full text] [doi: 10.2196/jmir.3358] [Medline: 24598704]
27. Gostin L. Public health law: a renaissance. J Law Med Ethics 2002;30(2):136-140. [doi: 10.1111/j.1748-720x.2002.tb00379.x] [Medline: 12066591]
28. Kass NE. An ethics framework for public health. Am J Public Health 2001 Nov;91(11):1776-1782. [doi: 10.2105/ajph.91.11.1776] [Medline: 11684600]
29. Buchanan D. Autonomy, paternalism, and justice: ethical priorities in public health. Am J Public Health 2008 Jan;98(1):15-21. [doi: 10.2105/ajph.2007.110361]
30. Wiley L, Berman ML, Blanke D. Who's your nanny? Choice, paternalism and public health in the age of personal responsibility. J Law Med Ethics 2013 Mar;41 Suppl 1:88-91. [doi: 10.1111/jlme.12048] [Medline: 23590750]
31. Jennings B. Public health and liberty: beyond the Millian paradigm. Public Health Ethics 2009 Jun 09;2(2):123-134. [doi: 10.1093/phe/php009]
32. Petrini C. Theoretical models and operational frameworks in public health ethics. Int J Environ Res Public Health 2010 Jan;7(1):189-202 [FREE Full text] [doi: 10.3390/ijerph7010189] [Medline: 20195441]
33. Lalonde M. A new perspective on the health of Canadians. Government of Canada. 1974. URL: https://www.canada.ca/en/public-health/services/health-promotion/population-health/a-new-perspective-on-health-canadians.html [accessed 2020-07-20]
34. Beauchamp D, Steinbeck B. New Ethics for the Public's Health. Oxford, UK: Oxford University Press; Jul 29, 1999.
35. Childress JF, Faden RR, Gaare RD, Gostin LO, Kahn J, Bonnie RJ, et al. Public health ethics: mapping the terrain. J Law Med Ethics 2002 Jun 24;30(2):170-178. [doi: 10.1111/j.1748-720x.2002.tb00384.x] [Medline: 12066595]
36. Jennings B, Khan J, Mastroianni A, Parker L. Ethics and public health: model curriculum. Georgetown University. 2003 Jul. URL: http://hdl.handle.net/10822/556779 [accessed 2020-07-20]

37. Floridi L. Tolerant paternalism: pro-ethical design as a resolution of the dilemma of toleration. Sci Eng Ethics 2016 Dec;22(6):1669-1688. [doi: 10.1007/s11948-015-9733-2] [Medline: 26649432]
38. Morley J, Floridi L. The limits of empowerment: how to reframe the role of mHealth tools in the healthcare ecosystem. Sci Eng Ethics 2020 Jun;26(3):1159-1183 [FREE Full text] [doi: 10.1007/s11948-019-00115-1] [Medline: 31172424]
39. Nieuwenhuis A. The Concept of Pluralism in the case law of the ECtHR. Eur Constitutional Law Rev 2007 Dec 18;3(3):367-384. [doi: 10.1017/s1574019607003677]
40. A 2019 snapshot: a brief look back at the last year at Public Health England. Public Health England. 2019. URL: https://publichealthengland.exposure.co/a-2019-snapshot [accessed 2020-07-20]
41. Pluviano S, Watt C, Della Sala S. Misinformation lingers in memory: failure of three pro-vaccination strategies. PLoS One 2017;12(7):e0181640 [FREE Full text] [doi: 10.1371/journal.pone.0181640] [Medline: 28749996]
42. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. Vaccine 2010 Feb 17;28(7):1709-1716. [doi: 10.1016/j.vaccine.2009.12.022] [Medline: 20045099]
43. Zittrain J, Palfrey J. Internet filtering: the politics and mechanisms of control. In: Deibert R, Palfrey J, Rohonzinski R, Zittrain J, editors. Access Denied: The Practice and Policy of Global Internet Filtering. Cambridge, Massachusetts: The MIT Press; 2008.
44. Brown I, Cowls J. Check the web: assessing the ethics and politics of policing the Internet for extremist material. Vox-Pol. 2015 Nov 23. URL: https://www.voxpol.eu/check-the-web/ [accessed 2020-07-20]
45. Macdonald S, Correia SG, Watkin A. Regulating terrorist content on social media: automation and the rule of law. Int J Law Context 2019 Jun 20;15(2):183-197. [doi: 10.1017/s1744552319000119]
46. Lupton D. 'Better understanding about what's going on': young Australians' use of digital technologies for health and fitness. Sport Education Soc 2018 Dec 06;25(1):1-13. [doi: 10.1080/13573322.2018.1555661]
47. Paige S, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C. Proposing a transactional model of eHealth literacy: concept analysis. J Med Internet Res 2018 Oct 02;20(10):e10175 [FREE Full text] [doi: 10.2196/10175] [Medline: 30279155]
48. Borah P, Xiao X. The importance of 'Likes': the interplay of message framing, source, and social endorsement on credibility perceptions of health information on Facebook. J Health Commun 2018;23(4):399-411. [doi: 10.1080/10810730.2018.1455770] [Medline: 29601271]
49. Chen X, Hay JL, Waters EA, Kiviniemi MT, Biddle C, Schofield E, et al. Health literacy and use and trust in health information. J Health Commun 2018;23(8):724-734 [FREE Full text] [doi: 10.1080/10810730.2018.1511658] [Medline: 30160641]
50. Huo C, Zhang M, Ma F. Factors influencing people's health knowledge adoption in social media. Library Hi Tech 2018 Mar 19;36(1):129-151. [doi: 10.1108/lht-04-2017-0074]
51. Kim J. They liked and shared: effects of social media virality metrics on perceptions of message influence and behavioral intentions. Comput Hum Behav 2018 Jul;84:153-161. [doi: 10.1016/j.chb.2018.01.030]
52. Tworek C. Freaks, geeks, norms and mores: why people use the status quo as a moral compass. The Conversation. 2016. URL: https://theconversation.com/freaks-geeks-norms-and-mores-why-people-use-the-status-quo-as-a-moral-compass-59046 [accessed 2020-07-20]
53. Gonzalez-Bailon S. Online social networks and bottom-up politics. SSRN J 2013:209-222. [doi: 10.2139/ssrn.2246663]
54. Broniatowski D, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/ajph.2018.304567]
55. Jin F, Wang W, Zhao L, Dougherty E, Cao Y, Lu C, et al. Misinformation propagation in the age of Twitter. Computer 2014 Dec;47(12):90-94. [doi: 10.1109/mc.2014.361]
56. Marmot M, Bell R. Fair society, healthy lives. Public Health 2012 Sep;126 Suppl 1:S4-S10. [doi: 10.1016/j.puhe.2012.05.014] [Medline: 22784581]
57. Floridi L. On the intrinsic value of information objects and the infosphere. Ethics Inf Technol 2002;4(4):287-304 [FREE Full text] [doi: 10.1023/A:1021342422699]
58. Infosphere. Wikipedia. URL: https://en.wikipedia.org/wiki/Infosphere
59. Turoldo F. Responsibility as an ethical framework for public health interventions. Am J Public Health 2009 Jul;99(7):1197-1202. [doi: 10.2105/ajph.2007.127514]
60. World Health Organization. Rio political declaration on social determinants of health. 2011 Presented at: World Conference on Social Determinants of Health; 2011; Rio, Brazil URL: https://www.who.int/sdhconference/declaration/en/
61. Dahlgren G, Whitehead M. Policies and strategies to promote social equity in health. Institute for Further Studies. Stockholm; 1991. URL: https://core.ac.uk/reader/6472456 [accessed 2020-07-20]
62. Zittrain J. The Future of the Internet and How to Stop it. London, England: Yale University Press; 2008.
63. Floridi L. The 4th Revolution: How the Infosphere Is Reshaping Human Reality. Oxford: Oxford University Press; 2014.
64. Fidler D, Gostin LO. The new international health regulations: an historic development for international law and public health. J Law Med Ethics 2016 Jun 23;34(1):85-94. [doi: 10.1111/j.1748-720x.2006.00011.x]

65. Public health. Royal College of Nursing. 2020. URL: https://www.rcn.org.uk/clinical-topics/public-health [accessed 2020-07-20]
66. Health matters: prevention - a life course approach. Public Health England. 2019. URL: https://www.gov.uk/government/publications/health-matters-life-course-approach-to-prevention/health-matters-prevention-a-life-course-approach [accessed 2020-07-20]
67. Vogel L. Viral misinformation threatens public health. CMAJ 2017 Dec 18;189(50):E1567 [FREE Full text] [doi: 10.1503/cmaj.109-5536] [Medline: 29255104]
68. Larson H. The biggest pandemic risk? Viral misinformation. Nature 2018 Oct;562(7727):309. [doi: 10.1038/d41586-018-07034-4] [Medline: 30327527]
69. Rothstein M. Rethinking the meaning of public health. J Law Med Ethics 2002;30(2):144-149. [doi: 10.1111/j.1748-720x.2002.tb00381.x] [Medline: 12066593]
70. Goldacre B. Media misinformation and health behaviours. Lancet Oncol 2009 Sep;10(9):848. [doi: 10.1016/s1470-2045(09)70252-9]
71. Christodoulou M. Pro-anorexia websites pose public health challenge. Lancet 2012 Jan;379(9811):110. [doi: 10.1016/s0140-6736(12)60048-8]
72. Fernández-Luque L, Bau T. Health and social media: perfect storm of information. Healthc Inform Res 2015 Apr;21(2):67-73 [FREE Full text] [doi: 10.4258/hir.2015.21.2.67] [Medline: 25995958]
73. Tierney S. The dangers and draw of online communication: pro-anorexia websites and their implications for users, practitioners, and researchers. Eat Disord 2006;14(3):181-190. [doi: 10.1080/10640260600638865] [Medline: 16807213]
74. Lau A, Gabarron E, Fernandez-Luque L, Armayones M. Social media in health--what are the safety concerns for health consumers? Health Inf Manag 2012;41(2):30-35. [doi: 10.1177/183335831204100204] [Medline: 23705132]
75. Armstrong P, Naylor CD. Counteracting health misinformation: a role for medical journals? JAMA 2019 May 21;321(19):1863-1864. [doi: 10.1001/jama.2019.5168] [Medline: 31009036]
76. Shen C, Wang MP, Wan A, Viswanath K, Chan SSC, Lam TH. Health information exposure from information and communication technologies and its associations with health behaviors: Population-based survey. Prev Med 2018 Aug;113:140-146. [doi: 10.1016/j.ypmed.2018.05.018] [Medline: 29787775]
77. Eriksson J, Giacomello G. Who controls the internet? Beyond the obstinacy or obsolescence of the state. Int Stud Rev 2009;11(1):205-230 [FREE Full text] [doi: 10.1111/j.1468-2486.2008.01841.x]
78. Floridi L. The method of levels of abstraction. Minds Machines 2008 Aug 15;18(3):303-329. [doi: 10.1007/s11023-008-9113-7]
79. Chou W, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
80. Ghenai A. Health misinformation in search and social media. 2017 Presented at: Proceedings of 40th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2017; Tokyo, Japan URL: https://dl.acm.org/doi/10.1145/3079452.3079483
81. Yang Y, Broniatowski DA, Reiss DR. Government role in regulating vaccine misinformation on social media platforms. JAMA Pediatr 2019 Sep 03:1011-1012. [doi: 10.1001/jamapediatrics.2019.2838] [Medline: 31479099]
82. Lee L, Heilig CM, White A. Ethical justification for conducting public health surveillance without patient consent. Am J Public Health 2012 Jan;102(1):38-44. [doi: 10.2105/ajph.2011.300297]
83. Solomonides AE, Mackey TK. Emerging ethical issues in digital health information. Camb Q Healthc Ethics 2015 Jun 10;24(3):311-322. [doi: 10.1017/s0963180114000632]
84. Dent S. Apple is rejecting coronavirus apps to limit virus misinformation. Engadget. 2020 Mar 06. URL: https://tinyurl.com/y42rdoll [accessed 2020-03-06]
85. Dowden O, Mordaunt P. Government cracks down on spread of false coronavirus information online. UK Cabinet Office, Department of Culture Media and Sport, Gov.uk. 2020 Mar 30. URL: https://www.gov.uk/government/news/government-cracks-down-on-spread-of-false-coronavirus-information-online [accessed 2020-03-30]
86. Milstien J, Griffin P, Lee J. Damage to immunisation programmes from misinformation on contraceptive vaccines. Reprod Health Matters 1995 Nov;3(6):24-28. [doi: 10.1016/0968-8080(95)90155-8]
87. Larson H, Jarrett C, Eckersberger E, Smith DMD, Paterson P. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007-2012. Vaccine 2014 Apr 17;32(19):2150-2159. [doi: 10.1016/j.vaccine.2014.01.081] [Medline: 24598724]
88. Taddeo M. Is cybersecurity a public good? Minds Machines 2019 Oct 5;29(3):349-354. [doi: 10.1007/s11023-019-09507-5]

Abbreviations

COVID-19: coronavirus disease
gTLD: general top-level domain
ICANN: Internet Corporation for Assigned Names and Numbers
LoA: level of abstraction

LoA_{FOR}: level of abstraction for the web
MDI: mis- and disinformation
NHS: National Health Service
OHMDI: online health-related misinformation and disinformation
OSP: online service provider
SDOH: social determinants of health
TLD: top-level domain
UN: United Nations
WHO: World Health Organization

Edited by G Eysenbach; submitted 13.04.20; peer-reviewed by A Sheon, I Mircheva; comments to author 08.06.20; revised version received 11.06.20; accepted 08.07.20; published 03.08.20.

Please cite as:
Morley J, Cowls J, Taddeo M, Floridi L
Public Health in the Information Age: Recognizing the Infosphere as a Social Determinant of Health
J Med Internet Res 2020;22(8):e19311
URL: https://www.jmir.org/2020/8/e19311
doi:10.2196/19311
PMID:32648850

Viewpoint

Data Heterogeneity: The Enzyme to Catalyze Translational Bioinformatics?

Eli M Cahan[1,2], BBA; Purvesh Khatri[1,3], PhD

[1]Department of Medicine, School of Medicine, Stanford University, Stanford, CA, United States

[2]School of Medicine, New York University, New York, NY, United States

[3]Department of Biomedical Data Sciences, School of Medicine, Stanford University, Stanford, CA, United States

Corresponding Author:
Purvesh Khatri, PhD
Department of Medicine
School of Medicine
Stanford University
1265 Welch Road
Medical School Office Building, X219
Stanford, CA, 94305
United States
Phone: 1 650 497 5281
Email: pkhatri@stanford.edu

Abstract

Up to 95% of novel interventions demonstrating significant effects at the bench fail to translate to the bedside. In recent years, the windfalls of "big data" have afforded investigators more substrate for research than ever before. However, issues with translation have persisted: although countless biomarkers for diagnostic and therapeutic targeting have been proposed, few of these generalize effectively. We assert that inadequate heterogeneity in datasets used for discovery and validation causes their nonrepresentativeness of the diversity observed in real-world patient populations. This nonrepresentativeness is contrasted with advantages rendered by the solicitation and utilization of data heterogeneity for multisystemic disease modeling. Accordingly, we propose the potential benefits of models premised on heterogeneity to promote the Institute for Healthcare Improvement's Triple Aim. In an era of personalized medicine, these models can confer higher quality clinical care for individuals, increased access to effective care across all populations, and lower costs for the health care system.

(J Med Internet Res 2020;22(8):e18044) doi:10.2196/18044

KEYWORDS

medical Informatics; health equity; health care disparities; population health; quality improvement; precision medicine

Background

Philosopher Karl Popper commented in 1934 that "non-reproducible single occurrences are of no significance to science" [1]. Yet, 85 years since this statement was made, science remains inundated with nonreproducible single occurrences. John Ioannidis famously wrote in 2005 that "most published research is false" [2]. Chalmers and Glasziou [3] later quantified the false positive rate of published science at 85%; the false positive rates in translational medicine may be even higher than this estimate. Up to 89% of studies demonstrating significant preclinical effects of novel molecules are nonreplicable [4], and the translation failure rate of novel interventions demonstrating significant effects preclinically that are never approved for clinical use reaches up to 95% [5]. These ranges may themselves be underestimates, since they are based on molecules assessed by pharmaceutical companies and in studies published in the highest-impact journals. The translation failure rate of less promising molecules is likely higher still.

In recent years, the emergence of multidimensional "big data" has endowed clinician investigators with more plentiful research substrate than ever before. However, issues with translation have persisted: despite innumerable statistically significant biomarkers identified in the preclinical setting, few of these generalize effectively. For example, 0% of proposed biomarkers for rheumatoid arthritis have demonstrated generalizability [6]. In addition, since enormous samples contribute sufficient statistical power capable of offsetting minute effect sizes, increasingly voluminous data may cause translation failure to become more rather than less of an endemic problem. Indeed,

recent studies have noted a 36% deterioration of clinical effectiveness for molecules in Phase II trials [5].

We do not believe that the "depth" of samples (ie, cohort size) is responsible for the observed patterns in translation failure associated with big data. Rather, we believe that the problem is insufficient "breadth"; that is, the datasets used for discovery and validation fail to represent the diversity observed in distinctive real-world patient populations. In other words, by failing to represent the extent of real-world population diversity, we can define these datasets as inadequately *heterogeneous*.

There is already evidence for the effectiveness of translational bioinformatics premised on heterogeneity for conditions previously plagued by generalization failures, such as in the derivation of host response–based gene panels to predict sepsis and tuberculosis. These panels have outperformed all precedents developed without accounting for heterogeneity (including those using the most sophisticated machine-learning techniques); have been validated across time points, disease severity cohorts, and comorbidities; and have been generalizable across multiple continents [7-9].

In this paper, we highlight the tendency toward homogeneity in translational discovery and illuminate its negative implications. In contrast, we present heterogeneity as an ally rather than an enemy of meaningful translation. Finally, we describe the potential impact of incorporating heterogeneity into the process of translational bioinformatics for addressing the Institute for Healthcare Improvement's Triple Aim: facilitating personalized medicine, alleviating a health care cost crisis, and resolving health disparities [10,11].

Homogeneity Inherent to "Big" Translational Datasets

The core benefits of big data can be summarized in terms of volume (how much data are available), velocity (how quickly data are accumulated), and variety (how heterogenous the data are) [12]. Although the former two benefits have been harnessed extensively in translational research, the latter has not.

Datasets used for translational research may lack variety owing to three mechanisms: it may be absent, unevenly distributed, or inaccessible. The absence of variety results from constricted sourcing of data, leading to the funneling of homogenous features. One example is the exclusive use of healthy subjects for benchmarking, such as in immunocellular profiling for autoimmune disease [13]. The uneven distribution of variety within a dataset can lead to unintentional clustering of homogeneity, thereby filtering out heterogeneous characteristics. This is a digitized form of sampling bias: since heritability and penetrance both vary within populations, the findings in genome-wide association studies (GWAS) depend markedly upon the sampled cohorts [14]. Finally, variety may be present in the raw data but difficult to access, sequestering the heterogeneity due to technical hurdles. As dataset complexity increases, the risk of sequestration is amplified [15].

This becomes problematic in translational genomics, such as by producing "missing heritability" that is unexplainable from the processed dataset. It has been theorized that much of this "dark matter" (ie, the factors invisible in the processed dataset) relates to environmental influences. These environmental influences produce endophenotypes (expression profiles remaining latent until specific triggering exposures), which are epigenetic traits that can have strong contributions to phenotypic variation [16].

Homogenous datasets account poorly for differential environmental exposures and thus tend to be unreflective of transcriptomic diversity in broader populations. In turn, findings derived from such datasets may not extrapolate routinely beyond the experimental setting, thus precipitating translation failure.

Homogeneity Rendered From "Big" Translational Datasets

Alternatively, homogeneity may be intentionally selected for within the dataset. The contemporary system of science is lubricated by two forms of currency—financial and academic—both of which present disincentives to embracing heterogeneity. On the one hand, budgetary constraints make inclusive, comprehensive methodologies (for instance, preclinical validation studies on multiple animal cohorts) either impractical or unaffordable [13]. On the other hand, the relentless pursuit of academic currency (reputation, garnered through publication) is more easily facilitated by exclusive, narrow methodologies. The inflation of effect sizes is readily conjured in well-controlled experimental populations subjected to investigator-dependent research methods [17].

This investigator-dependent variability—which produces what has been deemed the "vibration of effects"—fosters significant interstudy dissimilarity [17]. Investigator choices can fragment broad baseline populations into discrete clusters subjected to inconsistent exposures to create unbalanced terminal populations [7]. As Kaptchuk [18] pointed out:

> *Facts do not accumulate on the blank slates of researchers' minds and data simply do not speak for themselves…[the] evaluative process is never totally objective or completely independent of scientists' convictions or theoretical apparatus.*

Accordingly, one way to reframe the reproducibility crisis is as an *exclusivity* crisis. Intrinsic homogeneity (native to datasets) compounded by extrinsic homogeneity (rendered to datasets) yields a sort of "private epidemiology," in which discrete study clusters are nonrepresentative of clinical diversity. This has been observed both in vitro and in vivo, where physiologic models poorly recapitulate real-world biology; up to 100% of findings based on observational data (such as vast catalogs of genetic signals) are not replicable [2,5,19]. Poor reproducibility has also been observed in silico, as predictive models premised on these limited feature sets have low external validity [12].

In short, the forces molding experimental homogeneity sculpt what become N-of-none studies. These are reflective of realities contained neatly within digital cells in spreadsheets rather than organic realities in patients.

Heterogeneity in Translational Big Data: Today

More vivid depictions of organic (rather than spreadsheet) realities can be drawn from the introduction of heterogeneity to translational bioinformatics. Heterogeneity expands the analytical spectrum beyond the monochromatic shades of homogenous datasets to better represent real-world phenomena.

Just as meta-analyses mediate between-study biases in evaluation of treatment effects, the introduction of heterogeneity similarly allows for mediation of between-sample biases. Crucially, heterogeneity does not eliminate differences but rather synthesizes similarities [15]. The utility of heterogeneity comes from deriving commonality across diverse subgroups by including rather than excluding distinctive features.

This adheres to theories of systems biology (beyond Oslerian pathophysiology), which contextualize biological interactions in dynamic settings. Robust evidence has documented the inconsistent behavior of unique biological entities (ie, genomic, proteomic, and transcriptomic) "longitudinally" across time points and "latitudinally" across milieu [20]. Accordingly, cross-sectional studies in well-controlled samples seem to be ill-suited for explaining—much less, solving—polygenic diseases and polymechanistic syndromes.

Heterogeneity may be imputed experimentally by casting a wide net of investigators or of data samples. For the former, crowd-sourced collaboration has improved translational efforts compared with independent analyses across multiple indications (Table 1).

For the latter, construction of diverse datasets has yielded durable findings relevant for translation across numerous disorders previously plagued by false positives (Table 2). Protocols for introduction of heterogeneity by the use of multiple datasets are publicly available [21].

Table 1. Illustrative applications of crowd-sourced heterogeneity.

Title	Author	Year	Indication
Crowdsourced assessment of common genetic contribution to predicting anti-TNF treatment response in rheumatoid arthritis	Sieberts et al [22]	2016	Rheumatoid arthritis
Crowdsourced estimation of cognitive decline and resilience in Alzheimer's disease	Allen et al [23]	2016	Alzheimer disease
Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: development of a prognostic model through a crowdsourced challenge with open clinical trial data	Guinney et al [24]	2017	Prostate cancer
A community approach to mortality prediction in sepsis via gene expression analysis	Sweeney et al [25]	2018	Sepsis

Table 2. Illustrative applications of user-constructed heterogeneity.

Title	Author	Year	Indication
Leveraging heterogeneity across multiple datasets increases cell-mixture deconvolution accuracy and reduces biological and technical biases	Vallania et al [26]	2018	Autoimmune disease (systemic lupus erythematosus)
Identification of a common gene expression signature in dilated cardiomyopathy across independent microarray studies.	Barth et al [27]	2006	Cardiomyopathy
A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation.	Khatri et al [28]	2013	Organ transplantation
Robust classification of bacterial and viral infections via integrated host gene expression diagnostics.	Sweeney et al [29]	2016	Upper respiratory infection
A community approach to mortality prediction in sepsis via gene expression analysis	Sweeney et al [24]	2018	Sepsis
Integrated, multi-cohort analysis identifies conserved transcriptional signatures across multiple respiratory viruses.	Andres-Terre et al [30]	2015	Influenza
Integrated multi-cohort transcriptional meta-analysis of neurodegenerative diseases	Li et al [31]	2014	Neurodegenerative disease
Integrated, multicohort analysis of systemic sclerosis identifies robust transcriptional signature of disease severity.	Lofgren et al [32]	2016	Systemic sclerosis
Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis	Sweeney et al [8]	2016	(Pulmonary) tuberculosis
Meta-analysis of continuous phenotypes identifies a gene signature that correlates with COPD disease status.	Scott et al [33]	2017	Chronic obstructive pulmonary disease (COPD)
A comprehensive time-course–based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set	Sweeney et al [34]	2016	Sepsis

Benefits to these strategies are exemplified by the studies mentioned in the Background section addressing tuberculosis and sepsis, respectively. The imputation of heterogeneity allowed for a 3-gene tuberculosis panel to be generalizable across 10 African countries [8,35] and an 11-gene panel capable of forming distinctive sepsis patient clusters to be validated in multiple nations [36]. Both of these panels, with their ability to accurately guide care for diverse patient groups (within and between populations), symbolize truly personalized medicine [7].

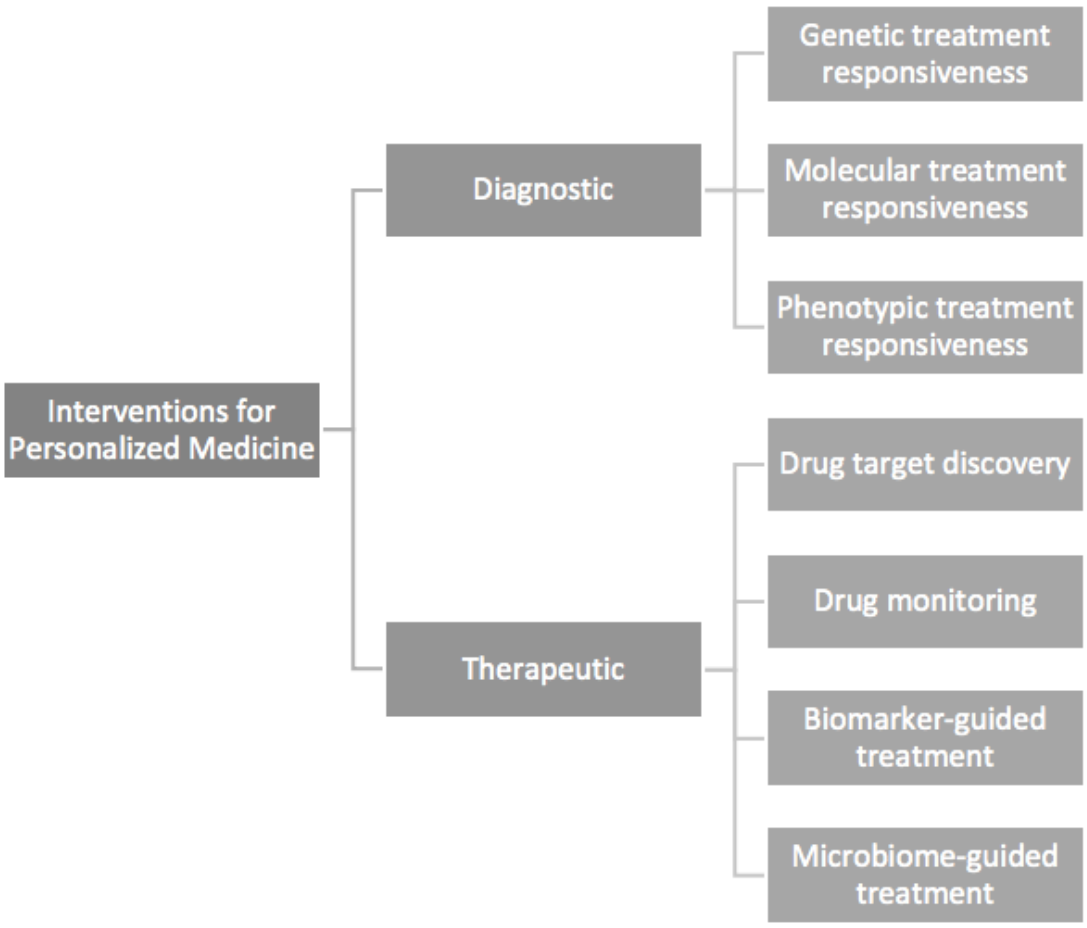
Heterogeneity in Translational Big Data: Tomorrow

General Prospects

Looking toward the future, the use of heterogeneity may play a prominent role in the advancement of translational bioinformatics by cultivating generalizability as a byproduct of representativeness. This bears substantial potential at the discovery stage, during which statistical significance is useful but not sufficient for predicting clinical effectiveness [2,19]. A myriad of diagnostic/prognostic and therapeutic modalities are being actively investigated for translation of personalized medicine, and validation will be crucial to distinguish the wheat from the chaff (Figure 1). Validation of novel diagnostics/prognostics (such as biomarkers) stands to benefit from heterogeneity given the aforementioned patient diversity across longitudinal and latitudinal scenarios [20]. Validation of novel therapeutics benefits from heterogeneity by enrichment of preclinical and clinical trials [37].

Establishment of data inclusiveness standards to supplement existing research guidelines (such as ARRIVE for preclinical studies and STROBE for observational studies) can accelerate the uptake of heterogeneity into best practices. The assimilation of heterogeneity into research practice in turn bears implications on personalized medicine, health care costs, and health disparities.

Figure 1. Modalities currently under investigation using translational bioinformatics to promote personalized medicine.



Personalized Medicine

Leveraging heterogeneity in translational medicine may offer the quickest path to personalized medicine. It has been noted that increasing the number of datasets included in GWAS samples, controlling for sample size, markedly improves the predictive power of the obtained gene panels to a much greater extent than expanding the sample size alone [15].

This model also incorporates "dark matter" contributing to "missing heritability," permitting the parsimonious identification of key biological pathways in spite of environmental differences between patient cohorts [16,38]. Moreover, observed differences may be informative rather than confounding: outliers bilaterally (such as weak or strong responders to interventions) are instructive and fertile sources for future investigation rather than "negligible." N-of-one study becomes feasible within this paradigm.

Finally, while heterogeneity is not necessarily a panacea for discovery—studies utilizing heterogeneity to address acute respiratory distress syndrome have failed to find robust biomarkers—the utility of negative findings is bolstered by the methodology [39]. Evidence-of-absence investigations benefit greatly from additional rigor that more conclusively redirects researchers toward clinically meaningful prospects [13].

Health Care Costs

Health care costs may be targeted from the sides of supply and demand alike. On the supply side, from the perspective of pharmaceutical companies, improved replicability of novel molecules reduces research and development costs devoted toward validation studies, which are currently estimated in the millions of dollars per agent tested [5]. Theoretically, this can allow for reduction in prices with preservation of profit margins. On the demand side, from the payor perspective, improved generalizability first enhances the cost-effectiveness of covered interventions, as clinical effects approach experimental effects [14]. Additionally, more reliable evidence-of-absence studies empower decision making for minimization of overutilized, misutilized, and ineffective interventions [13]. Finally, better understanding of "outlier" pathophysiology can promote the optimal management of "hot spotters"; that is, the oft-cited 1% of the population accounting for 33% of expenditures [40].

Health Disparities

Reductions in payor costs, if passed on to consumers, improve the accessibility of health care. For example, the demonstration of predictive power for tuberculosis diagnosis using 3-gene rather than 71-gene panels implies marked reductions in testing costs (presuming proportional and consistent marginal costs). Furthermore, to the extent that technological barriers for 3-gene sequencing are lower, these diagnostics become available to populations outside of high-resource settings alone [8]. As long as more parsimonious models are adequately representative and maintain predictive power across population groups (as was the case in [8]), accuracy would be preserved in an equitable way while access is simultaneously enhanced.

Heterogeneity may also support the resolution of health disparities by virtue of inclusiveness. As previously discussed, multiplicity of sample sets benefits all populations, with disproportionately greater benefits for traditionally excluded populations [15]. In this way, channeling the "wisdom of crowds" refers not only to wisdom pulled by collaboration between investigators but also to the wisdom pushed by the comprehensiveness of study populations.

Conclusion

In summary, we believe that research practices premised on sample homogeneity are important drivers of shortcomings in contemporary bench-to-bedside informatics. We assert that introduction of heterogeneity can favorably bend this trajectory. Uptake promoted by informal research culture change and formal inclusiveness criteria can lead to meaningful, sustainable, and equitable patient care in the future. In other words, the heterogeneity ethos echoes Osler's original invocation for personalized medicine: "Just listen to the patient. He is telling you the diagnosis!"

Authors' Contributions

Both authors (EC and PK) equally contributed to conceptualization, editing, and finalizing the manuscript. EC drafted the manuscript and created the figure. Both authors meet the following criteria: (1) substantial contributions to the conception or design of the work or the acquisition, analysis or interpretation of the data; (2) drafting the work or revising it critically for important intellectual content; (3) final approval of the completed version; and (4) accountability for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

References

1. Popper K. The logic of scientific discovery. New York: Basic Books; 1959.
2. Ioannidis J. Why most published research findings are false. PLoS Med 2005 Aug;2(8):e124 [FREE Full text] [doi: 10.1371/journal.pmed.0020124] [Medline: 16060722]
3. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet 2009 Jul 04;374(9683):86-89. [doi: 10.1016/S0140-6736(09)60329-9] [Medline: 19525005]

4. Begley C, Ellis L. Drug development: Raise standards for preclinical cancer research. Nature 2012 Mar 28;483(7391):531-533. [doi: 10.1038/483531a] [Medline: 22460880]
5. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov 2011 Aug 31;10(9):712. [doi: 10.1038/nrd3439-c1] [Medline: 21892149]
6. Lopez-Rodriguez R, Perez-Pampin E, Marquez A, Blanco F, Joven B, Carreira P, et al. Validation study of genetic biomarkers of response to TNF inhibitors in rheumatoid arthritis. PLoS One 2018;13(5):e0196793 [FREE Full text] [doi: 10.1371/journal.pone.0196793] [Medline: 29734345]
7. Sweeney TE, Azad TD, Donato M, Haynes WA, Perumal TM, Henao R, et al. Unsupervised Analysis of Transcriptomics in Bacterial Sepsis Across Multiple Datasets Reveals Three Robust Clusters. Crit Care Med 2018 Jun;46(6):915-925 [FREE Full text] [doi: 10.1097/CCM.0000000000003084] [Medline: 29537985]
8. Sweeney T, Braviak L, Tato C, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. Lancet Respir Med 2016 Mar;4(3):213-224 [FREE Full text] [doi: 10.1016/S2213-2600(16)00048-5] [Medline: 26907218]
9. Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. PLoS Med 2019 Apr;16(4):e1002786 [FREE Full text] [doi: 10.1371/journal.pmed.1002786] [Medline: 31013272]
10. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood) 2008;27(3):759-769. [doi: 10.1377/hlthaff.27.3.759] [Medline: 18474969]
11. IHI Triple Aim Initiative. Institute for Healthcare Improvement. URL: http://www.ihi.org/Engage/Initiatives/TripleAim/Pages/default.aspx [accessed 2020-07-15]
12. Cahan E, Hernandez-Boussard T, Thadaney-Israni S, Rubin D. Putting the data before the algorithm in big data addressing personalized healthcare. NPJ Digit Med 2019;2:78 [FREE Full text] [doi: 10.1038/s41746-019-0157-2] [Medline: 31453373]
13. Sweeney T, Khatri P. Generalizable Biomarkers in Critical Care: Toward Precision Medicine. Crit Care Med 2017 Jun;45(6):934-939 [FREE Full text] [doi: 10.1097/CCM.0000000000002402] [Medline: 28509729]
14. Vineis P, Schulte P, McMichael A. Misconceptions about the use of genetic tests in populations. Lancet 2001 Mar 03;357(9257):709-712. [doi: 10.1016/S0140-6736(00)04136-2] [Medline: 11247571]
15. Sweeney T, Haynes W, Vallania F, Ioannidis J, Khatri P. Methods to increase reproducibility in differential gene expression via meta-analysis. Nucleic Acids Res 2017 Jan 09;45(1):e1 [FREE Full text] [doi: 10.1093/nar/gkw797] [Medline: 27634930]
16. Reed L, Williams S, Springston M, Brown J, Freeman K, DesRoches C, et al. Genotype-by-diet interactions drive metabolic phenotype variation in Drosophila melanogaster. Genetics 2010 Jul;185(3):1009-1019 [FREE Full text] [doi: 10.1534/genetics.109.113571] [Medline: 20385784]
17. Patel C, Burford B, Ioannidis J. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. J Clin Epidemiol 2015 Sep;68(9):1046-1058 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.05.029] [Medline: 26279400]
18. Kaptchuk T. Effect of interpretive bias on research evidence. BMJ 2003 Jun 28;326(7404):1453-1455 [FREE Full text] [doi: 10.1136/bmj.326.7404.1453] [Medline: 12829562]
19. Ioannidis J. How to make more published research true. PLoS Med 2014 Oct;11(10):e1001747 [FREE Full text] [doi: 10.1371/journal.pmed.1001747] [Medline: 25334033]
20. Kwan A, Hubank M, Rashid A, Klein N, Peters M. Transcriptional instability during evolving sepsis may limit biomarker based risk stratification. PLoS One 2013;8(3):e60501 [FREE Full text] [doi: 10.1371/journal.pone.0060501] [Medline: 23544148]
21. Khatri Lab: Tools 2019. Khatri P. URL: https://khatrilab.stanford.edu/tools/ [accessed 2020-07-14]
22. Sieberts SK, Zhu F, García-García J, Stahl E, Pratap A, Pandey G, Members of the Rheumatoid Arthritis Challenge Consortium, et al. Crowdsourced assessment of common genetic contribution to predicting anti-TNF treatment response in rheumatoid arthritis. Nat Commun 2016 Aug 23;7:12460 [FREE Full text] [doi: 10.1038/ncomms12460] [Medline: 27549343]
23. Allen GI, Amoroso N, Anghel C, Balagurusamy V, Bare CJ, Beaton D, Alzheimer's Disease Neuroimaging Initiative. Crowdsourced estimation of cognitive decline and resilience in Alzheimer's disease. Alzheimers Dement 2016 Jun;12(6):645-653 [FREE Full text] [doi: 10.1016/j.jalz.2016.02.006] [Medline: 27079753]
24. Guinney J, Wang T, Laajala TD, Winner KK, Bare JC, Neto EC, Prostate Cancer Challenge DREAM Community. Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: development of a prognostic model through a crowdsourced challenge with open clinical trial data. Lancet Oncol 2017 Jan;18(1):132-142 [FREE Full text] [doi: 10.1016/S1470-2045(16)30560-5] [Medline: 27864015]
25. Sweeney TE, Perumal TM, Henao R, Nichols M, Howrylak JA, Choi AM, et al. A community approach to mortality prediction in sepsis via gene expression analysis. Nat Commun 2018 Feb 15;9(1):694 [FREE Full text] [doi: 10.1038/s41467-018-03078-2] [Medline: 29449546]
26. Vallania F, Tam A, Lofgren S, Schaffert S, Azad TD, Bongen E, et al. Leveraging heterogeneity across multiple datasets increases cell-mixture deconvolution accuracy and reduces biological and technical biases. Nat Commun 2018 Nov 09;9(1):4735 [FREE Full text] [doi: 10.1038/s41467-018-07242-6] [Medline: 30413720]

27. Barth AS, Kuner R, Buness A, Ruschhaupt M, Merk S, Zwermann L, et al. Identification of a common gene expression signature in dilated cardiomyopathy across independent microarray studies. J Am Coll Cardiol 2006 Oct 17;48(8):1610-1617 [FREE Full text] [doi: 10.1016/j.jacc.2006.07.026] [Medline: 17045896]
28. Khatri P, Roedder S, Kimura N, De Vusser K, Morgan AA, Gong Y, et al. A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation. J Exp Med 2013 Oct 21;210(11):2205-2221 [FREE Full text] [doi: 10.1084/jem.20122709] [Medline: 24127489]
29. Sweeney TE, Wong HR, Khatri P. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. Sci Transl Med 2016 Jul 06;8(346):346ra91 [FREE Full text] [doi: 10.1126/scitranslmed.aaf7165] [Medline: 27384347]
30. Andres-Terre M, McGuire HM, Pouliot Y, Bongen E, Sweeney TE, Tato CM, et al. Integrated, Multi-cohort Analysis Identifies Conserved Transcriptional Signatures across Multiple Respiratory Viruses. Immunity 2015 Dec 15;43(6):1199-1211 [FREE Full text] [doi: 10.1016/j.immuni.2015.11.003] [Medline: 26682989]
31. Li MD, Burns TC, Morgan AA, Khatri P. Integrated multi-cohort transcriptional meta-analysis of neurodegenerative diseases. Acta Neuropathol Commun 2014 Sep 04;2:93 [FREE Full text] [doi: 10.1186/s40478-014-0093-y] [Medline: 25187168]
32. Lofgren S, Hinchcliff M, Carns M, Wood T, Aren K, Arroyo E, et al. Integrated, multicohort analysis of systemic sclerosis identifies robust transcriptional signature of disease severity. JCI Insight 2016 Dec 22;1(21):e89073. [doi: 10.1172/jci.insight.89073] [Medline: 28018971]
33. Scott M, Vallania F, Khatri P. META-ANALYSIS OF CONTINUOUS PHENOTYPES IDENTIFIES A GENE SIGNATURE THAT CORRELATES WITH COPD DISEASE STATUS. Pac Symp Biocomput 2017;22:266-275. [doi: 10.1142/9789813207813_0026] [Medline: 27896981]
34. Sweeney TE, Shidham A, Wong HR, Khatri P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. Sci Transl Med 2015 May 13;7(287):287ra71 [FREE Full text] [doi: 10.1126/scitranslmed.aaa5993] [Medline: 25972003]
35. Francisco N, Fang Y, Ding L, Feng S, Yang Y, Wu M, et al. Diagnostic accuracy of a selected signature gene set that discriminates active pulmonary tuberculosis and other pulmonary diseases. J Infect 2017 Dec;75(6):499-510. [doi: 10.1016/j.jinf.2017.09.012] [Medline: 28941629]
36. Maslove D, Shapira T, Tyryshkin K, Veldhoen R, Marshall J, Muscedere J. Validation of diagnostic gene sets to identify critically ill patients with sepsis. J Crit Care 2019 Feb;49:92-98. [doi: 10.1016/j.jcrc.2018.10.028] [Medline: 30408726]
37. Temple R. Enrichment of clinical study populations. Clin Pharmacol Ther 2010 Dec;88(6):774-778. [doi: 10.1038/clpt.2010.233] [Medline: 20944560]
38. Raymond S, López MC, Baker H, Larson S, Efron P, Sweeney T, et al. Unique transcriptomic response to sepsis is observed among patients of different age groups. PLoS One 2017;12(9):e0184159 [FREE Full text] [doi: 10.1371/journal.pone.0184159] [Medline: 28886074]
39. Sweeney T, Thomas N, Howrylak J, Wong H, Rogers A, Khatri P. Multicohort Analysis of Whole-Blood Gene Expression Data Does Not Form a Robust Diagnostic for Acute Respiratory Distress Syndrome. Crit Care Med 2018 Feb;46(2):244-251 [FREE Full text] [doi: 10.1097/CCM.0000000000002839] [Medline: 29337789]
40. Gawande A. The hot spotters: can we lower medical costs by giving the neediest patients better care? New Yorker. The New Yorker. 2011 Jan 24. URL: https://www.newyorker.com/magazine/2011/01/24/the-hot-spotters [accessed 2020-07-14]

Abbreviations

GWAS: genome-wide association study

Edited by G Eysenbach; submitted 29.01.20; peer-reviewed by S Israni, N Benda; comments to author 19.03.20; revised version received 18.05.20; accepted 03.06.20; published 12.08.20.

Please cite as:
Cahan EM, Khatri P
Data Heterogeneity: The Enzyme to Catalyze Translational Bioinformatics?
J Med Internet Res 2020;22(8):e18044
URL: https://www.jmir.org/2020/8/e18044
doi:10.2196/18044
PMID:32784182

Viewpoint

Can Robots Improve Testing Capacity for SARS-CoV-2?

Kathrin Cresswell[1], BSc, MSc, PhD; Sandeep Ramalingam[1], MD, PhD, FRCPath; Aziz Sheikh[1], OBE, FRSE, FMedSci
The University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Kathrin Cresswell, BSc, MSc, PhD
The University of Edinburgh
Teviot Place
Edinburgh
United Kingdom
Phone: 44 131 651 4151
Email: kathrin.beyer@ed.ac.uk

Abstract

There is currently increasing interest internationally in deploying robotic applications for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) testing, as these can help to reduce the risk of transmission of the virus to health care staff and patients. We provide an overview of key recent developments in this area. We argue that, although there is some potential for deploying robots to help with SARS-CoV-2 testing, the potential of patient-facing applications is likely to be limited. This is due to the high costs associated with patient-facing functionality, and risks of potentially adverse impacts on health care staff work practices and patient interactions. In contrast, back-end laboratory-based robots dealing with sample extraction and amplification, that effectively integrate with established processes, software, and interfaces to process samples, are much more likely to result in safety and efficiency gains. Consideration should therefore be given to deploying these at scale.

(J Med Internet Res 2020;22(8):e20169) doi:10.2196/20169

KEYWORDS

robotics; testing; SARS-CoV-2; COVID-19; pandemic; virus; infectious disease

Introduction

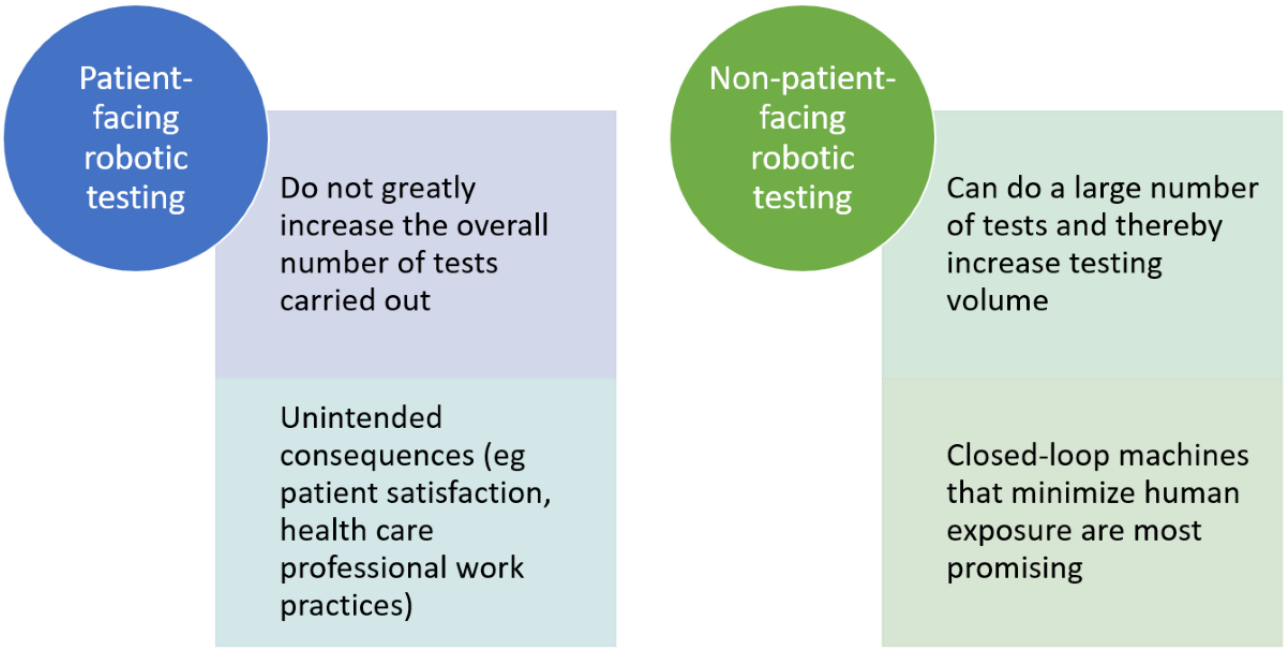
Testing is crucial to identify, curb the spread of, and contain severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Testing capacity will therefore need to increase substantially for the foreseeable future [1]. Robotic testing technologies may help to increase testing capacity and also minimize the risk of nosocomial transmission. There are currently two ways of testing for coronavirus disease (COVID-19): virological tests and serological tests. Virological methods work with genetic material obtained from nasal, throat, or saliva swabs and commonly use reverse transcription polymerase chain reaction (RT-PCR) technology to covert ribonucleic acid (RNA) to deoxyribonucleic acid (DNA). This technology is used to detect the presence of SARS-CoV-2. Serological tests use saliva, whole blood, serum, or plasma to look for antibodies. Various new testing methods that are variants of these two approaches are currently in development [2-6].

However, despite a general recognition that testing for SARS-CoV-2 is a key international priority, there is currently a lack of testing capacity contributing to the inadequate numbers of tests being undertaken [7]. In addition, existing testing procedures can endanger health care staff and laboratory technicians (ie, those who have to handle blood samples and swabs). We here provide an overview of key recent developments in robotic testing for SARS-CoV-2, which can help to reduce the exposure of health care and technical staff to the virus.

Overview of Current Developments in Deploying Robots for SARS-CoV-2 Testing

Robots for SARS-CoV-2 testing procedures can be either patient-facing (eg, collecting nasal swabs and thereby reducing exposure of those collecting swabs) or non–patient-facing (eg, liquid handling machines that reduce exposure for laboratory technicians; Figure 1).

Figure 1. Types of robots used for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) testing procedures.



To date, patient-facing testing robots have been mainly experimental, deployed as pilots and in limited settings. Examples include remote-controlled robots taking throat swabs that have been used in parts of China [8], a robotic arm handing out test tubes to drivers in cars for SARS-CoV-2 testing [9], and a 3D-printed robotic arm taking throat swabs that was developed in Denmark [10]. Such robots are expensive (the Chinese robot costs approximately £62,000 [US $81,069]), can only do a limited number of activities on 1 sample at a time (and therefore do not greatly increase the overall number of tests carried out). They can also cause patients to worry due to a lack of personal contact [8]. Previous research has further found that patient-facing robotic applications can have unintended consequences resulting from adverse impacts on health care professional work practices, and on patient satisfaction [11,12]. For example, frail older adults and isolated patients depend on face-to-face contact as a source of emotional support. Attempts at making these applications more human-like may only partly address this issue, as robots that look too human-like can be perceived as threatening [13,14]. Patient-facing robots may, however, play a role in high-risk settings where infection control needs to be prioritized. They cannot replace face-to-face interactions that are required to provide high-quality and safe care for the majority of patients.

Non–patient-facing testing robots including liquid handlers, especially those that do not involve contact dispensing, are more promising. These robots can move liquids using magnetic plates, aspirate, dispense, or transport liquid samples (sometimes using pipettes), and in some cases interpret biological or chemical events (eg, detect if a virus is present). This reduces exposure for laboratory technicians who have to have to handle blood samples and swabs, and interface with these machines for sample preparation. Many laboratories already utilize some degree of automation, and this mitigates the risks of adverse impacts on existing work practices of health care staff.

Automated testing robots also have a high throughput and are rapidly able to tackle the large volumes of tests required during the SARS-CoV-2 pandemic and in the “new normal,” as they can carry out numerous tests simultaneously [15]. For example, the Spanish Ministry of Health recently commissioned 4 COVID-19 testing robots that will be able to carry out 80,000 tests per day [16]. Similarly, a newly established COVID-19 testing laboratory at Berkeley’s Innovative Genomics Institute (IGI) uses a robotic liquid handling machine that uses pipetting to test up to 3000 cases per day [17]. Another example is a Danish pipetting robot that automates the sample preparation process and was originally used for testing for *Salmonella* before being repurposed to test for SARS-CoV-2 [18].

However, the functionality of these machines varies. Some RT-PCR liquid handlers only help up to the extraction and addition of samples for PCR/RT-PCR, while others also transfer the material to a thermal cycler where PCR/RT-PCR happens (ie, extraction and amplification). Clearly, a closed-loop process (where technicians input a sample and the machine prepares samples and tests) is preferable as it minimizes human contact with samples (including testing for multiple viruses).

Another potential issue is the interfacing with existing software and the associated communication of results. Some automated testing robots do not allow automatic downloading of results

from the robot to the main laboratory computer, which then renders the whole process impractical as the large number of results generated has to be manually entered by technicians. This may also introduce the risk of transcription errors, which may in turn have adverse consequences for patient care [19]. Additional integration software can help to address this issue, but adds to the overall cost (in relation to both acquisition and maintenance) and may require additional programming.

Conclusions

Overall, there is a lack of empirical evidence on patient-facing virological/serological testing robots, and even if such evidence was available, these technologies would be unlikely to tackle pressing issues around scaling of testing capability. Testing robots in laboratories, however, have the potential to bring significant rapid benefits at low cost as these technologies can fulfil multiple purposes (eg, handling other types of liquids). Moreover, they already exist in many laboratories and can therefore be readily repurposed to respond to COVID-19 (although this has to be done by the manufacturer). Most useful are likely to be non–patient-facing testing robots that tackle the whole extraction and amplification cycle, as this will eliminate the need to transfer material for the amplification stage and thereby minimize the risk of unintended consequences.

Where there are established processes (eg, back-end laboratory-based robots tackling extraction and amplification) and where these interface effectively with existing software to process the results, these should be scaled up. In parallel, there is a need to stimulate research and innovation initiatives to explore the feasibility of developing a scalable front-end testing robot for high-risk settings (eg, infectious disease wards).

Acknowledgments

This work was funded by a Scottish Government Chief Scientist Research Grant. AS is supported by BREATHE - The Health Data Research Hub for Respiratory Health (MC_PC_19004), which is funded through the UK Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research UK. The views expressed are those of the authors.

Conflicts of Interest

None declared.

References

1. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA 2020 Mar 03;323(14):1341. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
2. Lamb LE, Bartolone SN, Ward E, Chancellor MB. Rapid detection of novel coronavirus/Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) by reverse transcription-loop-mediated isothermal amplification. PLoS One 2020;15(6):e0234682 [FREE Full text] [doi: 10.1371/journal.pone.0234682] [Medline: 32530929]
3. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology 2020 Aug;296(2):E32-E40 [FREE Full text] [doi: 10.1148/radiol.2020200642] [Medline: 32101510]
4. Bai H, Cai X, Zhang X. Landscape Coronavirus Disease 2019 test (COVID-19 test) in vitro--A comparison of PCR vs Immunoassay vs Crispr-Based test. OSF Preprints 2020:1-13. [doi: 10.31219/osf.io/6eagn]
5. Fast, portable tests come online to curb coronavirus pandemic. URL: https://www.briancolemd.com/wp-content/themes/ypo-theme/pdf/fast-portable-tests-come-online-to-curb-coronavirus-pandemic.pdf [accessed 2020-05-10]
6. Is the future of blood testing robotic?. URL: https://www.medicaldevice-network.com/features/robotic-blood-sampling/ [accessed 2020-07-09]
7. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24;323(13):1239. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
8. Engineers develop robots to treat and test Covid-19 patients in a bid to protect health workers. URL: https://www.telegraph.co.uk/global-health/science-and-disease/engineers-develop-robots-treat-test-covid-19-patients-bid-protect/ [accessed 2020-05-10]
9. Robots help to make drive-through virus testing safer. URL: https://drivesncontrols.com/news/fullstory.php/aid/6334/Robots_help_to_make_drive-through_virus_testing_safer.html [accessed 2020-05-10]
10. Prototype robot that carries out throat swab tests for coronavirus using a disposable 3D-printed arm is developed by Danish researchers. URL: https://www.dailymail.co.uk/sciencetech/article-8365079/COVID-19-throat-swab-test-robot-developed-Danish-researchers.html [accessed 2020-07-09]
11. Cresswell K, Cunningham-Burley S, Sheikh A. Health Care Robotics: Qualitative Exploration of Key Challenges and Future Directions. J Med Internet Res 2018 Jul 04;20(7):e10410 [FREE Full text] [doi: 10.2196/10410] [Medline: 29973336]
12. Broadbent E, Stafford R, MacDonald B. Acceptance of Healthcare Robots for the Older Population: Review and Future Directions. Int J of Soc Robotics 2009 Oct 3;1(4):319-330. [doi: 10.1007/s12369-009-0030-6]

13. Seyama J, Nagayama RS. The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. Presence: Teleoperators and Virtual Environments 2007 Aug;16(4):337-351. [doi: 10.1162/pres.16.4.337]
14. S Korean researchers develop robots to remotely collect samples for COVID-19 tests. URL: https://www.youtube.com/watch?v=ss_JO0j2O44 [accessed 2020-07-09]
15. Kong F, Yuan L, Zheng YF, Chen W. Automatic liquid handling for life science: a critical review of the current state of the art. J Lab Autom 2012 Jun 06;17(3):169-185. [doi: 10.1177/2211068211435302] [Medline: 22357568]
16. Spain plans to use robots to test 80,000 people a day for the coronavirus. URL: https://thenextweb.com/neural/2020/03/23/spain-plans-to-use-robots-to-test-80000-people-a-day-for-the-coronavirus/ [accessed 2020-05-10]
17. UC Berkeley scientists spin up a robotic COVID-19 testing lab. URL: https://news.berkeley.edu/2020/03/30/uc-berkeley-scientists-spin-up-a-robotic-covid-19-testing-lab/ [accessed 2020-05-10]
18. Robot automates COVID-19 testing. URL: https://healthcare-in-europe.com/en/news/robot-automates-covid-19-testing.html [accessed 2020-07-09]
19. Kushniruk AW, Triola MM, Borycki EM, Stein B, Kannry JL. Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. Int J Med Inform 2005 Aug;74(7-8):519-526. [doi: 10.1016/j.ijmedinf.2005.01.003] [Medline: 16043081]

Abbreviations

COVID-19: coronavirus disease
RT-PCR: reverse transcription polymerase chain reaction
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 13.05.20; peer-reviewed by H Rjoub, JA Sánchez Margallo; comments to author 08.07.20; revised version received 12.07.20; accepted 22.07.20; published 12.08.20.

Please cite as:
Cresswell K, Ramalingam S, Sheikh A
Can Robots Improve Testing Capacity for SARS-CoV-2?
J Med Internet Res 2020;22(8):e20169
URL: http://www.jmir.org/2020/8/e20169/
doi:10.2196/20169
PMID:32735547

Viewpoint

Sijilli: A Scalable Model of Cloud-Based Electronic Health Records for Migrating Populations in Low-Resource Settings

Shadi Saleh[1,2], MPH, PhD; Nour El Arnaout[1], MPH; Lina Abdouni[1], MSc; Zeinab Jammoul[1], RN, MPH; Noha Hachach[1,3], MPH; Amlan Dasgupta[4], MSc

[1]Global Health Institute, American University of Beirut, Beirut, Lebanon

[2]Department of Health Management and Policy, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

[3]American University of Beirut Medical Center, Beirut, Lebanon

[4]Epic Systems Corporation, Verona, WI, United States

Corresponding Author:
Shadi Saleh, MPH, PhD
Global Health Institute
American University of Beirut
Riad El Solh
Beirut, 1107 2020
Lebanon
Phone: 961 1350000 ext 4832
Email: ss117@aub.edu.lb

Abstract

The world is witnessing an alarming rate of displacement and migration, with more than 70.8 million forcibly displaced individuals, including 26 million refugees. These populations are known to have increased vulnerability and susceptibility to mental and physical health problems due to the migration journey. Access of these individuals to health services, whether during their trajectory of displacement or in refugee-hosting countries, remains limited and challenging due to multiple factors, including language and cultural barriers and unavailability of the refugees' health records. Cloud-based electronic health records (EHRs) are considered among the top five health technologies integrated in humanitarian crisis preparedness and response during times of conflict. This viewpoint describes the design and implementation of a scalable and innovative cloud-based EHR named Sijilli, which targets refugees in low-resource settings. This paper discusses this solution compared with other similar practices, shedding light on its potential for scalability.

(J Med Internet Res 2020;22(8):e18183) doi:10.2196/18183

KEYWORDS

eHealth; digital health; innovation; refugees; low- and middle-income countries; technology

Introduction

Global Refugee Crisis

The number of forcibly displaced individuals, including refugees, has reached an alarming rate globally, with 70.8 million forcibly displaced individuals, including around 26 million refugees [1], of whom the majority are of Syrian nationality [2,3]. More than 80% of these refugees have been internally displaced or have fled to Syria's neighboring countries [4]. Notably, low- and middle-income countries (LMICs) host the highest number of refugees [4], with inconsistent percentages reported [5]. Figures from 2016 indicate that a total of 5 million registered Syrian refugees have fled to Syria's neighboring countries, namely Lebanon, Jordan, Egypt, Turkey, and Iraq [6].

Health Needs of Refugees

Refugees are often exposed to stressful conditions of vulnerability, poverty, poor nutrition, and marginalization prior to and during the journey of migration, a situation that increases their susceptibility of developing different health problems, including mental health problems [7-9]. From a health perspective, women, children, and older adults remain the most affected during the journey of migration [9]. Malnutrition, anemia, communicable and noncommunicable diseases (NCDs), women's health, and mental health emerge as the top priority health concerns of refugees [10]. However, several barriers impede refugees' access to the appropriate health services targeting these needs in the host countries. Examples of barriers include inadequate knowledge about the availability of health services, insufficient financial capacities, limited access to

transportation, cultural differences, and language barriers and the scarcity of interpreters [11,12]. Legal restrictions such as issues of registration could also hinder refugees from accessing adequate health services in the host countries [12,13].

Need for Innovative Solutions for Health System Strengthening in LMICs

The growing and continuous influx of refugees imposes challenges on the health systems of the refugee-hosting countries vis-à-vis catering to the needs of these populations and providing essential health services [9], especially because the majority of refugee-hosting countries are LMICs with already exhausted and fragile health care systems. Having said that, strengthening the fragile health systems of the refugee-hosting countries remains an urgent need to ensure equitable access to health services for refugee populations [9]. This in turn calls for the implementation of relatively low-cost, innovative, feasible, and contextualized solutions tailored to target the needs of these emotionally and physically distressed populations.

Electronic health records (EHRs) have been progressively adopted in resource-constrained settings [14] and conflict-affected areas [15] for the great promise they hold in providing coordinated care, sharing patient information among health care providers across different settings, reducing medical errors, and improving the quality and continuity of health care [16]. For instance, the use of EHRs avoids duplicate tests, prevents drug-drug interactions, and consequently improves patient care. In addition, EHRs allow patients' access to their health records remotely and permit better management of their health [17].

EHRs have evolved in the past decade [16] and have been defined as a longitudinal systematic collection of electronic health information of individuals or populations [18,19] created in any health care delivery setting [20]. EHRs generally store patients' health data, including demographic information, social history, health problems, active diagnoses and past medical history, diagnostic test results, medications, hospitalization information, consultant reports, immunizations, allergies, health screening study results, and progress notes [17].

A relatively recent type of EHRs is the cloud-based EHR system [21]. Cloud-based EHRs have been promoted in conflict-affected areas where the digital infrastructure, such as internet connectivity, is lacking, and they have been regarded as cost-effective methods that permit scalability and interoperability [22]. In fact, cloud-based EHRs not only track patients across time, but also permit the sharing of patients' data among health care providers so that patients' health care can be easily tracked and monitored [15]. In addition, cloud-based EHRs are considered among the top five health technologies integrated in humanitarian crisis preparedness and response during times of conflict [23]. The World Disasters Report 2013 stated that the use of cloud-based EHR is a successful approach for the recovery of health records in case of damage to physically held health records, a highly probable incidence in conflict settings [24]. Moreover, cloud-based EHRs have been regarded as resilient and effective tools that facilitate monitoring the health status of refugees over the long term, specifically in protracted crisis, and eventually improve the process of health care delivery [25-27].

Existing eHealth Systems Targeting Refugee Populations

A few types of EHRs have been implemented for migrants and refugees [28] and have shown their potential to address some of the challenges facing these populations in accessing health care services [28,29], yet they also underline areas that need improvement. The United Nations Relief and Works Agency for Palestinian Refugees (UNRWA) developed a computer-based (eHealth) system to improve the quality and efficiency of health care services provided to Palestinian refugees living in UNRWA's fields of operations, namely Lebanon, Jordan, Syria, West Bank, and Gaza [30]. The UNRWA's eHealth system was developed and introduced in 2010 in response to the increasing workload and to the high prevalence of NCDs [30]. The system was associated with an improvement in the quality of health care and was described as a secure and cost-effective tool supporting the health care services provided by the UNRWA's health program [30]. However, the UNRWA's eHealth system covers only Palestinian refugees that are registered with the UNRWA and excludes millions of unregistered Palestinian refugees and all refugees of other nationalities worldwide. Adding to the drawbacks of the UNRWA's eHealth system is its restricted use, whereby it is accessible solely at health facilities governed by UNRWA.

Similarly, the Refugee Assistance Information System (RAIS) is an EHR system developed by the United Nations High Commissioner for Refugees (UNHCR) to track, monitor, and provide assistance to refugees registered with UNHCR [25,31]. It collects basic demographic information and some health services data of UNHCR-registered refugees [25,31]. However, the RAIS remains incomprehensive regarding health data and is exclusively created for refugees registered with UNHCR, leaving behind millions of unregistered refugees.

The Context of Lebanon

Lebanon, one of the main LMICs housing refugees, hosts around 1 million registered Syrian refugees in addition to more than 400,000 unregistered refugees [32], rendering it the country with the highest number of refugees per capita worldwide [5]. This massive influx of refugees to Lebanon, coupled with their increased demand for health services, has consequently created an unprecedented strain on the local health care system, originally characterized by its fragility [33]. The situation is further aggravated by refugees residing mainly in underserved rural Lebanese areas, where health service delivery is relatively suboptimal and insufficient [34].

Regarding the use of eHealth systems for refugees in Lebanon, both UNRWA's eHealth system and the RAIS are applicable in the country, yet the use of both systems remains bounded by the aforementioned restrictions.

Another example of existing technologies in Lebanon is Phoenix. Phoenix is an EHR system developed by the Lebanese Ministry of Public Health (MoPH), and it functions exclusively among the 229 primary health care centers (PHCs) that operate under the supervision of the MoPH. Phoenix EHR targets

disadvantaged Lebanese and refugees from different nationalities who receive primary health care within the national network of PHCs governed by the MoPH. It contains vital patient information, including documentation, guidelines, orders, and results [35]. The system allows the exchange of patient health information among health care providers within each PHC and the MoPH; however, the health information included in Phoenix EHR cannot be shared across the different PHCs belonging to the network. Instead, a new health record has to be created again whenever a patient visits a new PHC for the first time. In addition, Phoenix EHR does not function at Lebanese hospitals or mobile health clinics, where a significant proportion of refugees seek health services.

Sijilli: A Cloud-Based EHR for Refugees in Low-Resource Settings

Conception of Sijilli

The pressing need for innovative digital solutions that would enhance refugees' access to needed health services in low-resource settings, regardless of their migration journey, triggered the development of Sijilli (meaning "my record" in Arabic).

This paper describes the design of the cloud-based mobile EHR system and reports on its implementation among Syrian refugees in Lebanon. Sijilli [36] was launched in 2018 as a collaboration between the Global Health Institute (GHI) at the American University of Beirut (AUB) in Beirut, Lebanon, and the health care software company Epic in the United States. Sijilli was designed in a contextualized approach to cater to refugees' need for universally accessible EHRs and to fill the several gaps identified in existing practices and available similar technologies. It aims to securely collect and preserve essential health data for refugees during and beyond their displacement journey, besides ensuring data security, universal access, and interoperability of the health record data. Of note, more than 10,000 Syrian refugees across Lebanon currently possess a Sijilli EHR.

The Concept Design and Implementation

Sijilli is designed to allow interaction and to support convenient workflows among different users, from the data entry personnel to the administrators and ultimately to the end users, namely health providers and refugees.

Data entry personnel collect the health information from refugees in different remote and underserved areas using data entry–friendly software run on tablet computers. Worth mentioning is that the use of the Sijilli software system by data entry personnel does not require internet connectivity and is done completely offline. The system adopts a simple user interface design to accommodate users and data collectors of diverse health backgrounds. Once the health information of the refugee is entered into the system, the Sijilli data collection software generates a password-protected and advanced encrypted standard (AES) PDF document of the health record of the refugee. The password of the PDF is generated exclusively and blindly (ie, without input from the data entry personnel) for the Sijilli holder and consists of the first letters of the refugee's first name, father's name, mother's name, and family name, in addition to the 4 numbers of their year of birth, entered in this specific order. This simple composition of the password allows Sijilli holders to recall their passwords by following simple instructions. The Sijilli health record in its PDF version is then downloaded on a key-shaped flash drive (Figure 1) and handed to the refugee for use across different health facilities worldwide without any restrictions. In parallel, an encrypted deidentified version of the generated health record is uploaded to the Sijilli cloud-based server. The cloud-based version of the health record of any Sijilli holder can be accessed globally by either the patient or the health provider via the Sijilli website [37] following a 2-step identity verification process. The multiple layers of security (ie, USBs, hardware used for data entry, and internal and external servers of the cloud-based system) are crucial to ensure data security and privacy of the Sijilli system.

Worth noting is that patients cannot alter any data contained in their Sijilli records, neither through the PDF on the flash drive nor through the patient's portal on the Sijilli website. On the other hand, health providers can have access to refugee patients' Sijilli records when provided with the personal identification number (PIN) and the security questions that are recognized exclusively by Sijilli holders/refugees. External health providers can see all the health data of the refugee, originally collected by the data entry personnel at the time of the creation of the health record, and can also add clinical notes online at the time of encounter with the patient using the cloud-based version of the Sijilli EHR. No modifications to the original patient information can be made by the health providers. This ensures traceability and protects the liability of the GHI and Epic team.

Figure 1. The key-shaped USB containing the Sijilli electronic health record handed to the refugee.

Architecture of Sijilli

Clinical Development

The Sijilli EHR is composed of 7 sections covering sociodemographic information, social and lifestyle habits, medical and surgical history, obstetrics and gynecological (OB/GYN) conditions for women only, medication use, vaccination history, and mental health screening. The data entered in the sociodemographic section allow the creation of the password and security questions used at times of access to the health record.

The sociodemographic section includes questions on basic sociodemographic information, such as the previous occupation in the country of origin where conflict is taking place and the current occupation in the refugee-hosting country, which allows the identification of potential occupational health hazards and risk factors. Other information collected includes the location of the settlement and the year of migration, among other information. The section also includes questions that may affect the exposure of the refugee to health risks, including parental consanguinity, given that consanguineous marriage is a deeply rooted social practice among the Syrian refugee population. Knowledge about parental consanguinity enables the health care provider to explore and identify genetic disorders such as thalassemia. Allergies are also entered in this section to indicate any source of allergy that the refugee has.

Further risk factors can be identified through data entered into the social and lifestyle section, which namely addresses smoking, alcohol drinking, and physical exercise. As the Syrian refugee crisis shifts from an acute emergency to a protracted crisis, NCDs and preventative medicine become a vital aspect of the health care response to refugee populations.

The medical and surgical history section was designed to allow the selection of one or multiple medical conditions or surgical procedures, if any, using a drop-down menu that includes the most common conditions and surgeries, derived from the International Classification of Diseases, 10th Revision developed by the World Health Organization [38]. Diseases not found in the drop-down list can be entered manually by data entry personnel. Worth noting is that the list of medical conditions and surgical procedures was constantly updated throughout the course of implementation according to the frequency of encounter of medical conditions to ensure a standardized method of data entry.

With regards to the OB/GYN section, it is worth noting that this section was not originally embedded in the system, given that OB/GYN-related conditions were included in the general medical and surgical history section. However, a stand-alone OB/GYN section was added during the project implementation phase to take into consideration that the majority of the refugees having their Sijilli records created were women of child-bearing age, many of whom were pregnant. Moreover, a high pregnancy complication rate was noted by the women, which required the addition of a section dedicated to OB/GYN that included pregnancy-related complications to ensure proper and comprehensive documentation. The aforementioned section includes questions on gravida, para, aborta, method of delivery and associated complications, breastfeeding practice, and use of contraception.

To record any medications that the refugee is taking, a section on medication use is included in the Sijilli software system. Medications included cover mainly those taken for chronic conditions, in a drop-down menu format. Categorization of medications was based on *Davis's Drug Guide for Nurses*, which classifies medications according to their type (eg, statins, antihypertensive drugs, etc) [39]. The data entry personnel are also able to enter any other medication taken beyond those in the list. The dosage, frequency of use, and start date of use for each entered medication are mandatory fields. This section remains necessary to indicate to the health provider any potential drug-drug interactions at the time of medication prescription.

The vaccination history section was included due to the outbreak of communicable diseases among the refugee population, such as the repetitive measles outbreaks witnessed among the refugee population in Lebanon. This section includes all required vaccines by the Lebanese MoPH. It is also designed to allow

data entry personnel to indicate if school vaccines were given in Syria, given that this was generally a routine practice.

The record then navigates to the mental health screening section, which uses the Patient Health Questionnaire-9 (PHQ-9). It is well established that refugees are at an increased risk of posttraumatic stress disorder, major depressive disorder (MDD), generalized anxiety disorder, and other behavioral and emotional problems. We elected to screen for MDD in particular, as previous research has shown that it is the most common and debilitating condition among Syrian refugees in Lebanon [40]. The PHQ-2 followed by the PHQ-9 were chosen as the screening tools for depression due to their ease of use, allowing for data collectors and data entry personnel with limited clinical experience.

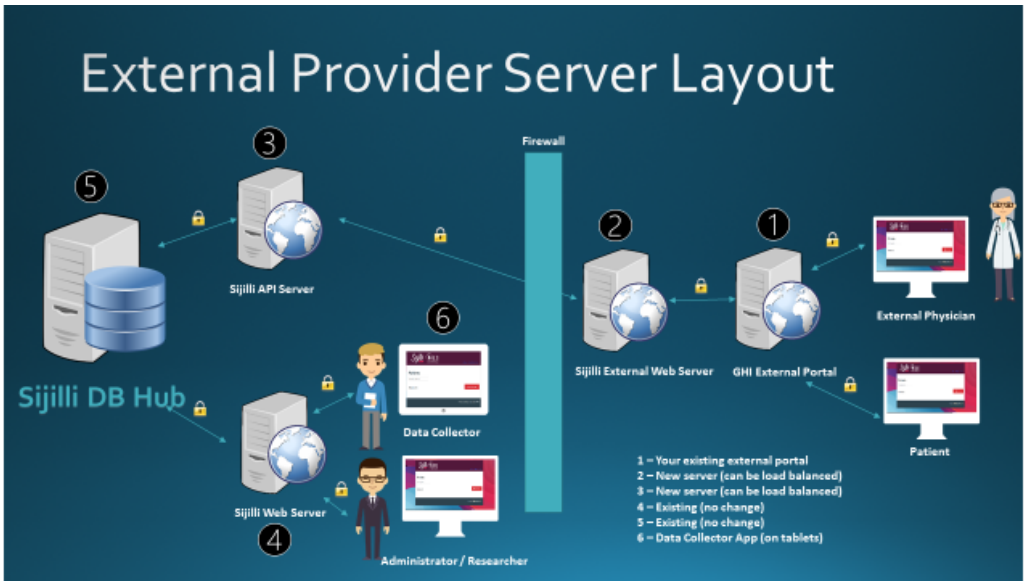
Technical Development

The Sijilli EHR system was designed to give different levels of access authorization to different users, such as data collectors and data entry personnel, administrators, external health providers, and refugees. Data entry personnel are on the front lines collecting information from the refugee population in areas often lacking network connectivity. All data from the data entry hardware (eg, laptops, tablets, etc) hosting the Sijilli data collection software can be wiped out after being synced back to the hub. The administrators, on the other hand, can access the Sijilli administrator web application, which is not available outside the internal network. The application can be used to authenticate hardware used for data collection and data entry, process synced requests, and update code sets that can be used in the Sijilli data collection software. Administrators can also generate a deidentified report of the health records for research purposes. The administrator web-based application can only be used by users authenticated in the AUB network and authorized by administrators through the web application. External health providers can access a specific refugee's Sijilli EHR by providing a PIN and answering security questions, which are only recognized by the refugee, examine a summary of the refugee patient's Sijilli EHR, and add clinical notes for future care sought by the refugee. In parallel, refugees who are Sijilli holders can also access their respective record through the password-protected PDF version of the EHR or through the portal on the Sijilli website by providing their PIN and answering security questions about their record.

Key Features of Sijilli's Technical Architecture

Figure 2 shows the technical architecture diagram of the Sijilli EHR, which is based on several key features.

Figure 2. Architecture diagram of the Sijilli electronic health record. API: application programming interface; DB: database; GHI: Global Health Institute.



Multilayer Security

Given that Sijilli is a cloud-based EHR, data security is of particular concern and has been considered by addressing several key points. These include access to data, which is driven by role-based user access whereby once the data get into the database, only external health providers and the refugees can access patient records by providing the PIN and answering specific security questions. The access of administrators is restricted to deidentified personal health information of the refugees following data upload to the Sijilli database. In terms of authentication, hardware used by data entry personnel are independently authenticated by the administrator. In cases where laptops are used for data collection and data entry, the disk on the Sijilli data collection application is encrypted to protect against theft. Data are then wiped out after being uploaded to the Sijilli hub database. The PDF files generated in the Sijilli data collection application are password protected with a refugee-friendly PIN and AES encryption. For access through the Sijilli portal, external health providers and refugees can only access one Sijilli record at a time, and a captcha check has been added to the external web applications to prevent robots from accessing multiple records. All data in motion were set to be network protected using transport layer security 1.2. An audit trail exists in the hub database to indicate the access patterns to the health records by external health providers.

Scalability

Sijilli is characterized by 3 independent server applications that are cloud enabled and can be independently scaled up based on load and usage patterns. The Sijilli data collection application can be installed on as many laptops or tablets as required for data collection/data entry, and the syncing of data occurs when the laptops or tablets used are within the internal network.

Extensibility and Maintainability

Administrators can manage the code sets and other metadata associated with the Sijilli application from the administrator application. This metadata can be synced to all the Sijilli data collection applications when the data from the data collector application are downloaded to the Sijilli hub. In terms of maintainability, the server teams have the authority to install the updates to the applications independently of each other.

Integration and Contextualization

A key feature of the Sijilli EHR is its contextualized design. The design of each of the Sijilli data collection applications, the EHR itself, and the Sijilli website portals was tailored to the needs and interest of each of the multiple users/stakeholders involved.

First, the design of the Sijilli data collection application used for primary creation of the Sijilli EHR for refugees was based on the input of several health providers, including physicians, nurses, community health workers, medical residents, and medical students. This was done to ensure that the platform developed is user-friendly at the time of Sijilli EHR creation and time efficient in terms of data entry and collection. The Sijilli data collection application then underwent beta testing on a sample of 100 refugees, after which the application underwent some calibration and fine-tuning steps. These mainly included changes in the units of categories used, the addition of some medical conditions to the drop-down menu, and the addition of a complete section on OB/GYN conditions for female refugees. Health providers were also involved in the design of the Sijilli website, the portal through which they can update the cloud-based EHR of refugees visiting a health facility. This is to ensure that the website/platform is user-friendly and smooth to use. The platform includes the Arabic translation of instructions to accommodate non–English-speaking health providers, who are common in the Syrian-Lebanese context.

Discussion

Summary

Migrating populations and refugees continue to face obstacles in accessing health care services [41] and experience gaps in the continuity of health care [29], which differs across host countries and during the stages of the migration process, a fact that poses a negative impact on their health status [41]. One of the biggest challenges that faces refugees during and beyond their migration journey is access to medical records [41]. In view of this, EHRs seem to have the potential to address some of these challenges and provide opportunities to increase access to health care services and eventually improve the quality and continuity of health care [29]. However, implementation of EHRs in LMICs is still limited [42].

Previous studies have examined the association between EHR adoption and improvement in the quality and continuity of health care in refugee settings and have emphasized the importance of the adoption of these innovative technologies for the great promise they hold in such resource-constrained settings [26,30,43-46]. A series of cohort studies by Khader et al [26,43-46] evaluated the effectiveness of the implementation of an electronic medical record (EMR) cohort monitoring system to follow up on Palestinian refugees with hypertension [26,44] and diabetes [43,45,47] in Jordan. Despite minimal challenges encountered, mainly operational challenges [47], the authors reported promising results and concluded that an EMR cohort monitoring system is an efficient tool in a refugee context in terms of management and follow-up of NCDs [26,43-46].

Comparison With Similar eHealth Practices

The previously described UNRWA's eHealth system [30] and the RAIS [25,31,47] have been considered simple, secure, and cost-effective tools that have been correlated with an improvement in the quality and continuity of health care among refugee populations. Although they both collect some health information, Sijilli EHR appears to be more comprehensive compared with the UNRWA's eHealth system and the RAIS, wherein it covers a more inclusive range of health data, including context-specific components, such as parental consanguinity and mental health, which are not covered by either of the other systems. For instance, consanguineous marriage, a common practice among refugees [48], is well known to be a risk factor for a variety of genetic diseases [49], while mental health disorders have been reportedly high among refugee populations [50]. Therefore, such context-specific components permit better evaluation and monitoring of the health status of refugees during their migration journey and beyond.

In terms of inclusiveness, the UNRWA's eHealth system is designed to target Palestinian refugees that are registered with the UNRWA, and it is only functional through the health centers in the 5 fields of UNRWA's operations [30]. Likewise, the RAIS is exclusively created for refugees registered with the UNHCR [25,31,47], and only UNHCR officials have access to the health record data [31]. Accordingly, both systems exclude millions of unregistered refugees worldwide. Conversely, although initially created for Syrian refugees across Lebanon, Sijilli is designed to take into consideration the potential for upscaling to cover refugees of any nationality worldwide and to provide them and their health care providers with global access to their health records wherever the migration journey takes them.

Potential for Upscaling

With the impending challenge of refugee migration, initially to neighboring host LMICs, followed by opportunistic migration to European and possibly North American countries, the importance of an innovative, universally accessible, and digital technology becomes more emphasized [23]. According to asylum-seeker data from UNHCR on refugee migration and

resettlement [51], more than half a million Syrians moved to Germany seeking asylum, approximately 100,000 moved to Sweden, and 50,000 moved to Austria. Additionally, about 100,000 have moved to North America, indicating the desire of refugees to seek final settlement in developed countries. Even beyond the example of Syrian refugees, the world is witnessing many other migration crises, such as the hundreds of thousands of Rohingya fleeing from Myanmar to Bangladesh and a similar number of Venezuelans seeking asylum in the United States, among others [52]. eHealth and EHRs are well known in developed countries and have proven to be impactful in enhancing the quality of clinical care. The obstacle to scaling such solutions to LMICs has been the limiting high costs to purchase and maintain such solutions and therefore, the development of low-cost cloud-based tools offers the needed solutions to support access and provision of health care to vulnerable populations [53]. Literature indicates that health care providers and governments are keen to improve access to, engagement with, and delivery of health care through the use of eHealth [23]. A recent World Bank Report highlighted that many donor agencies are currently focusing on scalable eHealth interventions for LMICs that are low cost and mobile, mainly to resolve the barriers of access [53]. Such innovative solutions can also account for issues related to continuity of care and language barriers, while providing users with adequate health services.

Sijilli's potential for scalability and interoperability lies in the fact that Sijilli is a low-cost cloud-based EHR that can be easily exported to conflict-affected or crisis-affected areas that lack the required digital infrastructure, and, more importantly, it omits the need for physically held patient records. Furthermore, such solutions can be transported with the migrating user across several host settings with guaranteed security that puts the refugee and the host service provider at ease. Countries with fragmented health care systems would also benefit from a parallel solution, which would allow service providers to exchange and consolidate the patient's health-related information and interventions in one comprehensive, secure, and easy-to-access EHR.

Conclusions

In conclusion, the Sijilli EHR represents a model for an innovative and scalable cloud-based solution that could be replicated for several vulnerable populations in different low-resource or crisis settings. Sijilli EHR's contextualized design, architecture with multilayered security, and process design and implementation highlight the potential applicability of this digital health solution to different populations and settings, with specific considerations for cases of different contexts and users. Sijilli EHR has a crucial role in enhancing refugees' access to health services throughout the displacement and migration journey.

Acknowledgments

The research team would like to acknowledge Dr Joe Max Wakim, Mr Joseph Moukhtafi, Dr Dana Nabulsi, Dr Julien Lahoud, Mr Jeff Harman, and Mr Phil Ridzewski for their valuable contribution to this work.

Conflicts of Interest

None declared.

References

1. The UN Refugee Agency. Figures at a glance. URL: https://www.unhcr.org/figures-at-a-glance.html [accessed 2020-02-13]
2. Cohen J, Nussenzweig V, Nussenzweig R, Vekemans J, Leach A. From the circumsporozoite protein to the RTS, S/AS candidate vaccine. Hum Vaccin 2010 Jan;6(1):90-96. [doi: 10.4161/hv.6.1.9677] [Medline: 19806009]
3. Casares S, Brumeanu T, Richie TL. The RTS,S malaria vaccine. Vaccine 2010 Jul 12;28(31):4880-4894. [doi: 10.1016/j.vaccine.2010.05.033] [Medline: 20553771]
4. Silove D, Ventevogel P, Rees S. The contemporary refugee crisis: an overview of mental health challenges. World Psychiatry 2017 Jun;16(2):130-139 [FREE Full text] [doi: 10.1002/wps.20438] [Medline: 28498581]
5. The UN Refugee Agency. The Global Report. URL: https://www.unhcr.org/the-global-report.html [accessed 2020-02-14]
6. The UN Refugee Agency. Syria Regional Refugee Response. URL: https://data2.unhcr.org/en/situations/syria [accessed 2020-02-13]
7. Hansson E, Tuck A, Lurie S, McKenzie K. Improving mental health services for immigrant, refugee, ethno-cultural and racialized groups: Issues and options for service improvement. Ottawa, ON, Canada: Mental Health Commission of Canada; 2009. URL: https://www.mentalhealthcommission.ca/sites/default/files/Diversity_Issues_Options_Report_ENG_0_1.pdf [accessed 2020-07-22]
8. Pottie K, Greenaway C, Feightner J, Welch V, Swinkels H, Rashid M, coauthors of the Canadian Collaboration for Immigrant and Refugee Health. Evidence-based clinical guidelines for immigrants and refugees. CMAJ 2011 Sep 06;183(12):E824-E925 [FREE Full text] [doi: 10.1503/cmaj.090313] [Medline: 20530168]
9. Langlois EV, Haines A, Tomson G, Ghaffar A. Refugees: towards better access to health-care services. Lancet 2016 Jan 23;387(10016):319-321 [FREE Full text] [doi: 10.1016/S0140-6736(16)00101-X] [Medline: 26842434]

10. El Arnaout N, Rutherford S, Zreik T, Nabulsi D, Yassin N, Saleh S. Assessment of the health needs of Syrian refugees in Lebanon and Syria's neighboring countries. Confl Health 2019 Jun 27;13(1):31 [FREE Full text] [doi: 10.1186/s13031-019-0211-3] [Medline: 31297141]
11. Pavli A, Maltezou H. Health problems of newly arrived migrants and refugees in Europe. J Travel Med 2017 Jul 01;24(4):e. [doi: 10.1093/jtm/tax016] [Medline: 28426115]
12. Lengeler C. Insecticide-treated bed nets and curtains for preventing malaria. Cochrane Database Syst Rev 2004(2):CD000363. [doi: 10.1002/14651858.CD000363.pub2] [Medline: 15106149]
13. Arie S. How Europe keeps migrants out of its health system. BMJ 2015 Apr 24;350:h2216. [doi: 10.1136/bmj.h2216] [Medline: 25911414]
14. Haskew J, Rø G, Saito K, Turner K, Odhiambo G, Wamae A, et al. Implementation of a cloud-based electronic medical record for maternal and child health in rural Kenya. Int J Med Inform 2015 May;84(5):349-354. [doi: 10.1016/j.ijmedinf.2015.01.005] [Medline: 25670229]
15. Asi YM, Williams C. The role of digital health in making progress toward Sustainable Development Goal (SDG) 3 in conflict-affected populations. Int J Med Inform 2018 Jun;114:114-120. [doi: 10.1016/j.ijmedinf.2017.11.003] [Medline: 29126701]
16. Mathematica Policy Research, Harvard School of Public Health, Robert Wood Johnson Foundation. Health Information Technology in the United States: Better Information Systems for Better Care.: Robert Wood Johnson Foundation; 2013. URL: https://www.rwjf.org/en/library/research/2013/07/health-information-technology-in-the-united-states-2013.html [accessed 2020-07-22]
17. Birkhead GS, Klompas M, Shah NR. Uses of electronic health records for public health surveillance to advance public health. Annu Rev Public Health 2015 Mar 18;36:345-359. [doi: 10.1146/annurev-publhealth-031914-122747] [Medline: 25581157]
18. Rodríguez-Vera FJ, Marín Y, Sánchez A, Borrachero C, Pujol E. Illegible handwriting in medical records. J R Soc Med 2002 Nov;95(11):545-546 [FREE Full text] [doi: 10.1258/jrsm.95.11.545] [Medline: 12411618]
19. Winslow EH, Nestor VA, Davidoff SK, Thompson PG, Borum JC. Legibility and completeness of physicians' handwritten medication orders. Heart Lung 1997;26(2):158-164. [doi: 10.1016/s0147-9563(97)90076-5] [Medline: 9090521]
20. HIMSS Analytics. Electronic Medical Records vs. Electronic Health Records: Yes, There Is a Difference. URL: https://foundation.himss.org/Dorenfest [accessed 2019-01-17]
21. Xhafa F, Li J, Zhao G, Li J, Chen X, Wong DS. Designing cloud-based electronic health record system with attribute-based encryption. Multimed Tools Appl 2014 Feb 11;74(10):3441-3458. [doi: 10.1007/s11042-013-1829-6]
22. Prasad RB, Choi E, Lumb I. A taxonomy and survey of cloud computing systems. 2009 Presented at: Fifth International Joint Conference on INC, IMS and IDC; Aug 25-27, 2009; Seoul, South Korea. [doi: 10.1109/NCM.2009.218]
23. Mesmar S, Talhouk R, Akik C, Olivier P, Elhajj IH, Elbassuoni S, et al. The impact of digital technology on health of populations affected by humanitarian crises: Recent innovations and current gaps. J Public Health Policy 2016 Nov;37(Suppl 2):167-200. [doi: 10.1057/s41271-016-0040-1] [Medline: 27899794]
24. International Federation of Red Cross and Red Crescent Societies. World Disasters Report:Focus on technology and the future of humanitarian action. 2013. URL: https://www.ifrc.org/PageFiles/134658/WDR%202013%20complete.pdf [accessed 2020-01-15]
25. Mateen FJ, Carone M, Nyce S, Ghosn J, Mutuerandu T, Al-Saedy H, et al. Neurological disorders in Iraqi refugees in Jordan: data from the United Nations Refugee Assistance Information System. J Neurol 2012 Apr;259(4):694-701. [doi: 10.1007/s00415-011-6248-x] [Medline: 21952870]
26. Khader A, Farajallah L, Shahin Y, Hababeh M, Abu-Zayed I, Zachariah R, et al. Hypertension and treatment outcomes in Palestine refugees in United Nations Relief and Works Agency primary health care clinics in Jordan. Trop Med Int Health 2014 Oct;19(10):1276-1283 [FREE Full text] [doi: 10.1111/tmi.12356] [Medline: 25039838]
27. Ruby A, Knight A, Perel P, Blanchet K, Roberts B. The Effectiveness of Interventions for Non-Communicable Diseases in Humanitarian Crises: A Systematic Review. PLoS One 2015;10(9):e0138303 [FREE Full text] [doi: 10.1371/journal.pone.0138303] [Medline: 26406317]
28. Chiesa V, Chiarenza A, Mosca D, Rechel B. Health records for migrants and refugees: A systematic review. Health Policy 2019 Sep;123(9):888-900 [FREE Full text] [doi: 10.1016/j.healthpol.2019.07.018] [Medline: 31439455]
29. Schoevers M, van den Muijsenbergh M, Lagro-Janssen A. Patient-held records for undocumented immigrants: a blind spot. A systematic review of patient-held records. Ethn Health 2009 Oct;14(5):497-508. [doi: 10.1080/13557850902923273] [Medline: 19462264]
30. Ballout G, Al-Shorbaji N, Abu-Kishk N, Turki Y, Zeidan W, Seita A. UNRWA's innovative e-Health for 5 million Palestine refugees in the Near East. BMJ Innov 2018 Jun 08;4(3):128-134. [doi: 10.1136/bmjinnov-2017-000262]
31. Mateen FJ, Carone M, Al-Saedy H, Nyce S, Ghosn J, Mutuerandu T, et al. Medical conditions among Iraqi refugees in Jordan: data from the United Nations Refugee Assistance Information System. Bull World Health Organ 2012 Jun 01;90(6):444-451 [FREE Full text] [doi: 10.2471/BLT.11.097048] [Medline: 22690034]

32. Norwegian Refugee Council/Internal Displacement Monitoring Centre (NRC/IDMC). Legal Status of Refugees from Syria: Challenges And Consequences Of Maintaining Legal Stay In Beirut And Mount Lebanon. URL: https://www.refworld.org/docid/56cabc8c4.html [accessed 2020-01-13]
33. Ministry of Public Health, Lebanon. Syrian Refugees Crisis Impact on Lebanese Public Hospitals. URL: https://www.moph.gov.lb/en/Pages/127/9287/syrian-refugees-crisis-impact-on-lebanese-public-hospitals-financial-impact-analysis-apis-report [accessed 2019-12-28]
34. United Nations High Commissioner for Refugees, United Nations Children's Fund, World Food Programme. Vulnerability Assessment of Syrian Refugees in Lebanon. URL: https://www.unhcr.org/lb/wp-content/uploads/sites/16/2018/12/VASyR-2018.pdf [accessed 2019-12-27]
35. Ministry of Public Health, Lebanon. National PHC Network. URL: https://moph.gov.lb/en/Pages/0/8116/national-phc-network [accessed 2019-12-26]
36. Saleh S, El Arnaout N, Faulkner JR, Sayegh MH. Sijilli: a mobile electronic health records system for refugees in low-resource settings. Lancet Glob Health 2019 Sep;7(9):e1168-e1169 [FREE Full text] [doi: 10.1016/S2214-109X(19)30334-1] [Medline: 31401996]
37. Global Health Institute at American University of Beirut. Sijilli. URL: https://ghi.aub.edu.lb/sijilli/ [accessed 2020-08-04]
38. World Health Organization. International Statistical Classification of Diseases and Related Health Problems (10th Revision). URL: https://www.who.int/classifications/icd/ICD10Volume2_en_2010.pdf [accessed 2019-12-26]
39. Deglin J, Vallerand AH, Sanoski CA. Davis's drug guide for nurses. Philadelphia, PA: FA Davis Company; 2012.
40. Naja WJ, Aoun MP, El Khoury EL, Abdallah FJB, Haddad RS. Prevalence of depression in Syrian refugees and the influence of religiosity. Compr Psychiatry 2016 Jul;68:78-85. [doi: 10.1016/j.comppsych.2016.04.002] [Medline: 27234187]
41. Winters M, Rechel B, de Jong L, Pavlova M. A systematic review on the use of healthcare services by undocumented migrants in Europe. BMC Health Serv Res 2018 Jan 18;18(1):30 [FREE Full text] [doi: 10.1186/s12913-018-2838-y] [Medline: 29347933]
42. Hassan S, Vikanes A, Laine K, Zimmo K, Zimmo M, Bjertness E, et al. Building a research registry for studying birth complications and outcomes in six Palestinian governmental hospitals. BMC Pregnancy Childbirth 2017 Apr 11;17(1):112 [FREE Full text] [doi: 10.1186/s12884-017-1296-6] [Medline: 28399841]
43. Khader A, Ballout G, Shahin Y, Hababeh M, Farajallah L, Zeidan W, et al. Diabetes mellitus and treatment outcomes in Palestine refugees in UNRWA primary health care clinics in Jordan. Public Health Action 2013 Dec 21;3(4):259-264 [FREE Full text] [doi: 10.5588/pha.13.0083] [Medline: 26393043]
44. Khader A, Farajallah L, Shahin Y, Hababeh M, Abu-Zayed I, Kochi A, et al. Cohort monitoring of persons with hypertension: an illustrated example from a primary healthcare clinic for Palestine refugees in Jordan. Trop Med Int Health 2012 Sep;17(9):1163-1170 [FREE Full text] [doi: 10.1111/j.1365-3156.2012.03048.x] [Medline: 22845700]
45. Khader A, Farajallah L, Shahin Y, Hababeh M, Abu-Zayed I, Kochi A, et al. Cohort monitoring of persons with diabetes mellitus in a primary healthcare clinic for Palestine refugees in Jordan. Trop Med Int Health 2012 Dec;17(12):1569-1576 [FREE Full text] [doi: 10.1111/j.1365-3156.2012.03097.x] [Medline: 23051859]
46. Khader A, Ballout G, Shahin Y, Hababeh M, Farajallah L, Zeidan W, et al. Treatment outcomes in a cohort of Palestine refugees with diabetes mellitus followed through use of E-Health over 3 years in Jordan. Trop Med Int Health 2014 Feb;19(2):219-223 [FREE Full text] [doi: 10.1111/tmi.12241] [Medline: 24341942]
47. The UN Refugee Agency. Inter-sectoral Meeting. URL: https://data2.unhcr.org/en/documents/details/42681 [accessed 2019-12-23]
48. Taylor EM, Yanni EA, Pezzi C, Guterbock M, Rothney E, Harton E, et al. Physical and mental health status of Iraqi refugees resettled in the United States. J Immigr Minor Health 2014 Dec;16(6):1130-1137 [FREE Full text] [doi: 10.1007/s10903-013-9893-6] [Medline: 23959695]
49. Yunis K, Rafei RE, Mumtaz G. International Perspectives: Consanguinity: Perinatal Outcomes and Prevention - A View from the Middle East. NeoReviews 2008 Feb 01;9(2):e59-e65. [doi: 10.1542/neo.9-2-e59]
50. Lindert J, Carta MG, Schäfer I, Mollica RF. Refugees mental health-A public mental health challenge. Eur J Public Health 2016 Jun 06;26(3):374-375. [doi: 10.1093/eurpub/ckw010] [Medline: 27053728]
51. Connor P. Most displaced Syrians are in the Middle East, and about a million are in Europe.: Pew Research Center; 2018 Jan 29. URL: https://www.pewresearch.org/fact-tank/2018/01/29/where-displaced-syrians-have-resettled/ [accessed 2020-07-22]
52. Gedan BN. Venezuelan Migration: Is the Western Hemisphere Prepared for a Refugee Crisis? SAIS Review of International Affairs 2017;37(2):57-64. [doi: 10.1353/sais.2017.0027]
53. Syzdykova A, Malta A, Zolfo M, Diro E, Oliveira JL. Open-Source Electronic Health Record Systems for Low-Resource Settings: Systematic Review. JMIR Med Inform 2017 Nov 13;5(4):e44 [FREE Full text] [doi: 10.2196/medinform.8131] [Medline: 29133283]

Abbreviations

AES: advanced encrypted standard

AUB: American University of Beirut
EHR: electronic health record
EMR: electronic medical record
GHI: Global Health Institute
LMIC: low- and middle-income countries
MDD: major depressive disorder
MoPH: Ministry of Public Health
NCD: noncommunicable disease
OB/GYN: obstetrics and gynecology
PHC: primary health care center
PHQ: Patient Health Questionnaire
PIN: personal identification number
RAIS: Refugee Assistance Information System
UNHCR: United Nations High Commissioner for Refugees
UNRWA: United Nations Relief and Works Agency for Palestinian Refugees

Edited by G Eysenbach; submitted 10.02.20; peer-reviewed by H Mwanyika, I Contreras; comments to author 25.03.20; revised version received 19.05.20; accepted 03.06.20; published 13.08.20.

Please cite as:
Saleh S, El Arnaout N, Abdouni L, Jammoul Z, Hachach N, Dasgupta A
Sijilli: A Scalable Model of Cloud-Based Electronic Health Records for Migrating Populations in Low-Resource Settings
J Med Internet Res 2020;22(8):e18183
URL: http://www.jmir.org/2020/8/e18183/
doi:10.2196/18183
PMID:32788145

Viewpoint

e-INEBRIA Special Interest Group Roadmap for Best Practices for Research on Brief Digital Interventions for Problematic Alcohol and Illicit Drug Use

Michael Patrick Schaub[1], PhD; Anne H Berman[2,3,4], PhD; Hugo López Pelayo[5,6,7], PhD; Nikolaos Boumparis[8], MA; Zarnie Khadjesari[9], PhD; Matthijs Blankers[10,11,12], PhD; Antoni Gual[13], MD; Heleen Riper[14], PhD; Lodewijk Pas[15], MD

[1]Swiss Research Institute for Public Health and Addiction ISGF, University of Zurich, Zurich, Switzerland

[2]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

[3]Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden

[4]Department of Psychology, Uppsala University, Uppsala, Sweden

[5]Grup Recerca Addicions Clínic (GRAC-GRE), Department of Psychiatry, Clinical Institute of Neuroscience, Hospital Clínic i Universitari de Barcelona, Universitat de Barcelona, Barcelona, Spain

[6]Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

[7]La Red de Trastornos Adictivos, RETICS, Madrid, Spain

[8]Department of Clinical, Neuro- and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[9]School of Health Sciences, University of East Anglia, Norwich, United Kingdom

[10]Trimbos Institute, The Netherlands Institute of Mental Health and Addiction, Utrecht, Netherlands

[11]Department of Research, Arkin Mental Health Care, Amsterdam, Netherlands

[12]Amsterdam UMC, Location AMC, Department of Psychiatry, University of Amsterdam, Amsterdam, Netherlands

[13]Hospital Clinic, Psychiatry Department, Neurosciences Institute, Barcelona, Spain

[14]Department of Clinical, Neuro- and Developmental Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[15]Academic Center General Practice, Catholic University Leuven, Leuven, Belgium

Corresponding Author:
Michael Patrick Schaub, PhD
Swiss Research Institute for Public Health and Addiction ISGF
University of Zurich
Konradstrasse 32
Zurich, 8005
Switzerland
Phone: 41 (0) 44 4481165
Email: michael.schaub@isgf.uzh.ch

Abstract

There is great potential for scaling up the delivery of brief interventions for alcohol and illicit drug use, given the increasing coverage of mobile devices and technologies for digital interventions, including apps for smartphones and tablets. However, while the number of digital interventions is increasing rapidly, the involvement of brief-intervention researchers and the development of good practices has just begun. In 2018, the Special Interest Group on digital interventions of the International Network on Brief Interventions for Alcohol & Other Drugs (e-INEBRIA SIG) initiated a conversation regarding possible avenues of future research, which subsequently became a roadmap for digital interventions. This roadmap consists of points considered relevant for future research, ongoing technological developments, and their implementation across a continuum of prevention and care. Moreover, it outlines starting points for the diversification of brief digital interventions, as well as next steps for quality improvement and implementation in public health and clinical practice. The roadmap of the e-INEBRIA SIG on digital interventions is a starting point that indicates relevant next steps and provides orientation for researchers and interested practitioners with regard to the ambiguous literature and the complexity of current digital interventions.

(J Med Internet Res 2020;22(8):e20368) doi:10.2196/20368

KEYWORDS

brief interventions; mobile applications; good practice; implementation research; quality assurance

Background

The Special Interest Group on digital interventions of the International Network on Brief Interventions for Alcohol & Other Drugs (e-INEBRIA SIG) was launched at the INEBRIA conference in Lausanne (2016); the group also took part in INEBRIA conferences in 2017 (New York) and 2018 (Santiago de Chile). During the conference in Santiago de Chile, e-INEBRIA SIG discussed possible avenues for further research that subsequently, in continued meetings, turned into a roadmap, as presented in this commentary. In particular, e-INEBRIA SIG is interested in the potential of electronic health (eHealth) technologies for brief interventions in health and other settings to reduce the harms produced by alcohol, drug use, and gambling.

There is great potential for scaling up the delivery of brief interventions, given the increasing coverage in high, middle, and low income countries of smartphones, wearables, and other connected devices worldwide, and the rise of new technologies like geotagging and chatbots. Knowing which technologies can be successfully applied in target groups of alcohol and drug users and which components of interventions are essential for reducing harm could help enhance their impact at a public health level.

One strategy to foster evidence-based developments is to keep track of ongoing scientific developments, including periodic reviews and meta-analyses on the effectiveness and implementation of internet interventions for alcohol and other substances. Recent meta-analyses have demonstrated the effectiveness of internet interventions at reducing alcohol use in alcohol misusers [1]; reducing cannabis use in cannabis misusers, at least in the short-term [2]; and reducing drug use among opioid and mixed-drug users [3]. Moreover, individual patient data (IPD) meta-analyses could, due to their large pooled samples, help to identify initial trends in subgroups (eg, [1]).

However, a limitation of most digital interventions developed as part of scientific studies is that they often lack funding for maintenance after study completion. A recent review on such interventions in the alcohol field identified 72 trials assessing alcohol interventions, among which only 8 (11%) remained accessible; these were mostly internet-based interventions [4].

While numerous digital interventions aiming to reduce substance use are available, most lack evidence-based brief intervention content, have never been evaluated for effectiveness, and have been developed without academic or specialist input, including some that are profit-oriented [5,6]. In the case of cannabis misuse, the situation is worse because the few genuine apps aimed at harm reduction and reducing cannabis use are forced to compete with hundreds of currently available apps that promote cannabis use [7]. Some attempts have been made to develop evidence-based health app quality rating scales, and one has been developed for researchers [8], with a version available for end users [9]; however, it is unlikely that alcohol and drug users make use of the latter. Thus, it is rational to foster research on genuine harm-reduction–oriented digital interventions, particularly apps, to promote their use and evidence-based quality development.

While it is rational to develop generalizable digital interventions to reduce harm from alcohol and drug use at a general population level, there is evidence from other mental health fields that tailored digital interventions for specific cultures or migrants [10] can increase their effectiveness. There are few systematic reviews of tailored digital interventions for specific alcohol- or drug-using groups and classical diversification topics (eg, different age groups, like the elderly, as well as different genders, educational levels, and cultures). The development of such digital interventions in the addiction field is still in its infancy relative to face-to-face interventions.

Roadmap for Brief Digital Interventions

Aims

Taking into consideration the complexity of current brief digital interventions, the aim of the roadmap is to propose further steps and provide orientation in the ambiguous literature for researchers and interested practitioners thus promoting the coordinated, systematic development of a knowledge base that will facilitate the effective development and application of digital interventions.

Methods

The members of e-INEBRIA SIG represent 6 European countries, have previously developed and evaluated digital interventions for alcohol and other drugs, and exchange information and plan further research at the annual INEBRIA conference. We developed a preliminary roadmap during the last INEBRIA preconference workshop in 2018, in Santiago, Chile, and refined it during subsequent bimonthly teleconferences and email exchanges. The roadmap was then organized into four main topic groups and, once again, discussed within the group.

Topics

Evaluation of Effective Implementation Modalities

- Continue evaluation research efforts (start to construct an ongoing collection of overall data sets for IPD meta-analyses; begin with meta-analyses for different interventions and target groups)
- Investigate which elements make digital interventions more effective, particularly to identify components that improve adherence, enhance motivation, and/or promote sustained behavior change
- Investigate how brief digital interventions should be integrated and can enhance different forms of blended or face-to-face treatments for alcohol and drug use disorders

(with continuous electronic monitoring or personalized feedback; investigate the potentials and limitations of motivational interviewing)
- Investigate the complementarity between anonymous applications for public health and applications for the clinical setting, including hazardous or harmful use recognition and identification; behavior change support; and continued harm reduction
- Investigate public health and clinical applications for specific vulnerable populations (eg, gender, age, cultural issues, personal differences, differences in contexts and settings).
- Develop a core set of validated outcome measures suitable for use in digital brief intervention research, depending on the targeted substance

Diversification and Cultural Sensitivity

- Develop and investigate easy-to-use and tablet-optimized guided and unguided brief interventions for the elderly, with extended content and adapted measures
- Foster the development of culturally generalizable interventions for newly industrialized countries, in collaboration with the World Health Organization (WHO) and other relevant international organizations
- Start to adapt brief digital interventions for specific cultures and minorities to investigate if cultural adaptation makes a difference in an intervention's acceptance and effectiveness

Accommodation of New Technology

- Encourage research on novel technologies, such as chatbots, for brief internet interventions to allow for more interactivity, including geotagging functions (eg, to warn against places that are risky for relapses or to network with supporting peers or social workers in the immediate proximity at concert events [5])
- Clarify whether and how brief digital interventions can profit from gamification and for which target groups
- Develop and investigate just-in-time adaptive brief digital interventions
- Make use of big data technologies (eg, as an additional data source for missing data within a classical randomized controlled trial; to derive a natural additional control group in addition to existing groups in a classical randomized controlled trial design; or to gain better access to risk groups for brief digital interventions at population levels).
- Develop an evidence base of effective implementation strategies, where training, support, and financial incentives are the most effective strategies for implementing face-to-face brief interventions

Intervention Quality and Safety Management

- Develop quality criteria for digital interventions to change alcohol use and illicit drug use and promote harm reduction
- Develop a directive on collaboration between research and commercial interests, particularly for publicly funded organizations and private collaborations
- Further develop the taxonomy for describing brief digital interventions in a standardized way
- Develop best practices for professionals working with brief digital interventions
- Keep abreast of ongoing quality developments, like genuine health app rating websites

Discussion

While this roadmap provides orientation to current developments in this field, it is also considered a starting point. It will need updating and ongoing promotion at future meetings and conferences, as well as in publications. Moreover, we must ensure that the points listed in our roadmap will be followed by actions. Therefore, we plan to continue our discussions between future INEBRIA conferences, during symposia and workshops at these conferences, and at related conferences like those of the International Society for Research on Internet Interventions (ISRII) and the International Society of Behavioral Medicine (ISBM), where alcohol and drug use can be linked to relevant developments in digital interventions (ISRII), epidemiology, and interventions for health-related behaviors and chronic conditions (ISBM).

Conclusions

The roadmap proposed by e-INEBRIA SIG on digital interventions is a starting point that indicates relevant next steps. It also provides orientation for researchers and interested practitioners with regard to the ambiguous literature and the complexity of current digital interventions. Moreover, it is a call for action to coordinate our efforts toward research and evidence-based implementation in the context of rapid technological development.

Authors' Contributions

MPS and HR had the initial idea to construct this roadmap; MPS prepared the first draft of the manuscript; and all authors contributed to the manuscript.

Conflicts of Interest

AHB is the co-owner of a company (TeleCoach AB, not currently active) that aims to disseminate digital interventions for problematic behaviors, including hazardous and harmful alcohol use. AG has received a grant from Novartis. All other authors declare no conflicts of interest.

References

1. Riper H, Hoogendoorn A, Cuijpers P, Karyotaki E, Boumparis N, Mira A, et al. Effectiveness and treatment moderators of internet interventions for adult problem drinking: An individual patient data meta-analysis of 19 randomised controlled trials. PLoS Med 2018 Dec;15(12):e1002714 [FREE Full text] [doi: 10.1371/journal.pmed.1002714] [Medline: 30562347]
2. Boumparis N, Loheide-Niesmann L, Blankers M, Ebert DD, Korf D, Schaub MP, et al. Short- and long-term effects of digital prevention and treatment interventions for cannabis use reduction: A systematic review and meta-analysis. Drug Alcohol Depend 2019 Jul 01;200:82-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.03.016] [Medline: 31112834]
3. Boumparis N, Karyotaki E, Schaub MP, Cuijpers P, Riper H. Internet interventions for adult illicit substance users: a meta-analysis. Addiction 2017 Sep;112(9):1521-1532 [FREE Full text] [doi: 10.1111/add.13819] [Medline: 28295758]
4. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-Delivered Health Interventions That Work: Systematic Review of Meta-Analyses and Evaluation of Website Availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
5. Schaub M, Yi-Chen LJ, Pirona A. m-Health applications for responding to drug use and associated harms. In: EMCDDA PAPERS. Brussels: Publications Office of the European Union; 2018.
6. Tofighi B, Chemi C, Ruiz-Valcarcel J, Hein P, Hu L. Smartphone Apps Targeting Alcohol and Illicit Substance Use: Systematic Search in in Commercial App Stores and Critical Content Analysis. JMIR mHealth uHealth 2019 Apr 22;7(4):e11831 [FREE Full text] [doi: 10.2196/11831] [Medline: 31008713]
7. Ramo DE, Popova L, Grana R, Zhao S, Chavez K. Cannabis Mobile Apps: A Content Analysis. JMIR mHealth uHealth 2015 Aug 12;3(3):e81 [FREE Full text] [doi: 10.2196/mhealth.4405] [Medline: 26268634]
8. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
9. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR mHealth uHealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
10. Harper Shehadeh M, Heim E, Chowdhary N, Maercker A, Albanese E. Cultural Adaptation of Minimally Guided Interventions for Common Mental Disorders: A Systematic Review and Meta-Analysis. JMIR Ment Health 2016 Sep 26;3(3):e44 [FREE Full text] [doi: 10.2196/mental.5776] [Medline: 27670598]

Abbreviations

eHealth: electronic health
e-INEBRIA SIG: Special Interest Group on digital interventions of the International Network on Brief Interventions for Alcohol & Other Drugs (INEBRIA)
IPD: individual patient data
ISBM: International Society of Behavioral Medicine
ISRII: International Society for Research on Internet Interventions
WHO: World Health Organization

Edited by G Eysenbach; submitted 18.05.20; peer-reviewed by E Grekin, A Moyer; comments to author 12.06.20; accepted 25.06.20; published 14.08.20.

Please cite as:
Schaub MP, Berman AH, López Pelayo H, Boumparis N, Khadjesari Z, Blankers M, Gual A, Riper H, Pas L
e-INEBRIA Special Interest Group Roadmap for Best Practices for Research on Brief Digital Interventions for Problematic Alcohol and Illicit Drug Use
J Med Internet Res 2020;22(8):e20368
URL: http://www.jmir.org/2020/8/e20368/
doi:10.2196/20368
PMID:32586786

Viewpoint

Is Artificial Intelligence Better Than Human Clinicians in Predicting Patient Outcomes?

Joon Lee[1,2,3], PhD

[1]Data Intelligence for Health Lab, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[2]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[3]Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Joon Lee, PhD
Data Intelligence for Health Lab
Cumming School of Medicine
University of Calgary
3280 Hospital Dr NW
TRW 5E17
Calgary, AB, T2N 4Z6
Canada
Phone: 1 403 220 2968
Email: joonwu.lee@ucalgary.ca

Abstract

In contrast with medical imaging diagnostics powered by artificial intelligence (AI), in which deep learning has led to breakthroughs in recent years, patient outcome prediction poses an inherently challenging problem because it focuses on events that have not yet occurred. Interestingly, the performance of machine learning–based patient outcome prediction models has rarely been compared with that of human clinicians in the literature. Human intuition and insight may be sources of underused predictive information that AI will not be able to identify in electronic data. Both human and AI predictions should be investigated together with the aim of achieving a human-AI symbiosis that synergistically and complementarily combines AI with the predictive abilities of clinicians.

(J Med Internet Res 2020;22(8):e19918) doi:10.2196/19918

KEYWORDS

patient outcome prediction; artificial intelligence; machine learning; human-generated predictions; human-AI symbiosis

Absence of Human-Generated Predictions in Patient Outcome Research

In recent years, there has been a proliferation of patient outcome prediction research that applies machine learning (ML) and artificial intelligence (AI) to electronic health records (EHRs) and other clinical and administrative health data. The central premises are that 1) complex health data contains predictive information that ML can effectively extract and transform into a predictive algorithm and 2) accurate prediction of patient outcomes can facilitate early, preventative intervention and more efficient health care resource allocation through identification of high-risk patients. For example, predicting which intensive care unit patients are likely to develop sepsis can prompt early initiation of fluid resuscitation, vasopressor therapy, or antibiotics, which can reduce damage from insufficient organ perfusion [1,2]. Although AI has been enormously successful in medical imaging diagnostics, where the medical condition of interest is already present or absent in the images (eg, diagnosis of diabetic retinopathy [3] and classification of skin legions [4]), patient outcome prediction poses an inherent challenge of predicting events that have *notyet* occurred (eg, mortality, length of stay, and readmission) [5]. This challenge is common to both AI and human clinicians.

Interestingly, while human and AI predictions are often directly compared in medical imaging research [6-8], patient outcome prediction studies tend to focus only on ML and seldom investigate human predictions. This is corroborated by a number of systematic reviews and meta-analyses, which target only ML methods [9-14] or empirical methods [15-19]. This gap in the literature is coherent across a wide range of medical specialties and diseases, including trauma [9], cancer [11], neurosurgery [10], depression [12], acute gastrointestinal bleeding [13], sepsis [14], acute liver failure [15], ischemic stroke [16], thermal injury

[17], and cardiovascular disease [18,19]. The absence of human predictions appears to be a recent trend, as older literature prior to the current widespread use of modern ML and EHRs includes more comparisons of human and AI predictions [20-23].

There are several possible reasons why human performance is more frequently studied in medical imaging than in patient outcome prediction. First, radiologists are trained to analyze, interpret, and classify images, whereas most other medical specialists are not trained to directly predict patient outcomes. While accurate prognostic information can certainly be helpful in any medical specialty, it is usually generated by empirical risk scoring systems such as the Framingham Risk Score [24] or Acute Physiology and Chronic Health Evaluation (APACHE) [25] rather than by human clinicians. Second, human predictions in medical imaging are readily available from routine clinical practice or can be generated systematically by trained radiologists. Conversely, it is rare for clinicians in other medical specialties to record patient outcome predictions that *they* generate on a regular basis. Third, the implicit assumption is that humans cannot accurately predict patient outcomes because analysis of complex, high-dimensional clinical data may be required; moreover, recall bias is rampant in the human mind.

Humans and AI Should Work as a Team

However, there is no reason to rule out the possibility that human clinicians can outperform AI in patient outcome prediction, at least in some clinical scenarios. While AI can only access information that can be recorded in the form of electronic data, human clinicians interact face-to-face with their patients and have access to both clinical and contextual information. The qualitative information collected via clinicians' five senses can be critical in patient outcome prediction; however, this information is mostly absent in EHRs, if it is possible to record it at all. Although some qualitative observations can be recorded in EHRs as free-text notes, such as nursing notes, these data are logged in a limited, inconsistent fashion. Human intuition and insight may well be the most underused resources in patient outcome prediction.

While the performance of ML-based patient outcome prediction models appears impressive on paper, the most accurately predicted cases tend to be "easy" cases where the likely outcomes are already obvious to human clinicians [26]. This further supports the hypothesis that human clinicians perform well in patient outcome prediction.

On the other hand, AI easily outperforms humans in processing, analyzing, and finding patterns in complex, high-dimensional data [27]. As demonstrated by IBM Watson [28] and AlphaGo [29], the memory, attention, and information processing abilities of AI vastly exceed the capabilities of human cognition [30]. This AI advantage is crucial for extracting and using data-driven insights from big data [31]; it is also key to the recent successful breakthroughs in ML, particularly in deep learning [32], in a number of problem domains, including medical imaging [33]. In addition, AI does not suffer from fatigue [34] or cognitive biases (eg, recall bias) [35] as humans do. However, even if AI outperforms human clinicians in patient outcome prediction, human performance represents a more meaningful benchmark that puts AI performance in better perspective. Understanding the superiority of AI in comparison with humans can facilitate adoption of AI technology in real patient care.

The bottom line is that both AI and humans can make unique contributions to patient outcome prediction, and they should help each other to maximize predictive performance. Patient outcome prediction research should aim for human-AI symbiosis, where the respective predictive abilities of AI and human clinicians are combined in a synergistic and complementary way [36]. Given the challenging nature of patient outcome prediction, creating an AI to act alone without human help will simply lead to suboptimal predictive performance because even state-of-the-art ML technology cannot leverage information that is not present in the data [26].

Another way for AI and humans to work together is via the human-in-the-loop model, where humans directly inform machines on how to learn from the data at hand by providing guidance based on human intuition and knowledge. The term "interactive machine learning" [37] was coined to describe this paradigm; it encompasses more well-known branches of ML, such as active learning, where humans select which data points should be labelled. This human-in-the-loop approach can greatly reduce the computational complexity of some ML problems; for example, it has shown promising results in protein folding [38]. Moreover, in the field of human-computer interaction, the human-in-the-loop concept has been studied in the context of vehicle control [39], security [40,41], and decision-making [40,42]. Knowledge from these application areas can potentially inform the design of human-AI symbiosis in patient outcome prediction.

AI and human prediction performance may vary across different types of patients. Complex patterns in data can be more predictive than human intuition in certain patient subgroups, and the opposite may be true in other subpopulations. An investigation of how AI and human predictions can be optimally combined for different types of patients could directly contribute to advancing precision medicine. A better understanding of the respective predictive powers of AI and humans in various clinical scenarios can also help increase human trust in AI (eg, "For this type of patient, I need to trust AI more because most predictive information is buried in the complex data"). This can facilitate evidence-based adoption of AI technology.

For human clinicians to completely trust AI, it is necessary to understand *why* an algorithm arrives at a given conclusion; this requires transparency, traceability, and causality. The active field of explainable AI has been producing useful methods, such as SHapley Additive exPlanations (SHAP) [43], that can help explain how ML models work at an algorithmic level (this explanation is almost always based on correlation rather than causation); however, human clinicians ultimately want to elevate this algorithmic explainability to a model that is understandable by humans with sufficient causal understanding, also known as causability [44]. Therefore, mapping explainability to causability will be key in achieving true human-AI symbiosis.

One major roadblock to the proposed human-AI symbiosis is the need to collect a large number of human predictions in a variety of clinical scenarios, which is labor-intensive and adds

to clinicians' workloads. Seamlessly integrated electronic prediction collection platforms (eg, embedded in a multi-center EHR system) can minimize this burden and enable large-scale prediction collection.

From Patient Outcome Prediction to Real Impact

Once predictive performance is optimized via human-AI symbiosis, the next important step is to formulate clinical guidelines so that the predictive information is actionable. This is a crucial step, as accurate predictions alone will not lead to any real impact; rather, the combination of accurate predictions and appropriate interventions by clinicians will have a greater effect [5,26].

The ultimate goal of patient outcome prediction is to improve patient outcomes and decrease health care costs through early intervention and efficient use of health care resources. To prove that this goal has been met, we will need to perform randomized clinical trials of AI-driven patient care [45], such as that conducted by Wijnberge and colleagues [46]. In addition to simply comparing AI with human work alone, these randomized clinical trials should investigate a promising third species: human-AI symbiosis.

Acknowledgments

The author would like to thank the University of Calgary for institutional support.

Conflicts of Interest

None declared.

References

1. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. JMIR Med Inform 2016 Sep 30;4(3):e28 [FREE Full text] [doi: 10.2196/medinform.5909] [Medline: 27694098]
2. Lee J, Mark RG. An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care. Biomed Eng Online 2010;9(1):62. [doi: 10.1186/1475-925x-9-62]
3. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
4. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Jan 25;542(7639):115-118. [doi: 10.1038/nature21056]
5. Lindsell CJ, Stead WW, Johnson KB. Action-Informed Artificial Intelligence-Matching the Algorithm to the Problem. JAMA 2020 May 01;323(21):2141. [doi: 10.1001/jama.2020.5035] [Medline: 32356878]
6. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 2019 Oct;1(6):e271-e297. [doi: 10.1016/s2589-7500(19)30123-2]
7. Emblem KE, Nedregaard B, Hald JK, Nome T, Due-Tonnessen P, Bjornerud A. Automatic glioma characterization from dynamic susceptibility contrast imaging: brain tumor segmentation using knowledge-based fuzzy clustering. J Magn Reson Imaging 2009 Jul;30(1):1-10. [doi: 10.1002/jmri.21815] [Medline: 19557840]
8. Emblem KE, Pinho MC, Zöllner FG, Due-Tonnessen P, Hald JK, Schad LR, et al. A generic support vector machine model for preoperative glioma survival associations. Radiology 2015 Apr;275(1):228-234. [doi: 10.1148/radiol.14140770] [Medline: 25486589]
9. Liu NT, Salinas J. Machine Learning for Predicting Outcomes in Trauma. Shock 2017 Nov;48(5):504-510. [doi: 10.1097/SHK.0000000000000898] [Medline: 28498299]
10. Senders JT, Staples PC, Karhade AV, Zaki MM, Gormley WB, Broekman ML, et al. Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review. World Neurosurg 2018 Jan;109:476-486.e1. [doi: 10.1016/j.wneu.2017.09.149] [Medline: 28986230]
11. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8-17 [FREE Full text] [doi: 10.1016/j.csbj.2014.11.005] [Medline: 25750696]
12. Lee Y, Ragguett R, Mansur R, Boutilier J, Rosenblat J, Trevizol A, et al. Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review. J Affect Disord 2018 Dec 01;241:519-532. [doi: 10.1016/j.jad.2018.08.073] [Medline: 30153635]
13. Shung D, Simonov M, Gentry M, Au B, Laine L. Machine Learning to Predict Outcomes in Patients with Acute Gastrointestinal Bleeding: A Systematic Review. Dig Dis Sci 2019 Aug;64(8):2078-2087. [doi: 10.1007/s10620-019-05645-z] [Medline: 31055722]

14. Islam M, Nasrin T, Walther B, Wu C, Yang H, Li Y. Prediction of sepsis patients using machine learning approach: A meta-analysis. Comput Methods Programs Biomed 2019 Mar;170:1-9. [doi: 10.1016/j.cmpb.2018.12.027] [Medline: 30712598]
15. Wlodzimirow KA, Eslami S, Chamuleau RAFM, Nieuwoudt M, Abu-Hanna A. Prediction of poor outcome in patients with acute liver failure-systematic review of prediction models. PLoS One 2012;7(12):e50952 [FREE Full text] [doi: 10.1371/journal.pone.0050952] [Medline: 23272081]
16. Fahey M, Crayton E, Wolfe C, Douiri A. Clinical prediction models for mortality and functional outcome following ischemic stroke: A systematic review and meta-analysis. PLoS One 2018;13(1):e0185402 [FREE Full text] [doi: 10.1371/journal.pone.0185402] [Medline: 29377923]
17. Hussain A, Choukairi F, Dunn K. Predicting survival in thermal injury: a systematic review of methodology of composite prediction models. Burns 2013 Aug;39(5):835-850. [doi: 10.1016/j.burns.2012.12.010] [Medline: 23384617]
18. van Dieren S, Beulens J, Kengne A, Peelen L, Rutten G, Woodward M, et al. Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review. Heart 2012:360-369. [doi: 10.1136/heartjnl-2011-300734]
19. Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ 2016 May 16;353:i2416 [FREE Full text] [doi: 10.1136/bmj.i2416] [Medline: 27184143]
20. Grove W, Zald D, Lebow B, Snitz B, Nelson C. Clinical versus mechanical prediction: A meta-analysis. Psychol Assess 2000;12(1):19-30. [doi: 10.1037/1040-3590.12.1.19]
21. Gardner W, Lidz CW, Mulvey EP, Shaw EC. Clinical versus actuarial predictions of violence in patients with mental illnesses. J Consult Clin Psychol 1996;64(3):602-609. [doi: 10.1037/0022-006x.64.3.602]
22. Marchese MC. Clinical versus actuarial prediction: a review of the literature. Percept Mot Skills 1992 Oct;75(2):583-594. [Medline: 1408625]
23. Bandiera G, Stiell IG, Wells GA, Clement C, De Maio V, Vandemheen KL, et al. The Canadian C-Spine rule performs better than unstructured physician judgment. Ann Emerg Med 2003 Sep;42(3):395-402. [doi: 10.1016/s0196-0644(03)00422-0]
24. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008 Feb 12;117(6):743-753. [doi: 10.1161/CIRCULATIONAHA.107.699579] [Medline: 18212285]
25. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients*. Crit Care Med 2006;34(5):1297-1310. [doi: 10.1097/01.ccm.0000215112.84523.f0]
26. Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. N Engl J Med 2017 Jun 29;376(26):2507-2509 [FREE Full text] [doi: 10.1056/NEJMp1702071] [Medline: 28657867]
27. Jarrahi MH. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Bus Horiz 2018 Jul;61(4):577-586. [doi: 10.1016/j.bushor.2018.03.007]
28. Ferrucci D, Brown E, Chu-Carroll J, Fan J, Gondek D, Kalyanpur A, et al. Building Watson: An Overview of the DeepQA Project. AIMag 2010 Jul 28;31(3):59-79. [doi: 10.1609/aimag.v31i3.2303]
29. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016 Jan 28;529(7587):484-489. [doi: 10.1038/nature16961] [Medline: 26819042]
30. Grigsby SS. Artificial intelligence for advanced human-machine symbiosis. In: Augmented Cognition: Intelligent Technologies. AC 2018. Lecture Notes in Computer Science. 2018 Presented at: International Conference on Augmented Cognition; July 15-20, 2018; Las Vegas, NV p. 255-266. [doi: 10.1007/978-3-319-91470-1_22]
31. L'Heureux A, Grolinger K, Elyamany H, Capretz M. Machine Learning With Big Data: Challenges and Approaches. IEEE Access 2017;5:7776-7797. [doi: 10.1109/access.2017.2696365]
32. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
33. Martín Noguerol T, Paulano-Godino F, Martín-Valdivia MT, Menias CO, Luna A. Strengths, Weaknesses, Opportunities, and Threats Analysis of Artificial Intelligence and Machine Learning Applications in Radiology. J Am Coll Radiol 2019 Sep;16(9 Pt B):1239-1247. [doi: 10.1016/j.jacr.2019.05.047] [Medline: 31492401]
34. Hockey G, Wiethoff M. Cognitive fatigue in complex decision-making. Adv Space Biol Med 1993:139-150. [doi: 10.1016/S1569-2574(08)60101-X]
35. Dawson NV, Arkes HR. Systematic errors in medical decision making:. J Gen Intern Med 1987 May;2(3):183-187. [doi: 10.1007/bf02596149]
36. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan 7;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
37. Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inform 2016 Jun;3(2):119-131 [FREE Full text] [doi: 10.1007/s40708-016-0042-6] [Medline: 27747607]
38. Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, et al. Predicting protein structures with a multiplayer online game. Nature 2010 Aug 05;466(7307):756-760 [FREE Full text] [doi: 10.1038/nature09304] [Medline: 20686574]

39. Driggs-Campbell K, Shia V, Bajcsy R. Improved driver modeling for human-in-the-loop vehicular control. In: 2015 IEEE International Conference on Robotics and Automation (ICRA). 2015 Presented at: IEEE International Conference on Robotics and Automation (ICRA); 2015; Seattle, WA p. 1654-1661. [doi: 10.1109/icra.2015.7139410]
40. Schumann MA, Drusinsky D, Michael JB, Wijesekera D. Modeling Human-in-the-Loop Security Analysis and Decision-Making Processes. IIEEE Trans. Software Eng 2014 Feb;40(2):154-166. [doi: 10.1109/tse.2014.2302433]
41. Cranor LF. A Framework for Reasoning About the Human in the Loop. usenix.org. 2008. URL: https://www.usenix.org/legacy/event/upsec/tech/full_papers/cranor/cranor.pdf [accessed 2020-07-24]
42. Subramania HS, Khare VR. Pattern classification driven enhancements for human-in-the-loop decision support systems. Decision Support Systems 2011 Jan;50(2):460-468. [doi: 10.1016/j.dss.2010.11.003]
43. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell 2020 Jan;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
44. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 2019;9(4):e1312 [FREE Full text] [doi: 10.1002/widm.1312] [Medline: 32089788]
45. Angus DC. Randomized Clinical Trials of Artificial Intelligence. JAMA 2020 Feb 17;323(11):1043-1045. [doi: 10.1001/jama.2020.1039] [Medline: 32065828]
46. Wijnberge M, Geerts B, Hol L, Lemmers N, Mulder M, Berge P, et al. Effect of a Machine Learning-Derived Early Warning System for Intraoperative Hypotension vs Standard Care on Depth and Duration of Intraoperative Hypotension During Elective Noncardiac Surgery: The HYPE Randomized Clinical Trial. JAMA 2020 Feb 17:1052-1060. [doi: 10.1001/jama.2020.0592] [Medline: 32065827]

Abbreviations

AI: artificial intelligence
APACHE: Acute Physiology And Chronic Health Evaluation
EHR: electronic health record
ML: machine learning
SHAP: SHapley Additive exPlanations

Edited by G Eysenbach; submitted 06.05.20; peer-reviewed by A Holzinger, Z Ge; comments to author 12.06.20; revised version received 24.06.20; accepted 25.06.20; published 26.08.20.

Please cite as:
Lee J
Is Artificial Intelligence Better Than Human Clinicians in Predicting Patient Outcomes?
J Med Internet Res 2020;22(8):e19918
URL: http://www.jmir.org/2020/8/e19918/
doi:10.2196/19918
PMID:32845249

Viewpoint

The Need to Develop Standard Measures of Patient Adherence for Big Data: Viewpoint

Przemyslaw Kardas[1], Prof Dr, MD, PhD; Isabel Aguilar-Palacio[2,3], MD, MPH, PhD; Marta Almada[4], PhD; Caitriona Cahir[5], BA, HDip, PhD; Elisio Costa[4], Prof Dr, PhD; Anna Giardini[6], PsyD, MSc; Sara Malo[2,3], PharmB, PhD; Mireia Massot Mesquida[7], PharmB; Enrica Menditto[8,9], PharmD, PhD; Luís Midão[4], MSc; Carlos Luis Parra-Calderón[10], MSc; Enrique Pepiol Salom[11], MPharm, PGCert; Bernard Vrijens[12,13], PhD

[1]Department of Family Medicine, Medical University of Lodz, Lodz, Poland

[2]Preventive Medicine and Public Health Department, Zaragoza University, Zaragoza, Spain

[3]Fundación Instituto de Investigación Sanitaria de Aragón (IIS Aragón), Zaragoza, Spain

[4]UCIBIO REQUIMTE, ICBAS, Porto4Ageing - Competences Center on Active and Healthy Ageing, Faculty of Pharmacy, University of Porto, Porto, Portugal

[5]Division of Population Health Sciences, Royal College of Surgeons in Ireland, Dublin, Ireland

[6]IT Department, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy

[7]Servei d'Atenció Primària Vallès Occidental, Institut Català de la Salut, Barcelona, Spain

[8]CIRFF, Center of Pharmacoeconomics, University of Naples Federico II, Naples, Italy

[9]Department of Pharmacy, University of Naples Federico II, Naples, Italy

[10]Group of Research and Innovation in Biomedical Informatics, Biomedical Engineering and Health Economy, Institute of Biomedicine of Seville, IBiS / Virgen del Rocío University Hospital / CSIC / University of Seville, Seville, Spain

[11]International Commitee, Muy Ilustre Colegio Oficial de Farmacéuticos, Valencia, Spain

[12]AARDEX Group, Seraing, Belgium

[13]Liège University, Liège, Belgium

Corresponding Author:
Przemyslaw Kardas, Prof Dr, MD, PhD
Department of Family Medicine
Medical University of Lodz
60, Narutowicza St.
Lodz, 90-136
Poland
Phone: 48 6787210
Email: pkardas@csk.am.lodz.pl

Abstract

Despite half a century of dedicated studies, medication adherence remains far from perfect, with many patients not taking their medications as prescribed. The magnitude of this problem is rising, jeopardizing the effectiveness of evidence-based therapies. An important reason for this is the unprecedented demographic change at the beginning of the 21st century. Aging leads to multimorbidity and complex therapeutic regimens that create a fertile ground for nonadherence. As this scenario is a global problem, it needs a worldwide answer. Could this answer be provided, given the new opportunities created by the digitization of health care? Daily, health-related information is being collected in electronic health records, pharmacy dispensing databases, health insurance systems, and national health system records. These big data repositories offer a unique chance to study adherence both retrospectively and prospectively at the population level, as well as its related factors. In order to make full use of this opportunity, there is a need to develop standardized measures of adherence, which can be applied globally to big data and will inform scientific research, clinical practice, and public health. These standardized measures may also enable a better understanding of the relationship between adherence and clinical outcomes, and allow for fair benchmarking of the effectiveness and cost-effectiveness of adherence-targeting interventions. Unfortunately, despite this obvious need, such standards are still lacking. Therefore, the aim of this paper is to call for a consensus on global standards for measuring adherence with big data. More specifically, sound standards of formatting and analyzing big data are needed in order to assess, uniformly present, and compare

patterns of medication adherence across studies. Wide use of these standards may improve adherence and make health care systems more effective and sustainable.

(J Med Internet Res 2020;22(8):e18150) doi:10.2196/18150

KEYWORDS

patient adherence; big data; metrics; consensus

Introduction

Despite half a century of dedicated studies, medication adherence remains far from perfect. In fact, nonadherence to medication (ie, the scenario in which patients are not taking their medications as prescribed) is still very prevalent. According to the World Health Organization (WHO), 50% of patients are estimated to deviate from their chronic treatments [1]. It has been shown to lead to poor health outcomes, increased use of health services, and increased costs (both direct and indirect, eg, due to absenteeism, lost productivity, etc), jeopardizing the effectiveness of evidence-based therapies [2]. In a recent meta-analysis, medication nonadherence was found to be associated with all-cause hospitalization (adjusted odds ratio 1.17, 95% CI 1.12-1.21) and mortality (good adherence was associated with a 21% reduction in long-term mortality risk in comparison with medication nonadherence; adjusted hazard ratio 0.79, 95% CI 0.63-0.98) in older people [3]. Thus, nonadherence is an important determinant of individual health. At the population level, it also seriously affects public health and the economy. Considering all these effects, nonadherence has been recognized by the WHO as a "problem of striking magnitude" [1].

Unfortunately, the seriousness of this problem is ever increasing. An important reason for this is the unprecedented demographic change that is taking place at the beginning of the 21st century. It affects the whole world and is particularly pronounced in Europe. According to Eurostat, currently, persons aged 65 years or older comprise 20% of the EU-28 population, and this proportion is expected to rise up to 31% by 2100. Even more striking are the statistics for very old citizens; the proportion of persons aged 80 years or older is expected to rise within the same time period from 6% (current) of the EU-28 population to 15% [4].

Longer lifespan results in an increase in the prevalence of noncommunicable chronic conditions and multimorbidity (usually defined as the coexistence of two or more chronic conditions in an individual). This, in turn, leads to the frequent use of complex therapeutic regimens and creates a fertile ground for nonadherence [5].

In order to prevent nonadherence, one needs to know the major drivers of this phenomenon. The WHO developed a model of the determinants affecting adherence and grouped them into five main sets of factors as follows: health system–related factors, therapy-related factors, condition-related factors, patient-related factors, and socioeconomic factors [1]. Based on this model of adherence, multiple nonadherence-targeting interventions have been designed and tested, but unfortunately, only few have been successful. As stated in a recent Cochrane systematic review, "current methods of improving medication adherence for chronic health problems are mostly complex and not very effective, so the full benefits of treatment cannot be realized" [6].

Indeed, medication taking is a complex behavior, and diverse determinants play different roles at the individual level. As a consequence, it seems to be unrealistic to expect that one uniform intervention will solve the problem of nonadherence in each and every case. On the other hand, there is a rising body of evidence that nonadherence could be effectively managed through the use of various innovative digital solutions [7]. Successful examples include web-based education and monitoring programs [8], clinical decision support systems using data from electronic health records (EHRs) to produce alerts [9], mobile technologies (mobile health [mHealth]), dedicated apps providing various combinations of patient monitoring, education, and facilitation of adherence [10,11], etc. Owing to digitization, for the first time in history, nonadherence may also become precisely measurable on a mass scale owing to the availability of large health care databases. This is particularly important for older populations, which is a group usually understudied in clinical trials for various reasons (eg, multimorbidity and related polymedication).

However, these promising opportunities are not fully utilized yet owing to a lack of basic widely accepted standards for measuring and managing adherence in big data. The discussion that started in 2019 at the forums of professional bodies active in the area of patient adherence research (ie, Action Group A1 "Adherence to prescription and medical plans" of the European Innovative Partnership on Active and Healthy Ageing and International Society for Medication Adherence ESPACOMP), to which the authors of this publication belong, led to the conclusion that this scenario needs to be changed. This idea corresponds very well with recent recommendations of the Heads of Medicines Agencies-European Medicines Agency joint Big Data Task Force, which called for the development of skills and the creation of capabilities to analyze big data [12]. Therefore, the aim of this publication is to establish a call for a consensus on global standards for measuring adherence with big data. More specifically, sound standards of formatting and analyzing big data are needed in order to assess, uniformly present, and compare patterns of medication adherence across studies, and thus, help scientific research, clinical practice, and public health.

Opportunities Created for Adherence Owing to Digitization of the Health Care Sector

Digitization is a new opportunity that has, in recent times, become more frequently adopted in the health care sector.

Interestingly, to date, digital solutions have been widely used outside health care, but they have only recently been employed in the field of medicine and offer great promise toward improved and more efficient care. In order to speed up this process, in 2018, the European Commission developed a plan for the digital transformation of health and care in the Digital Single Market. The plan is based on three pillars as follows: (1) securing data access and sharing; (2) connecting and sharing health data for research, faster diagnosis, and improved individualized health care services and health outcomes; and (3) strengthening citizen empowerment and individual care through digital services [13]. This plan is a part of the overarching European Strategy for Data [14]. Digitization of the health care sector creates an opportunity for using big data analytics tools and methods to assess nonadherence, improve clinical practice or health care services, and promote the use of tailored interventions. Routinely collected information on prescribing and dispensing, which are available in EHRs, pharmacy dispensing databases, health insurance claims systems, and national health systems records, enables a more thorough exploration of the relationship between adherence and health outcomes. The rising use of mHealth by patients for self-monitoring and disease management is also another potential data source for analysis. Thus, big data may represent a powerful and relatively low-cost resource for investigating important public health concerns in real-life scenarios, including the prevalence of nonadherence, its drivers, and the consequences of nonadherence. Big data can also be used to provide information for designing new interventions and targeting both prevention and management of nonadherence. Moreover, big data allow research on an incomparable scale, covering large populations (eg, primary nonadherence, a measure of unfilled prescriptions, was recently assessed in a cohort of 1.6 million Catalonian primary care patients [15] and in a national population based study in Poland [16]). Unlike medical trials, big data also provide an opportunity to assess adherence longitudinally (eg, an Estonian study analyzed a national database over a period of 15 years [17]). All this is possible without typical limitations in terms of cost, intrusiveness, and bias, which are characteristic of studies employing other sorts of data for adherence measurement and monitoring. However, at present, uniform and accepted standards of adherence measurement for big data are still lacking. Moreover, currently, big data collection is not uniformly formatted or structured for adherence measurement, which means that nontrivial operations are needed to allow for this kind of analysis. Therefore, to build a solid evidence base for adherence management across clinical settings, it is necessary to standardize adherence estimation and facilitate the appropriate use of these standards [18].

Adherence research is not the only area of research facing problems with standardization when it comes to digital health. For example, there are no global standards for EHRs [19]. Various historical, cultural, economic, and political reasons could be cited as causative factors, and despite activities of several interoperability initiatives, both public and private, this is still the case [20]. Several standardization development organizations have developed very mature and widely implemented standards, such as the Clinical Document Architecture [21] and Fast Healthcare Interoperability Resources [22] by HL7, and CEN-ISO 13606 [23], but they are limited in their interoperability. Interestingly, ISO 10781:2015 provides a reference list of functions that may be present in an EHR system, and of these, the following function tackles adherence assessment: Care Patient Support CPS 3.1 Function on “Support for Standard Assessments” [24].

A systemic review of the challenges for the use of big data in health care identified issues in data structure, security, data standardization, storage and transfer, and managerial skills, such as data governance, to be the most often provided in the current literature [25]. Practical challenges included data preprocessing and curation, model training, refinement of systems, ethical and legal issues, data privacy and security, end users understanding acceptance, etc [26,27].

Certainly, to overcome all these challenges is not easy. However, it is quite easy to illustrate why this scenario urgently needs to be changed.

Need for Standard Big Data–Related Adherence Metrics for Research

The introduction of the Ascertaining Barriers for Compliance taxonomy (ABC taxonomy) and new adherence terminology (named after the dedicated European research project “*Ascertaining Barriers for Adherence*”) made the first big step forward in terms of standardization by defining three essential components of adherence. These components are as follows: (1) initiation (taking the first dose of the prescribed medication); (2) implementation (taking medication as prescribed); and (3) discontinuation (stopping treatment) [28]. Following this, the recently introduced EMERGE (*ESPACOMP Medication Adherence Reporting Guideline* developed under the umbrella of the International Society for Medication Adherence, COMpliance, and Persistence [ESPACOMP]) provides guidance, along with a checklist for reporting results of studies on medication adherence [29]. Other interesting activities in this area include the work initiated with the support of the Government of Spain and the European project StandICT.eu to generate an extension of the SNOMED CT terminology for the domain of adherence [30] and the proposal to include terms of adherence in the amendment to ISO 13940 (system of concepts to support continuity of care) currently in progress in ISO TC 215 [31].

However, there are still many challenges with the use of big data for adherence assessment. Without standard metrics, the same data may lead to diverse results, as clearly depicted in the study by Malo et al, which found different mean adherence values and proportions of adherent patients when using medication possession ratio versus proportion of days covered [32].

To date, numerous studies on adherence have been undertaken, using diverse approaches to data analysis, which have led to mixed results. Most often, pharmacy records are used to measure adherence in terms of implementation and discontinuation [33]. EHR data have also been used for effective prediction of medication adherence trajectories [34], which has evoked certain discussions [35]. Menditto et al managed to integrate and

analyze six databases from three countries, which allowed for a fair comparison of medication adherence across the various countries [36]. Another study managed to assess and compare adherence to chronic medication across three European cohorts of older people by developing a common protocol and using structured documents for sharing and applying methodologies [37].

Some attempts to introduce standards to adherence assessment in big data have been made by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). The ISPOR Medication Adherence and Persistence Special Interest Group produced recommendations for assessment of initial medication adherence [38] and proposed a checklist for medication adherence studies using retrospective databases [39]. Arnet et al [40] and Raebel et al [41] proposed standard definitions and their operationalization to quantify adherence to medication from electronic databases. Lehnman et al [42] and Williams et al [43] provided some basic guidance regarding the use of pharmacy refill data to assess adherence. At the same time, a systematic review of publications on adherence in older Americans identified as many as 20 differently named measures of adherence derived from pharmacy claims data [44]. Even more interestingly, some adherence measures derived from big data are already in use for incentivizing health care providers to consider long-term health outcomes. In the United States, the Centers for Medicare & Medicaid Services adopted several quality measures using the threshold of the proportion of days covered of ≥0.8 for the drugs under measurement, for a period of 12 consecutive months [45]. In fact, there is evidence that this improves adherence [46].

Moreover, another important question that needs to be addressed in future standardized measures of adherence in big data is "what is the subject of adherence assessment: a drug, a condition, or a patient?" In other words, how to measure adherence to multiple medications prescribed for the same condition and/or to various conditions in patients with multimorbidity [47]. Indicators designed and widely used to evaluate single-medication adherence are not necessarily valid for the assessment of adherence to polypharmacy regimens [48]. For example, a study assessing adherence in individuals belonging to the Epichron cohort returned highly diverse results for various drug classes as follows: 72.4% for antidiabetics versus only 44.3% for lipid-lowering drugs [49]. A recent systematic review found serious inconsistency in the measures used to estimate adherence and persistence to multiple cardiometabolic medications [50], while another review concluded that "there appears to be no standardized method to measure multiple medication adherence" [51]. For sure, further research is needed in this respect and is particularly important given the aging population.

In summary, the major disadvantage of the current lack of widely accepted standards for adherence assessment is the difficulty in comparing and interpreting scientific studies' results. Uniform adherence measurement and a common ontology urgently need to be developed in order to support research and enable real-life implementation of study findings [40,52,53]. This is also necessary for cross-study comparisons and fair benchmarking of adherence-targeting interventions.

Need for Standard Big Data–Related Adherence Metrics for Clinical Practice, Public Health, and Health Policy

Big data and the development of a standardized measure of adherence may enable more reliable and valid investigations into the association between nonadherence and health outcomes. To date, there is no consensual standard for what constitutes adequate adherence. In practice, 80% is often used as a cutoff to classify "good adherence," but scientific evidence for this threshold is unclear. In fact, a systematic review investigated medication adherence thresholds in relation to clinical outcomes and found the included studies to be highly heterogeneous, and it could not confirm or reject the validity of the historical 80% cutoff threshold for adherence [54]. Moreover, many treatments are also preventative, and it may take a very long time to determine any therapeutic benefit at all from such treatments.

Various interventions have been designed to prevent and manage nonadherence in real-life settings. Unfortunately, despite objective need, these interventions are generally underused. Lack of standardized comparable measures of adherence is one of the major barriers to the objective selection of the most effective and cost-effective interventions [2] and the scaling-up of the best practice. Only with reliable and valid measures can nonadherence be tracked along a timeline, allowing the assessment of the long-term effects of particular interventions and the benchmarking of their effectiveness. Standard measures and guidelines to assess adherence could also facilitate the introduction as well as the assessment of the effectiveness of incentives to promote adherence at the patient, provider, and payer levels and the ability to target individual risk factors at the various health care provider levels [55].

Standardized adherence measures employed in big data sets may also provide insights into the reasons why patients do not adhere to their prescribed medication regimens. This is of utmost importance as a review of systematic reviews identified 771 individual factor items as possible determinants of nonadherence, concluding that "lack of standardized adherence definitions and use of poor measurement methods resulted in many inconsistencies in the findings and many of the identified factors had an inconsistent effect on adherence" [56].

Thus, big data sets are useful to assess adherence in different profiles of drug users, analyze the factors related to adherence, explore causes of discontinuation, and compare results across different populations. With this information, drug users at the highest risk of nonadherence can be identified, and tailored interventions can be designed and implemented. It is of paramount importance to consider that most of the current interventions to address adherence are using or are based on information technology (IT) solutions. These, however, are not currently based on standardized measures of adherence [55]. Another potential technology approach for prediction and assessment of adherence is artificial intelligence (AI). With AI, big data may be analyzed both retrospectively and in real time, allowing for more personalized health care. However, a major hindrance to the adoption of AI is, yet again, the lack of sound

operational measures of adherence to properly train the algorithms.

Along with individual health, public health is sure to benefit from the introduction of standardized measures of adherence. Such measures will enable comparison of adherence rates within and across different countries, populations, and disease groups, allowing for fair benchmarking of interventions, better planning, and practical implementation. In fact, more often, medication adherence is accepted as a measure of the quality of care provided by physicians, as well as the quality and effectiveness of the entire health care system [55,57]. Moreover, in this area, a lack of standardized measures causes the problem of nonadherence to be often overlooked in national agendas. Currently, only few countries systematically monitor adherence. Therefore, the recent Organization for Economic Co-operation and Development (OECD) report calls for standardization in order to allow for international benchmarking [55].

The recent outbreak of the COVID-19 pandemic has shown the extraordinary role that infoepidemiology (ie, information epidemiology) can play in the management of major public health problems [58]. Let this lesson be an inspiration for the wider adoption of digitization in health care in general and the faster utilization of the potential of big data for the management of adherence in particular.

Conclusions

Ongoing digitization of the health care sector and availability of big data repositories create an unprecedented opportunity to study patient adherence on a mass scale both retrospectively and in real time. Obvious benefits that could be derived for science, as well as individual and public health, are hindered by the current lack of standards for adherence-related data analysis. What sort of standards need to be agreed upon to change this scenario? First, there needs to be a standard format for the data collected in big data databases, such as EHRs and prescribing and dispensing registers, to allow for smooth and effective assessment of adherence parameters. Second, sound metrics need to be developed to process the raw data. Third, the standards of presentation of adherence measures being assessed within big data need to be agreed upon.

Bearing in mind the troublesome history of health care sector digitization, this plan may appear to be ambitious. However, the authors of this paper are motivated to face the challenge and develop these highly needed global standards, discuss them with the scientific world, and finally, agree on a common consensus. Therefore, interested individuals are invited to join our efforts within the new initiative that we want to call DIGI.PASs (Patient Adherence Standard measures to be used with big data collections available in DIGItal repositories).

Conflicts of Interest

PK received speaker honoraria from Aflofarm, Fresenius, Polpharma, and Sandoz and received grant funding from European Union's Horizon 2020 for GATEKEEPER project (grant agreement number 857223) and The European Commission ERASMUS+ Project Skills4Adherence (grant agreement number: 2017-1-PL01-KA202-038672), outside of this work. BV is the CEO and shareholder of AARDEX Group.

References

1. Sabaté E. Adherence to Long-term Therapies: Evidence for Action. Geneva: World Health Organization; 2003.
2. Cutler RL, Fernandez-Llimos F, Frommer M, Benrimoj C, Garcia-Cardenas V. Economic impact of medication non-adherence by disease groups: a systematic review. BMJ Open 2018 Jan 21;8(1):e016982 [FREE Full text] [doi: 10.1136/bmjopen-2017-016982] [Medline: 29358417]
3. Walsh CA, Cahir C, Tecklenborg S, Byrne C, Culbertson MA, Bennett KE. The association between medication non-adherence and adverse health outcomes in ageing populations: A systematic review and meta-analysis. Br J Clin Pharmacol 2019 Nov;85(11):2464-2478. [doi: 10.1111/bcp.14075] [Medline: 31486099]
4. European Comission. The EU's population projected up to 2100. Eurostat. URL: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20190710-1 [accessed 2020-04-13]
5. Wong MC, Liu J, Zhou S, Li S, Su X, Wang HH, et al. The association between multimorbidity and poor adherence with cardiovascular medications. Int J Cardiol 2014 Dec 15;177(2):477-482. [doi: 10.1016/j.ijcard.2014.09.103] [Medline: 25443249]
6. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014 Nov 20(11):CD000011 [FREE Full text] [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
7. Car J, Tan WS, Huang Z, Sloot P, Franklin BD. eHealth in the future of medications management: personalisation, monitoring and adherence. BMC Med 2017 Apr 05;15(1):73 [FREE Full text] [doi: 10.1186/s12916-017-0838-0] [Medline: 28376771]
8. Linn AJ, Vervloet M, van Dijk L, Smit EG, Van Weert JC. Effects of eHealth interventions on medication adherence: a systematic review of the literature. J Med Internet Res 2011 Dec 05;13(4):e103 [FREE Full text] [doi: 10.2196/jmir.1738] [Medline: 22138112]
9. Bailey SC, Wallia A, Wright S, Wismer GA, Infanzon AC, Curtis LM, et al. Electronic Health Record-Based Strategy to Promote Medication Adherence Among Patients With Diabetes: Longitudinal Observational Study. J Med Internet Res 2019 Oct 21;21(10):e13499 [FREE Full text] [doi: 10.2196/13499] [Medline: 31638592]

10. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
11. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ. Mobile Apps for Increasing Treatment Adherence: Systematic Review. J Med Internet Res 2019 Jun 18;21(6):e12505 [FREE Full text] [doi: 10.2196/12505] [Medline: 31215517]
12. Priority Recommendations of the HMA-EMA joint Big Data Task Force. European Medicines Agency. URL: https://www.ema.europa.eu/en/documents/other/priority-recommendations-hma-ema-joint-big-data-task-force_en.pdf [accessed 2020-04-13]
13. eHealth: Digital health and care. European Commission. URL: https://ec.europa.eu/health/ehealth/overview_en [accessed 2020-04-13]
14. European data strategy: Making the EU a role model for a society empowered by data. European Commission. URL: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy [accessed 2020-04-13]
15. Aznar-Lou I, Fernández A, Gil-Girbau M, Fajó-Pascual M, Moreno-Peral P, Peñarrubia-María MT, et al. Initial medication non-adherence: prevalence and predictive factors in a cohort of 1.6 million primary care patients. Br J Clin Pharmacol 2017 Jun;83(6):1328-1340 [FREE Full text] [doi: 10.1111/bcp.13215] [Medline: 28229476]
16. Kardas P, Cieszyński J, Czech M, Banaś I, Lewek P. Primary nonadherence to medication and its drivers in Poland: findings from the electronic prescription pilot analysis. Pol Arch Intern Med 2020 Jan 31;130(1):8-16 [FREE Full text] [doi: 10.20452/pamw.14994] [Medline: 31559971]
17. Laius O, Pisarev H, Maasalu K, Kõks S, Märtson A. Adherence to osteoporosis medicines in Estonia-a comprehensive 15-year retrospective prescriptions database study. Arch Osteoporos 2017 Dec;12(1):59. [doi: 10.1007/s11657-017-0354-z] [Medline: 28643266]
18. Dima A, Allemann S, Dediu D. AdhereR: An Open Science Approach to Estimating Adherence to Medications Using Electronic Healthcare Databases. Stud Health Technol Inform 2019 Aug 21;264:1451-1452. [doi: 10.3233/SHTI190479] [Medline: 31438176]
19. OECD. Health in the 21st Century: Putting Data to Work for Stronger Health System. In: OECD Health Policy Studies. Paris: OECD Publishing; 2019.
20. Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, et al. Making sense of big data in health research: Towards an EU action plan. Genome Med 2016 Jun 23;8(1):71 [FREE Full text] [doi: 10.1186/s13073-016-0323-y] [Medline: 27338147]
21. Master Grid of Standards. Health Level Seven International. URL: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=7V [accessed 2020-04-13]
22. Welcome to FHIR. HL7. URL: https://www.hl7.org/fhir/ [accessed 2020-04-13]
23. ISO 13606-1:2019 Health informatics — Electronic health record communication — Part 1: Reference model. ISO. URL: https://www.iso.org/standard/67868.html [accessed 2020-04-13]
24. ISO/HL7 10781:2015 Health Informatics — HL7 Electronic Health Records-System Functional Model, Release 2 (EHR FM). ISO. URL: https://www.iso.org/standard/57757.html [accessed 2020-04-13]
25. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and Opportunities of Big Data in Health Care: A Systematic Review. JMIR Med Inform 2016 Nov 21;4(4):e38 [FREE Full text] [doi: 10.2196/medinform.5359] [Medline: 27872036]
26. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. The Lancet Oncology 2019 May;20(5):e262-e273. [doi: 10.1016/S1470-2045(19)30149-4]
27. Shortreed SM, Cook AJ, Coley RY, Bobb JF, Nelson JC. Challenges and Opportunities for Using Big Health Care Data to Advance Medical Science and Public Health. Am J Epidemiol 2019 May 01;188(5):851-861. [doi: 10.1093/aje/kwy292] [Medline: 30877288]
28. Vrijens B, De Geest S, Hughes DA, Przemyslaw K, Demonceau J, Ruppar T, Project Team ABC. A new taxonomy for describing and defining adherence to medications. Br J Clin Pharmacol 2012 May;73(5):691-705 [FREE Full text] [doi: 10.1111/j.1365-2125.2012.04167.x] [Medline: 22486599]
29. De Geest S, Zullig LL, Dunbar-Jacob J, Helmy R, Hughes DA, Wilson IB, et al. ESPACOMP Medication Adherence Reporting Guideline (EMERGE). Ann Intern Med 2018 Jul 03;169(1):30-35. [doi: 10.7326/M18-0543] [Medline: 29946690]
30. Final Impact Report: Europe's Contributions To ICT Standardisation Globally - Impacts From StandICT.Eu Funding. StandICT.eu. URL: https://www.standict.eu/publications/final-impact-report-europe%E2%80%99s-contributions-ict-standardisation-globally-impacts [accessed 2020-04-13]
31. Adherence and Integrated Care. BioPortal. URL: https://bioportal.bioontology.org/ontologies/ADHER_INTCARE_EN [accessed 2020-04-13]
32. Malo S, Aguilar-Palacio I, Feja C, Lallana MJ, Rabanaque MJ, Armesto J, et al. Different approaches to the assessment of adherence and persistence with cardiovascular-disease preventive medications. Curr Med Res Opin 2017 Jul;33(7):1329-1336. [doi: 10.1080/03007995.2017.1321534] [Medline: 28422521]
33. Lima-Dellamora ED, Osorio-de-Castro CG, Madruga LG, Azeredo TB. Use of pharmacy records to measure treatment adherence: a critical review of the literature. Cad Saude Publica 2017 Apr 20;33(3):e00136216 [FREE Full text] [doi: 10.1590/0102-311X00136216] [Medline: 28444026]

34. Franklin JM, Gopalakrishnan C, Krumme AA, Singh K, Rogers JR, Kimura J, et al. The relative benefits of claims and electronic health record data for predicting medication adherence trajectory. Am Heart J 2018 Mar;197:153-162. [doi: 10.1016/j.ahj.2017.09.019] [Medline: 29447776]
35. Zullig LL, Curtis LH. A population health perspective on a claims and electronic health record-based tool to screen for suboptimal medication adherence. Am Heart J 2018 Mar;197:150-152. [doi: 10.1016/j.ahj.2017.10.012] [Medline: 29447775]
36. Menditto E, Bolufer De Gea A, Cahir C, Marengoni A, Riegler S, Fico G, et al. Scaling up health knowledge at European level requires sharing integrated data: an approach for collection of database specification. Clinicoecon Outcomes Res 2016;8:253-265 [FREE Full text] [doi: 10.2147/CEOR.S97548] [Medline: 27358570]
37. Menditto E, Cahir C, Aza-Pascual-Salcedo M, Bruzzese D, Poblador-Plou B, Malo S, et al. Adherence to chronic medication in older populations: application of a common protocol among three European cohorts. Patient Prefer Adherence 2018;12:1975-1987 [FREE Full text] [doi: 10.2147/PPA.S164819] [Medline: 30323567]
38. Hutchins DS, Zeber JE, Roberts CS, Williams AF, Manias E, Peterson AM, IPSOR Medication AdherencePersistence Special Interest Group. Initial Medication Adherence-Review and Recommendations for Good Practices in Outcomes Research: An ISPOR Medication Adherence and Persistence Special Interest Group Report. Value Health 2015 Jul;18(5):690-699 [FREE Full text] [doi: 10.1016/j.jval.2015.02.015] [Medline: 26297098]
39. Peterson AM, Nau DP, Cramer JA, Benner J, Gwadry-Sridhar F, Nichol M. A checklist for medication compliance and persistence studies using retrospective databases. Value Health 2007;10(1):3-12 [FREE Full text] [doi: 10.1111/j.1524-4733.2006.00139.x] [Medline: 17261111]
40. Arnet I, Kooij MJ, Messerli M, Hersberger KE, Heerdink ER, Bouvy M. Proposal of Standardization to Assess Adherence With Medication Records: Methodology Matters. Ann Pharmacother 2016 May;50(5):360-368. [doi: 10.1177/1060028016634106] [Medline: 26917817]
41. Raebel MA, Schmittdiel J, Karter AJ, Konieczny JL, Steiner JF. Standardizing terminology and definitions of medication adherence and persistence in research employing electronic databases. Med Care 2013 Aug;51(8 Suppl 3):S11-S21 [FREE Full text] [doi: 10.1097/MLR.0b013e31829b1d2a] [Medline: 23774515]
42. Lehmann A, Aslani P, Ahmed R, Celio J, Gauchet A, Bedouch P, et al. Assessing medication adherence: options to consider. Int J Clin Pharm 2014 Feb;36(1):55-69. [doi: 10.1007/s11096-013-9865-x] [Medline: 24166659]
43. Williams AB, Amico KR, Bova C, Womack JA. A proposal for quality standards for measuring medication adherence in research. AIDS Behav 2013 Jan;17(1):284-297 [FREE Full text] [doi: 10.1007/s10461-012-0172-7] [Medline: 22407465]
44. Sattler EL, Lee JS, Perri M. Medication (re)fill adherence measures derived from pharmacy claims data in older Americans: a review of the literature. Drugs Aging 2013 Jun;30(6):383-399. [doi: 10.1007/s40266-013-0074-z] [Medline: 23553512]
45. Quality Measures. U.S. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/index?redirect=/qualitymeasures/ [accessed 2020-04-13]
46. Parekh N, Munshi KD, Hernandez I, Gellad WF, Henderson R, Shrank WH. Impact of Star Rating Medication Adherence Measures on Adherence for Targeted and Nontargeted Medications. Value Health 2019 Nov;22(11):1266-1274. [doi: 10.1016/j.jval.2019.06.009] [Medline: 31708063]
47. Kim S, Bennett K, Wallace E, Fahey T, Cahir C. Measuring medication adherence in older community-dwelling patients with multimorbidity. Eur J Clin Pharmacol 2018 Mar;74(3):357-364. [doi: 10.1007/s00228-017-2388-y] [Medline: 29199370]
48. Arnet I, Greenland M, Knuiman MW, Rankin JM, Hung J, Nedkoff L, et al. Operationalization and validation of a novel method to calculate adherence to polypharmacy with refill data from the Australian pharmaceutical benefits scheme (PBS) database. Clin Epidemiol 2018;10:1181-1194 [FREE Full text] [doi: 10.2147/CLEP.S153496] [Medline: 30233252]
49. Moreno Juste A, Gimeno Miguel A, Poblador Plou B, González Rubio F, Aza Pascual-Salcedo MM, Menditto E, et al. Adherence to treatment of hypertension, hypercholesterolaemia and diabetes in an elderly population of a Spanish cohort. Med Clin (Barc) 2019 Jul 05;153(1):1-5. [doi: 10.1016/j.medcli.2018.10.023] [Medline: 30503066]
50. Alfian SD, Pradipta IS, Hak E, Denig P. A systematic review finds inconsistency in the measures used to estimate adherence and persistence to multiple cardiometabolic medications. J Clin Epidemiol 2019 Apr;108:44-53 [FREE Full text] [doi: 10.1016/j.jclinepi.2018.12.003] [Medline: 30537541]
51. Pednekar PP, Ágh T, Malmenäs M, Raval AD, Bennett BM, Borah BJ, et al. Methods for Measuring Multiple Medication Adherence: A Systematic Review-Report of the ISPOR Medication Adherence and Persistence Special Interest Group. Value Health 2019 Feb;22(2):139-156 [FREE Full text] [doi: 10.1016/j.jval.2018.08.006] [Medline: 30711058]
52. Dima AL, Dediu D. Computation of adherence to medication and visualization of medication histories in R with AdhereR: Towards transparent and reproducible use of electronic healthcare data. PLoS One 2017;12(4):e0174426 [FREE Full text] [doi: 10.1371/journal.pone.0174426] [Medline: 28445530]
53. Román-Villarán E, Pérez-Leon FP, Escobar-Rodriguez GA, Parra-Calderón CL. EIP on AHA Ontology for adherence: Knowledge representation advanced tools. Transl Med UniSa 2019;19:49-53 [FREE Full text] [Medline: 31360667]
54. Baumgartner PC, Haynes RB, Hersberger KE, Arnet I. A Systematic Review of Medication Adherence Thresholds Dependent of Clinical Outcomes. Front Pharmacol 2018;9:1290 [FREE Full text] [doi: 10.3389/fphar.2018.01290] [Medline: 30524276]
55. Khan R, Socha-Dietrich K. Investing in medication adherence improves health outcomes and health system efficiency: Adherence to medicines for diabetes, hypertension, and hyperlipidaemia. In: OECD Health Working Papers No. 105. Paris: OECD Publishing; 2018:1-40.

56. Kardas P, Lewek P, Matyjaszczyk M. Determinants of patient adherence: a review of systematic reviews. Front Pharmacol 2013;4:91 [FREE Full text] [doi: 10.3389/fphar.2013.00091] [Medline: 23898295]
57. Seabury SA, Dougherty JS, Sullivan J. Medication adherence as a measure of the quality of care provided by physicians. Am J Manag Care 2019 Feb;25(2):78-83 [FREE Full text] [Medline: 30763038]
58. Mavragani A. Tracking COVID-19 in Europe: Infodemiology Approach. JMIR Public Health Surveill 2020 Apr 20;6(2):e18941 [FREE Full text] [doi: 10.2196/18941] [Medline: 32250957]

Abbreviations

AI: artificial intelligence
EHR: electronic health record
EMERGE: ESPACOMP Medication Adherence Reporting Guideline
ESPACOMP: European Society for Patient Adherence, COMpliance, and Persistence
ISO: International Organization for Standardization
ISPOR: International Society for Pharmacoeconomics and Outcomes Research
IT: information technology
WHO: World Health Organization

Edited by G Eysenbach; submitted 08.02.20; peer-reviewed by TT Meum, PM Matricardi; comments to author 18.03.20; revised version received 26.04.20; accepted 22.06.20; published 27.08.20.

Please cite as:
Kardas P, Aguilar-Palacio I, Almada M, Cahir C, Costa E, Giardini A, Malo S, Massot Mesquida M, Menditto E, Midão L, Parra-Calderón CL, Pepiol Salom E, Vrijens B
The Need to Develop Standard Measures of Patient Adherence for Big Data: Viewpoint
J Med Internet Res 2020;22(8):e18150
URL: http://www.jmir.org/2020/8/e18150/
doi:10.2196/18150
PMID:32663138

Viewpoint

The Art of Surgery: Balancing Compassionate With Virtual Care

Elisheva Tamar Anne Nemetz[1], BA, MBE; David Robert Urbach[2,3,4,5], MSc, MD, FRCSC, FACS; Karen Michelle Devon[2,4,5,6], MDCM, MSc, FRCSC, FACS

[1]Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[2]Department of Surgery, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[3]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[4]University Health Network, Toronto, ON, Canada
[5]Women's College Hospital, Toronto, ON, Canada
[6]Joint Centre for Bioethics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Karen Michelle Devon, MDCM, MSc, FRCSC, FACS
Women's College Hospital
76 Grenville Street
Toronto, ON, M5S 1B2
Canada
Phone: 1 (416) 323 6400 ext 7352
Email: karen.devon@wchospital.ca

Abstract

The recent drive to include virtual care in surgical practice has been accelerated due to the COVID-19 pandemic. Many physicians feel that communicating via telehealth is unlike traditional methods of providing health care, and thus guidance on maintaining excellence in communication is necessary, especially as academic literature on virtual care in surgery is nonexistent. Challenges faced in transitioning to virtual care include the inability to utilize body language, barriers to traditional physical examination, exacerbation of existing vulnerabilities and inequities in patient groups, the declining quality of medical education, and the fragmentation of the multidisciplinary health care team. This paper seeks to resolve these challenges by focusing on the pillars of good communication, including preparation, professionalism, empathy, respect, and the virtual physical examination.

(J Med Internet Res 2020;22(8):e22417) doi:10.2196/22417

KEYWORDS

bioethics; medical ethics; virtual care; telehealth; virtual care in surgery; video care in surgery; telehealth in surgery; surgical communication; COVID-19 and virtual care; consent; privacy; medical education; surgery

Introduction

There has been a desire in the past few years to include virtual care in surgical practice to enhance communication and improve patient access to care. With the emergence of the COVID-19 pandemic, this transition has been accelerated and brings with it several challenges, including significantly different methods of communication and connection. Effective communication has always been a pillar of surgical practice and the surgeon-patient relationship. Excellent communication fosters a relationship based on trust and, therefore, superior medical care and outcomes [1,2]. Evidence suggests that surgeons who communicate with their patients more empathically and are active listeners also have better diagnostic acumen [3,4]. Better communicators also face fewer medical malpractice suits and, more importantly, have more satisfied patients [4]. The literature on virtual surgical care is scarce, and guidance in the form of academic articles, manuals, and lectures on effective virtual medical communication and compassionate virtual care is missing from the surgical discourse. Many physicians feel that communicating via telehealth is too unlike traditional methods of providing health care, and thus guidance on maintaining excellence in communication is necessary.

Challenges in Telephone and Video Communication

Virtual communication inherently brings challenges. First, medical care provided over the phone does not permit the use of body language as a form of non-verbal communication. Additionally, a critical part of every medical encounter and physical examination is evaluating the patient's general appearance, which begins the moment the patient is greeted and is not feasible over the phone.

Video-based medical care, while potentially mitigating some of these problems, also has limitations. Although video communication allows for non-verbal communication, the use of touch is nonetheless limited. A hand on a shoulder when performing an examination or simply being seated and present with a patient through a difficult diagnosis cannot be replicated in a two-dimensional medium. Body language speaks volumes, and surgeons rely on body language as an essential form of communication, even aside from the necessity of a physical exam.

Virtual care can also exacerbate existing vulnerabilities and inequities in patient groups. For example, those that have difficulty with hearing, who rely on sign language, or for whom English is a second language, face even greater barriers. While interpretation services can be coordinated, it becomes even more difficult to ensure compassionate care when words are lost in translation without a physical presence. Moreover, access to a phone or computer is a privilege often taken for granted. Telephone and video-based care assume that each patient has access to technology, thereby perpetuating a socioeconomic divide. While more individuals have internet access at home, most families own only one computer [5]. It is not uncommon for a family to rely on a single computer for children's schoolwork, family entertainment, and parental work, making confidential and safe medical care via video difficult. High-speed internet is also difficult to access for many rural Americans, due to fewer providers, less wiring for broadband services, and limited options [6]. Slow internet can interrupt the provision of virtual care and necessitate reliance on phone consultation or booking a new appointment when technology fails. Monthly billing for cellular access may be too expensive for families, impairing access to mobile phones [7]. Knowledge of how to utilize technology is also often taken for granted. Many patients, particularly the elderly, have difficulty learning how to access and use video communication. Without appropriate education and assistive social support, they become disadvantaged through reliance on phone consultations, especially when video communication might be essential, such as with postoperative wound care or gait assessments.

Amid this new reality, the quality of medical education is suffering. Many surgeons are engaged in teaching fellows, residents, and medical students and find that incorporating the education of students into the existing demands and limitations of virtual care is especially challenging. Fellows, who are honing their subspecialty and independent practice skills, may lack the opportunity to see patients before the faculty surgeon. Residents may experience delays in transitioning from medical student to practicing physician, and many medical schools have removed students from hospitals, as they are considered non-essential to patient care.

Finally, surgeons often provide care in a team-based setting requiring communication not only between patient and surgeon but also with other members of the multidisciplinary health care team such as internal medicine physicians, social workers, and nurses. The involvement of team members in a collaborative and coordinated manner is essential for good medical and surgical care. Determinants of successful collaborative care include physical space, temporal arrangements, schedules, processes, and communication tools. Conversely, spatial distance, asynchronous schedules, and dependence on virtual communication between colleagues reduce the ability to engage in collaborative care [8]. Interprofessional education focusing on training and education that promotes collaboration becomes significantly reduced when moved to an online format. Configuring a virtual multidisciplinary health care team requires mitigation of distance limitations, enhancing authentic communication, integrating schedules, and modifying interprofessional education opportunities.

The challenges posed by virtual care should be weighed against the shortcomings of in-person appointments. Patients face monetary costs associated with transportation and hospital parking fees. Additional obstacles include taking time off from work, hiring a babysitter for children at home, and long wait times. Furthermore, postoperative appointments may pose difficulties for patients who become too immobilized following surgery to travel to in-person appointments. This problem is exacerbated for those who reside in rural communities, far from their surgeon's practice.

Textbox 1. Consultation request.

Case: Patient 1 is a 75-year-old woman referred to a thyroid surgeon by her family physician due to a large goiter, which is suspicious for thyroid cancer. She lives alone and speaks Arabic fluently. She is a retired pharmacist who trained in Morocco and immigrated to Canada in 1975. Her children and grandchildren live nearby, and she spends a lot of her time enjoying their company, cooking her grandchildren's favorite foods, and reading.

Principles of Good Communication

Providing compassionate virtual care includes the following principles: (1) preparation, (2) professionalism, (3) empathy, (4) respect, and (5) a virtual physical examination (refer to Textbox 1 for clinical case).

Preparation

Before booking patients for virtual medical care, it is necessary to ensure the modality is appropriate for the patient. If a patient requires a diagnostic or therapeutic intervention or physical examination, virtual medical care should not be offered, and the surgeon should work with the patient to provide that care in compliance with current hospital practices. After ensuring a patient is appropriate for virtual care, proper education of the modality of care provided should be shared with easy-to-understand instructions. For example, if the surgeon uses phone-based care from a blocked number, ensure the patient understands that they should expect a call from "No Caller ID" or "Blocked Number." Alternatively, for example, if a surgeon utilizes an online video-conferencing system for medical care, ensure the patient has access to the meeting ID and password and recommend the patient tests computer audio and video before the appointment. Encourage the patient to access instructional videos or guides and to log on at least five minutes prior in case issues arise and alert the patient that they may

receive a call within a specific time range. Patients should be provided the option to obtain technological education regarding the virtual care system the hospital utilizes. Doing so permits the patient to receive video-based medical care when they otherwise would be unable to do so. Technology-readiness can take many forms from online resources to pretesting with a mock appointment. Surgeons should ensure that patients are aware that trainees may be involved in their phone or video-based care.

The standards applicable to the physician-patient relationship established in-person must be applied to virtual care. Consent for email communication needs to be obtained, and the patient informed that personal health information will be protected. This information will not be collected, disclosed, or utilized more than reasonably necessary and will be obtained securely and privately [9]. Care provided through virtual means does not replace the need for physical examination or an in-person visit for certain disorders or urgent problems. The patient should acknowledge the need to seek urgent care in an emergency department as necessary [10]. The surgeon makes the professional calculus that the risks of virtual care do not outweigh the potential benefits and is in the patient's best interest (refer to Textbox 2 for preparation case resolution) [9].

Textbox 2. Coordinating the virtual consultation.

After receiving the referral, the surgeon's administrative assistant contacts Patient 1 to arrange an appointment. Because she lacks internet access, she is scheduled for a phone appointment and is given instructions to expect a call with the hospital's name and is provided a half-hour time slot. Consent for communication is obtained, and it is explained to her that her medical information will be safeguarded, and her privacy is of utmost importance. She acknowledges the need to seek urgent care in the emergency department as necessary and that virtual care may not replace the need for physical examination or an in-person visit. She is informed that because it is a teaching hospital, medical trainees may be present on the call. Patient 1 understands and is asked if she would benefit from a hospital translator that speaks Arabic. She consents to trainee involvement and requests a translator, whose assistance is arranged.

Professionalism

Expectations should be set before clinical encounters. Surgeons should choose a private, quiet, and uncluttered location. With phone and video care, visits are often shorter and more focused than in person [7,11]. Therefore, it is important to ensure that patients and their families can share their thoughts and feelings freely. Allow for patient autonomy within the clinical encounter by beginning with open-ended questions, such as "How can I help you?" or "Tell me what you would like to discuss today?" [4] This provides the patient with the opportunity to guide the beginning of the appointment by focusing on what they view as most important and promotes patient-centered care. Physicians should consider optimizing the hospital's electronic medical record system to provide a summary of the clinical plan, a wrap-up note, or access to lab results and imaging, a compassionate approach common to in-person clinical care.

When engaging in sensitive or emotional conversations over video or phone, surgeons should utilize appropriate points in the conversation and acknowledge the patient's verbal or non-verbal cues. In place of body language, practitioners are required to rely on their tone of voice to display empathy, warmth, reassurance, and presence. It is also beneficial to use summarizing and signposting, a statement that clarifies for the patient what to expect in the next part of the conversation. Signposting signals a transition from one phase of a conversation to another, especially before a shift in the sensitivity or seriousness of a topic. For example, after obtaining the patient's past medical history and before assessing the disease's impact on the patient's activities of daily living, acknowledgment of previous hospitalizations or illnesses with empathy is paramount. The simultaneous utilization of a transition statement is essential in signaling a progression to a more sensitive discussion.

Many institutions have recognized the limitations of phone and video medical care for patients who have difficulty hearing, rely on sign language, or for whom English is a second language. Implementing virtual interpretation services is an essential component in the new virtual delivery of care. These services can include on-demand phone and video interpretation, prescheduled interpretation booking services via phone or online booking portals, and sign language interpretation. The transition of in-hospital interpretation services to an online or phone modality is crucial for accountability, confidentiality, and equitable patient care.

The methods surgeons choose for clinical encounters should ensure patient privacy. The physical space from which the surgeon is conducting the visit should be private with minimal potential for disruption. Computer calling that utilizes Facebook, FaceTime, or unencrypted methods is discouraged [9]. It is better to utilize technology such as Direct Inward System Access (DISA), which allows staff and physicians to make outbound calls that appear to come from the clinic's phone number. Another option is to use personal phones with the number blocked by altering the phone's caller ID settings [12].

The surgeon should prepare to respond appropriately if technology fails or is interrupted. The surgeon should calmly instruct the patient to switch from video to a phone-based modality. Multiple means of communication should be made available before beginning the appointment in case one is interrupted. Surgeons should also be educated on their hospital's technology services and should seek opportunities to learn to engage the technology in their practice.

Amid the time-sensitive transition to virtual care during the beginning of the COVID-19 pandemic, providers were permitted to utilize the services of virtual communication companies that were previously unauthorized [13]. HIPAA (Health Insurance Portability and Accountability Act) waivers were provided in order to permit the use of these virtual communication technology companies [13]. However, hospitals and providers are continuously advised to weigh the benefits of virtual care with patient privacy. Providers must ensure that privacy regulations and principles are placed as top priorities during

virtual care (refer to Textbox 3 for professionalism case resolution).

Textbox 3. Initiating the virtual encounter.

Before calling Patient 1, the surgeon chooses to call from her office, which is quiet and private. The surgical fellow, resident, and hospital translator are added to the call before the patient joins, and the surgeon shares the goals of the appointment. Following this collaboration, the patient is then added in order to mitigate any technical difficulties. The surgeon introduces the surgical fellow, resident, and translator present at Patient 1's appointment. The surgeon outlines the appointment to her and shares that the resident may ask some questions part way through the clinical encounter, to which she agrees. The surgeon begins with an open-ended question to gain an understanding of her perception of her illness, "Can you tell me what you know about why you're here today?" Although the surgeon has a plan and is guiding the appointment, she lets Patient 1 complete statements without interruption. When moving from one section to another, the surgeon engages in signposting. For example, when the surgeon elicits the past medical history and wants a greater understanding of how the disease is affecting her quality of life, she says, "I appreciate you sharing your recent hospitalization and how that was difficult for you. I'm now going to ask you a couple of questions regarding how this may be impacting your everyday life and the activities you enjoy doing." The surgeon is, therefore, able to engage in a conversation where she obtains the information she needs while simultaneously allowing the patient to understand and respond appropriately to the flow of the conversation.

Empathy

Patients have feelings and fears surrounding medical decisions and illness. Compassionate patient-centered care requires the provider to assess what the patient's feelings or fears are, what their ideas are of the illness or medical decisions they face, what the effect of the medical decisions or illness is on their life, and what the patient's expectations are for the clinical encounter [14].

Screen sharing tools should be utilized to permit the patient to obtain a visual representation of what to expect. This tool can help demonstrate scar size with surgery, quell patients' concerns regarding the cosmetic results of surgery, and to provide or a visual depiction of the procedure. The surgeon should gauge the patient's comfort level through questions such as, "How much information would you like to know?" By this approach, the provider can avoid sharing imagery or excessive procedural details concerning the surgery when the patient is uncomfortable with that information. Acknowledge the patient's flexibility for engaging in virtual care and consider sharing appreciation for the time they put aside for the visit (refer to Textbox 4 for empathy case resolution).

Textbox 4. Empathic communication.

The surgeon relies on her tone to convey empathy and warmth. The surgeon assesses the patient's idea of illness, her feelings towards it, and expectations of the clinical encounter. When the patient hesitates from the suggestion of a biopsy as part of the medical plan moving forward, the surgeon gently says, "I can see you may have concerns about the biopsy, can you tell me a little more about how you're feeling?" Patient 1 shares that her brother had died from thyroid cancer. Through pausing and listening and being empathetic and nonjudgmental towards the patient's emotional reaction, the surgeon conveys that they are a team and that she will do everything she can to ensure the patient's comfort in this process. The surgeon also suggests having a family member present at future meetings if Patient 1 believes it would be beneficial.

Respect

With phone and video clinical encounters, there is limited opportunity for body language and movements. When engaged in video care, use eye contact to demonstrate active listening [15]. Looking off-camera, or being distracted by a notebook, computer, or phone can indicate a lack of attentiveness. Nonverbal communication can convey empathy, support, and reassurance [2]. Even slight nodding or modifying facial expressions is enough to demonstrate active listening.

In the course of a phone or video session, allow the patient to speak and complete statements without interruption. Leave space for the patient to think and pause before answering any questions. Refrain from interrupting and redirecting the patient too quickly. Additionally, active listening involves genuinely concentrating on what the patient is saying, instead of passively "hearing" [14].

Health-related beliefs, values, and behaviors are shaped by race, ethnicity, socioeconomic status, sexual orientation, and physical and mental ability [16]. Surgeons need to be culturally competent to integrate these factors into the delivery of health care. Just as surgery progresses to virtual care, the component of cultural competency needs to also. This adjustment can include incorporating traditional healers, thereby recognizing the importance of various cultures in the medical experience to expanding hours of operation in order to make clinical settings more accessible to patients (refer to Textbox 5 for respect case resolution) [16].

Textbox 5. Demonstrating respect.

The surgeon focuses on being present and not rushing through any step or question, as she infuses this distant modality with humanity. She allows for natural pauses and is not too quick to redirect the patient. When reviewing imaging, the surgeon states, "I am going to look at your imaging right now so there will be some silence" to demonstrate attentiveness and keep the patient informed. After the meeting concludes, the surgeon notes on the patient's chart that she requires interpretation services, which would be set up easily and automatically moving forward. She notes that the patient has a preference for her daughter to be present at future meetings and includes the contact information of the daughter provided. Additionally, as some aspects of treatment will require visits to the hospital, social work will be contacted to arrange resources that would benefit the patient and ensure she receives patient-centered care.

Virtual Physical Examination

The traditional physical examination is difficult to conduct through virtual care, and surgeons require creativity, in collaboration with the patient, to complete a physical examination. This creativity can take various forms, for example utilizing at-home blood pressure cuffs or asking patients to assess pulse strength, symmetry, regularity, and quality through guidance. Some examinations can technically be implemented over video but would be socially irresponsible to carry out. For example, assessing a postoperative inguinal hernia wound over video might expose the patient and compromise patient dignity. The medical profession is careful to put into practice rituals that demonstrate respect for patients, and many of these practices are impossible over video. These include leaving the room while the patient gowns and covering a patient's side while examining another, amongst others. When the surgeon is unsure about the appropriateness of the physical examination being completed virtually due to the efficacy of a virtual physical examination or out of respect for the patient, arrangements should be made for the patient to obtain in-person care (refer to Textbox 6 for virtual physical exam case resolution).

Textbox 6. Physical examination in the virtual encounter.

The surgeon shares that she would like to get a sense of Patient 1's health through an examination. The surgeon communicates that the resident will ask her some questions in order to obtain more information about the changes she is experiencing and that some questions may require movement by the patient. She understands, and the resident guides her through checking her height, weight, temperature, blood pressure, and pulse rate. The resident also acknowledges the patient's thyroid enlargement by asking if she or her family has noticed any change in her thyroid, and she responds that her daughter has commented on her neck getting bigger. In this particular case, the resident asks if she can move her neck and if she can touch her chin to her chest in order to determine if she exhibits a full range of motion. The virtual examination of Patient 1 elicited all the information required at this stage.

Conclusion

The literature on virtual surgical care is sparse, and guidance in the form of academic articles, manuals, and lectures on compassionate virtual care is absent from surgical discourse thus far. Through focusing on pillars of good communication, including preparation, professionalism, empathy, respect, and the virtual physical examination, this paper sought to address these challenges. Communication is the foundation of compassionate virtual care, thus more research is required to ensure that the modality is mastered in surgical care.

Conflicts of Interest

None declared.

References

1. Weng H, Steed JF, Yu S, Liu Y, Hsu C, Yu T, et al. The effect of surgeon empathy and emotional intelligence on patient satisfaction. Adv Health Sci Educ Theory Pract 2011 Dec;16(5):591-600. [doi: 10.1007/s10459-011-9278-3] [Medline: 21287265]
2. Berman AC, Chutka DS. Assessing effective physician-patient communication skills:. Korean J Med Educ 2016 Jun;28(2):243-249 [FREE Full text] [doi: 10.3946/kjme.2016.21] [Medline: 26913771]
3. Ranjan P, Kumari A, Chakrawarty A. How can Doctors Improve their Communication Skills? J Clin Diagn Res 2015 Mar;9(3):JE01-JE04 [FREE Full text] [doi: 10.7860/JCDR/2015/12072.5712] [Medline: 25954636]
4. Kuehl SP. Communication Tools for the Modern Doctor Bag. Physician Patient Communication Part 1: Beginning of a medical interview. J Community Hosp Intern Med Perspect 2011;1(3):8428 [FREE Full text] [doi: 10.3402/jchimp.v1i3.8428] [Medline: 23882333]
5. Pew Research Center 2016 Nov. The Typical American Household Contains Multiple Connected Devices. URL: https://www.pewresearch.org/fact-tank/2017/05/25/a-third-of-americans-live-in-a-household-with-three-or-more-smartphones/ft_17-05-03_devices_share/ [accessed 2020-07-31]
6. Anderson M. About A Quarter of Rural Americans Say Access to High Speed Internet is a Major Problem. Pew Research Center Sep. 2018. URL: https://www.pewresearch.org/fact-tank/2018/09/10/about-a-quarter-of-rural-americans-say-access-to-high-speed-internet-is-a-major-problem/ [accessed 2020-07-31]
7. Calton B, Abedini N, Fratkin M. Telemedicine in the Time of Coronavirus. J Pain Symptom Manage 2020 Mar 31:e12. [doi: 10.1016/j.jpainsymman.2020.03.019] [Medline: 32240756]
8. Morley L, Cashell A. Collaboration in Health Care. J Med Imaging Radiat Sci 2017 Jun;48(2):207-216. [doi: 10.1016/j.jmir.2017.02.071] [Medline: 31047370]
9. College OPOO. Protecting Personal Health Information. Policies 11-15. 2000. URL: https://www.cpso.on.ca/Physicians/Policies-Guidance/Policies/Protecting-Personal-Health-Information [accessed 2020-07-10]
10. Virtual Care Telephone and Video Codes Frequently Asked Questions 2020 Mar. Ontario Medical Association (OMA). URL: https://www.oha.com/Bulletins/OMA%20FAQ%20re%20virtual%20care%20April%202020%20see%20q%2034-38%20%281%29.pdf [accessed 2020-07-10]

11. Levinson W, Chaumeton N. Communication between surgeons and patients in routine office visits. Surgery 1999 Feb;125(2):127-134. [Medline: 10026744]
12. Rosh J, Mehta N. Magenta Health Case Study - Google Docs 2020 Apr. URL: https://docs.google.com/document/d/1yWdlhhq44QeRk1Wl7yGF9pjE-hy0aO4QzLQksnMCzOU/edit
13. U.S. Department of Health & Human Services 2020 Mar. Notification of Enforcement Discretion for Telehealth Remote Communications During the COVID-19 Nationwide Public Health Emergency. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/emergency-preparedness/notification-enforcement-discretion-telehealth/index.html [accessed 2020-07-04]
14. Grek A, Hudson J. Communication Skills 1 Initiate the Encounter. In: Tzanetos K, Wong D, editors. University of Toronto Faculty of Medicine. PDF. Toronto: University of Toronto Faculty of Medicine; Sep 06, 2019.
15. Calton BA, Rabow MW, Branagan L, Dionne-Odom JN, Parker Oliver D, Bakitas MA, et al. Top Ten Tips Palliative Care Clinicians Should Know About Telepalliative Care. J Palliat Med 2019 Aug;22(8):981-985. [doi: 10.1089/jpm.2019.0278] [Medline: 31237467]
16. Cultural Competence in Health Care: Is it important for people with chronic conditions?. URL: https://hpi.georgetown.edu/cultural/ [accessed 2020-07-31]

Abbreviations

DISA: Direct Inward System Access
HIPPA: Health Insurance Portability and Accountability Act

Edited by G Eysenbach; submitted 10.07.20; peer-reviewed by B Liu, R Hendricks-Sturrup; comments to author 28.07.20; revised version received 06.08.20; accepted 06.08.20; published 27.08.20.

Please cite as:
Nemetz ETA, Urbach DR, Devon KM
The Art of Surgery: Balancing Compassionate With Virtual Care
J Med Internet Res 2020;22(8):e22417
URL: http://www.jmir.org/2020/8/e22417/
doi:10.2196/22417
PMID:32852276

Viewpoint

The P Value Line Dance: When Does the Music Stop?

Marcus Bendtsen[1], PhD

Department of Health, Medicine and Caring Sciences, Division of Society and Health, Linköping, Sweden

Corresponding Author:
Marcus Bendtsen, PhD
Department of Health, Medicine and Caring Sciences
Division of Society and Health
Linköping University
Linköping, 581 83
Sweden
Phone: 46 13 28 10 00
Email: marcus.bendtsen@liu.se

Abstract

When should a trial stop? Such a seemingly innocent question evokes concerns of type I and II errors among those who believe that certainty can be the product of uncertainty and among researchers who have been told that they need to carefully calculate sample sizes, consider multiplicity, and not spend *P* values on interim analyses. However, the endeavor to dichotomize evidence into significant and nonsignificant has led to the basic driving force of science, namely uncertainty, to take a back seat. In this viewpoint we discuss that if testing the null hypothesis is the ultimate goal of science, then we need not worry about writing protocols, consider ethics, apply for funding, or run any experiments at all—all null hypotheses will be rejected at some point—everything has an effect. The job of science should be to unearth the uncertainties of the effects of treatments, not to test their difference from zero. We also show the fickleness of *P* values, how they may one day point to statistically significant results; and after a few more participants have been recruited, the once statistically significant effect suddenly disappears. We show plots which we hope would intuitively highlight that all assessments of evidence will fluctuate over time. Finally, we discuss the remedy in the form of Bayesian methods, where uncertainty leads; and which allows for continuous decision making to stop or continue recruitment, as new data from a trial is accumulated.

(J Med Internet Res 2020;22(8):e21345) doi:10.2196/21345

KEYWORDS

sample size; randomized controlled trial; Bayesian analysis; P value; dichotomization; dichotomy; error; uncertainty

Introduction

The (ab-)use of *P* values—the great divider of evidence, minds, and hearts—is, despite a great deal of critique [1-4], still going strong. It is remarkable that less than 60 years ago Hill wrote: “Fortunately I believe we have not yet gone so far as our friends in the USA where, I am told, some editors of journals will return an article because tests of significance have not been applied” [5]. The pendulum has unfortunately swung, as statistical significance has become the arbiter in many scientific disciplines, taking precedence over real-world impact of results, model critique, data quality, etc.

But is it not of the upmost importance to science to have a method to decide if an intervention has an effect? The answer is, to some rather surprisingly, a resounding No. There is no need to spend endless hours writing grant applications, thoughtfully designing experiments, tirelessly recruiting participants, and then chasing follow-up data to reduce attrition—if all you want to know is if an intervention has an effect, then the answer is Yes - all interventions have an effect and you can prove it using *P* value dichotomization as long as you have enough data [6]. The smaller effect you wish to identify, the larger the required sample size will be [7]; and at some point, the sample size required may be greater than the population at hand, which makes the experiment impossible. However, the null hypothesis will always fall given enough data.

This viewpoint will present 2 real-world examples, which will hopefully convince the reader that *P* value dichotomization is not helping scientific discovery and that the praxis needs to be reconsidered. The two trials discussed in this viewpoint have received ethical approval: Regional Ethical Committee in Linköping, Sweden (Dnr 2017/388-31 and Dnr 2018/417-31).

If we Could Only Recruit Some More People, This Smoking Cessation Intervention Would Become Effective!

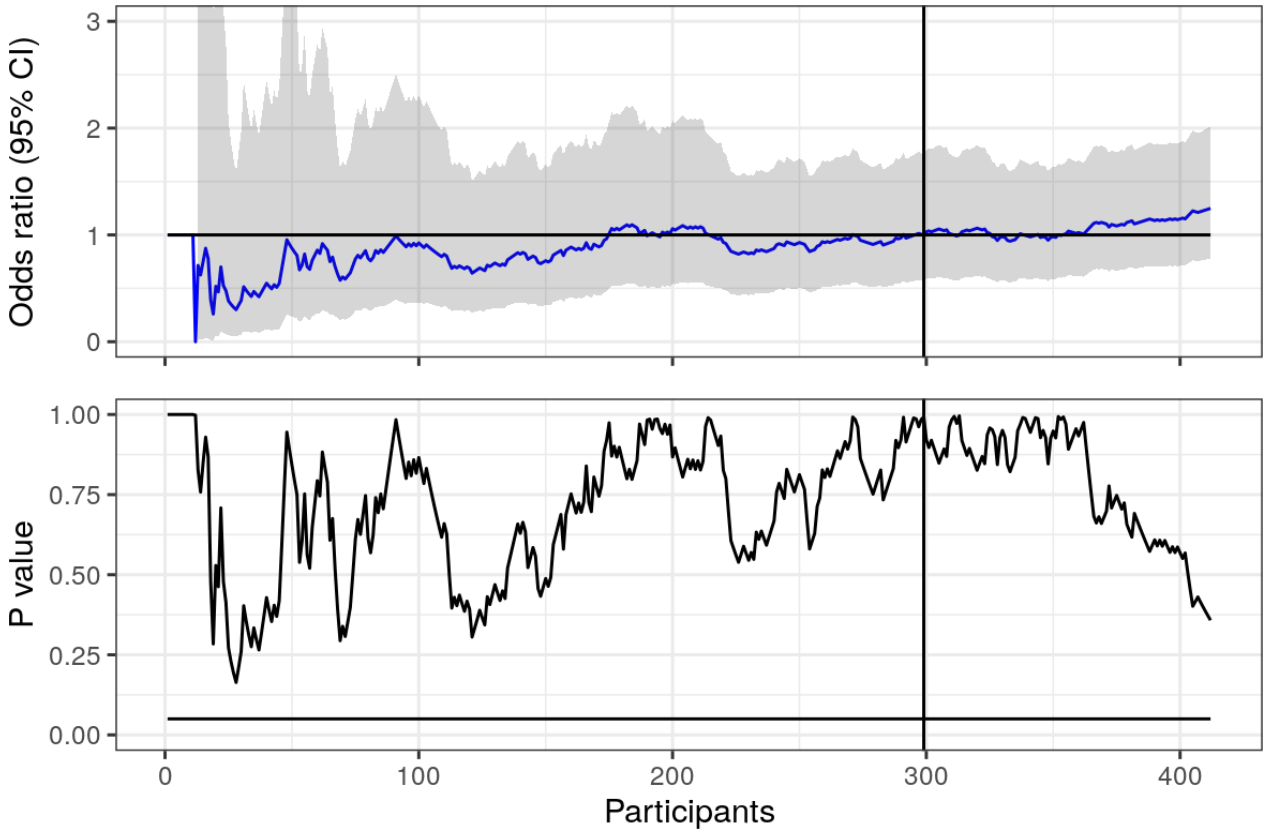
In our first example, we will look back at an experiment conducted among high school students in Sweden [8,9]. The effects of a text messaging intervention on smoking cessation was being estimated, in comparison to a waiting list control group. A 2-arm single blind trial design was used, with equal probability randomization to both groups.

There were 2 outcome measures: prolonged abstinence of smoking (not smoking more than 5 cigarettes over the past 8 weeks) and point prevalence of smoking cessation (not smoking any cigarette the past 4 weeks). Findings suggested, 3 months after randomization, that the effect of the intervention on prolonged abstinence could not be distinguished from the null (OR 1.25, 95% CI 0.78-2.01; *P*=.36); however, the effect on point prevalence could be (OR 1.83, 95% CI 1.12-3.02; *P*=.017).

Recruitment was initially planned to last for 6 months [9]. However, after this time had elapsed, only 386 students had been recruited, less than the prespecified goal of 558. Therefore, it was decided to extend the recruitment period by another 6 months, after which recruitment stopped; and a total of 535 students were recruited.

What would our null hypothesis–focused analyses have looked like had we decided to stop after recruiting 2 participants? 4? 50? 400? In Figure 1 and Figure 2 we have drawn plots, which represent our analyses of prolonged abstinence (Figure 1) and point prevalence (Figure 2) given a certain number of responders. Follow-up attrition was relatively high in this trial, however, for now we will use responders and participants interchangeably.

Figure 1. Prolonged abstinence: odds ratios and *P* values calculated using actual data from trial, plotted over time (number of responders in the study). Horizontal lines represent null value (OR 1) and the .05 statistical significance line. Vertical line represents where the first 6 months of recruitment ended.

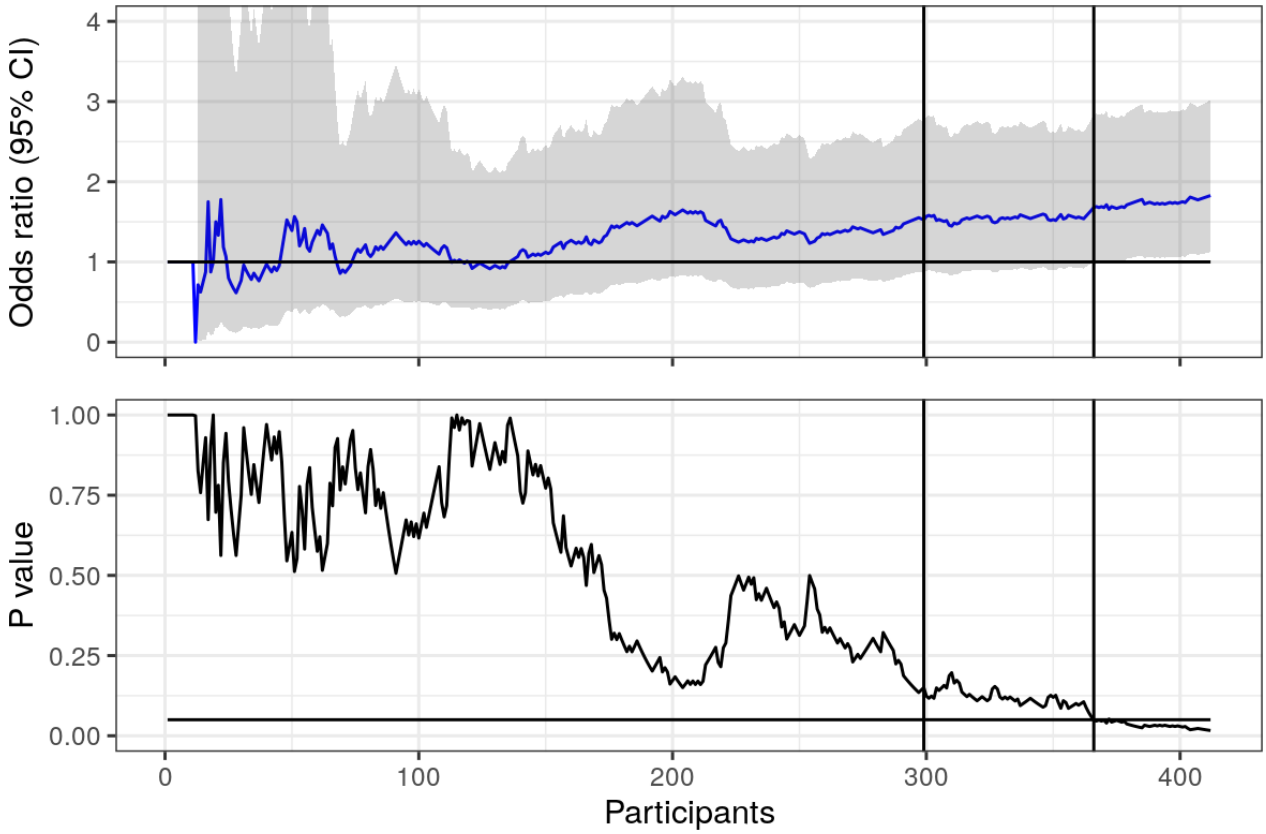


Looking at Figure 1, we can see that odds ratios for prolonged abstinence fluctuate heavily when there are few responders, but then seem to settle a bit as the number increases. We should expect this from point estimates, as when there are few data, each point plays a larger role in the estimate. The *P* values are highly unstable, fluctuating even when the number of responders is large, but never crossing the magic .05 significance line. The vertical line represents the 6-month mark, when the trial was initially planned to stop recruitment. Looking only at the *P* value, our conclusions would not have been much different had we decided to stop at this point. Since it was not significant, the OR is irrelevant (or is it?). However, the estimated odds ratios were different after 6 and 12 months (OR 1.00 vs OR 1.21).

Focusing instead on Figure 2, where point prevalence is analyzed, we see a similar story early on, ORs and *P* values fluctuate, but then things seem to settle a bit. Had the trial ended at the 6-month mark (the left most vertical line), we would have concluded that the effects of the intervention were not distinguishable from the null, thus not rejected the null hypothesis. However, at the end of 12 months, we can see that the *P* value was below the .05 significance line, suggesting

statistical significance and that the effect of the intervention was distinguishable from the null. Finally! After respondent 366 (right most vertical line), we can look at our OR in new light - the OR of 1.68 is statistically significant, there is an effect! Sadly, had we stopped at the previous respondent (number 365), the OR of 1.64 would not have been significant, indistinguishable from the null and not a measure which we should interpret as an effect of the intervention.

Figure 2. Point prevalence: odds ratios and *P* values calculated using actual data from trial, plotted over time (number of responders in the study). Horizontal lines represent null value (OR 1) and the .05 statistical significance line. First vertical line represents where the first 6 months of recruitment ended, second vertical line represents when point prevalence became statistically significant.



What if we had continued recruitment? What if we had another 400 students take part in our trial? Well, we cannot possibly know exactly how these students would have responded; but for the sake of this experiment, it is not strictly necessary. We can pretend that the new 400 participants are similar to the participants we already have, and therefore, sample 400 respondents from those already in our trial (with replacement). The new OR and *P* value timeline can be seen in Figure 3 and Figure 4.

As we can see in Figure 3, it turns out that the intervention actually did have an effect on prolonged abstinence. We just did not have enough respondents to distinguish it from the null using the .05 threshold, but now we do. We could argue that resampling from a sample with a nontrivial OR and recalculating the *P* value will of course result in a lower *P* value, but that is exactly the point! Statistical significance is a function of the sample size, so any effect can be statistically significant if there are enough participants; and all interventions have an effect [6,7]. Note that there is a lot of crossing the significance line between 600 and 800 respondents, many opportunities to end the trial and cry wolf. In Figure 4, the *P* value has essentially flatlined.

Figure 3. Prolonged abstinence (with sampled data): odds ratios and *P* values calculated using actual and sampled data from trial, plotted over time (number of responders in the study). Horizontal lines represent null value (OR 1) and the .05 statistical significance line. The vertical line represents where the first 6 months of recruitment ended.

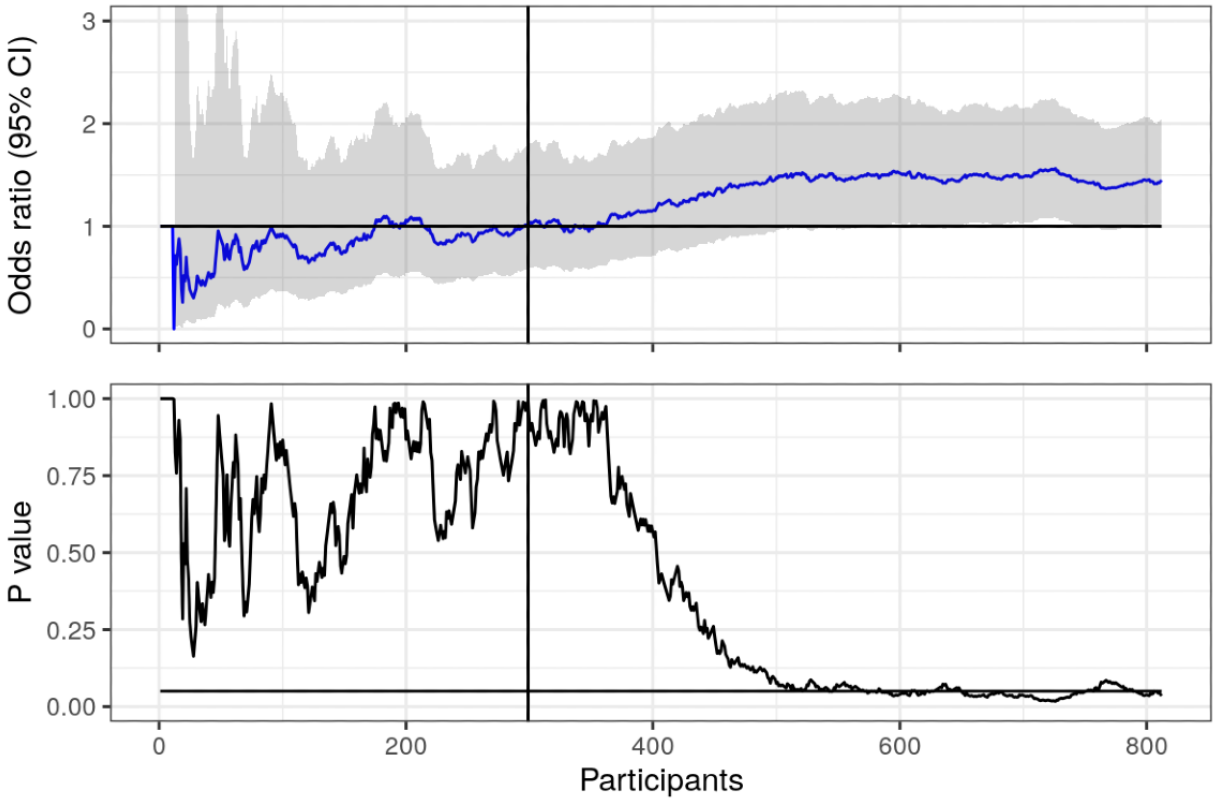
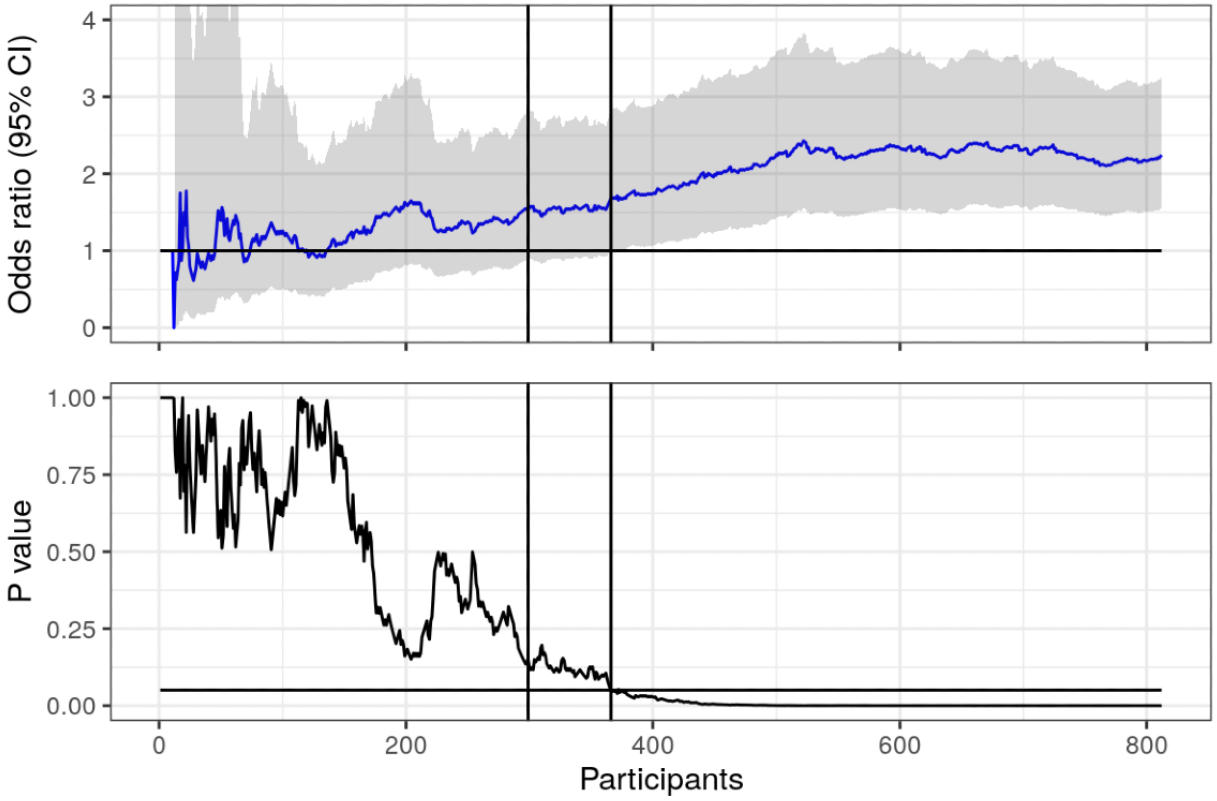


Figure 4. Point prevalence (with sampled data): odds ratios and *P* values calculated using actual and sampled data from trial, plotted over time (number of responders in the study). Horizontal lines represent null value (OR 1) and the .05 statistical significance line. First vertical line represents where the first 6 months of recruitment ended, second vertical line represents when point prevalence became statistically significant.



So, should we simply continue recruiting until our *P* values flatline, forgoing a prespecified sample size? That is likely a bad idea if there are harms and costs involved, which would make it ethically irresponsible. One may argue that if you cannot reject the null given your prespecified effect and sample size, then the value of flatlining the *P* value at some smaller effect size is not worth the risk. However, such thinking may be irresponsible, as we shall see.

But it was Statistically Significant Yesterday!

Our second example concerns an experiment of estimating the effects of receiving 1 of 2 different text messages with alcohol and health information. This experiment was nested within a larger (ongoing) trial of a digital alcohol intervention [10]. Participants who were randomized to the control arm of the trial were randomized further into 2 arms. The first arm received a public health message regarding alcohol, violence, and cancer. The second arm also received information about alcohol, violence, and cancer, but the information was worded in an alcohol industry manner, focusing on responsible drinking and downplaying the evidence on the risks of alcohol. At the end of both text messages was a hyperlink, which lead to more information about alcohol and health, and the experiment outcome was whether or not participants pressed the hyperlink.

After having recruited 150 participants in the experiment, we were curious to see how things were progressing. Interim analyses were interesting. It turned out that participants in the public health arm were far more likely to press the link (OR 2.26, 95% CI 1.18-4.42; *P*=.015). Figure 5 shows, as before, ORs and *P* values given a certain number of participants. The *P* value does at first glance look like it has flatlined.

However, as the trial progressed, and more participants were recruited, the fickleness [11] of the *P* value became apparent. After one year, 560 participants had been recruited, and the ORs and *P* values plot (Figure 6) looked markedly different. Now, there was no statistical significance (OR 1.14, 95% CI 0.82-1.59; *P*=.43). However, there were plenty of times where the trial might have ended, and a statistically significant result would have been the result; and if the trial continues recruitment, we will eventually have a significant result again (as discussed in the previous example).

Figure 5. Pressed link in text message: odds ratios and *P* values calculated and plotted over time (number of participants in the study). Horizontal lines represent null value (OR 1) and the .05 statistical significance line.

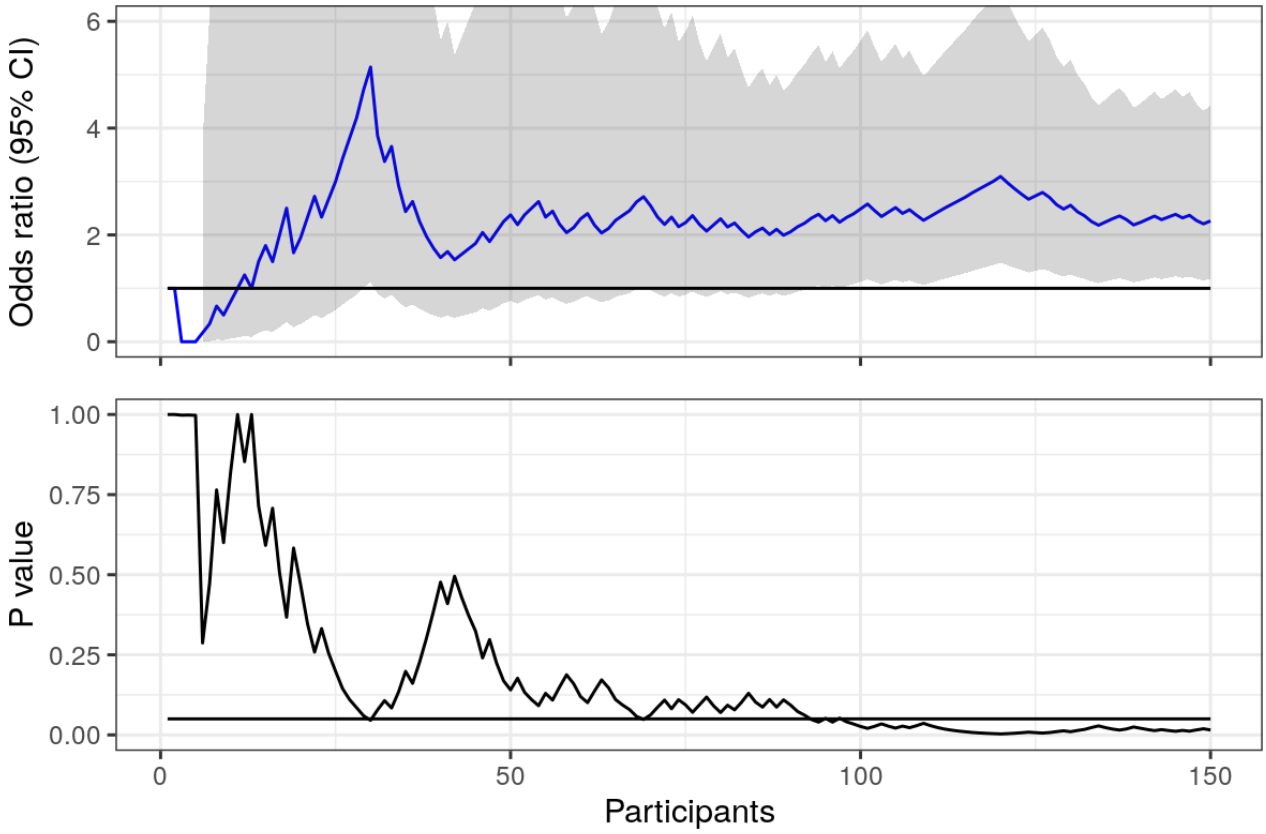
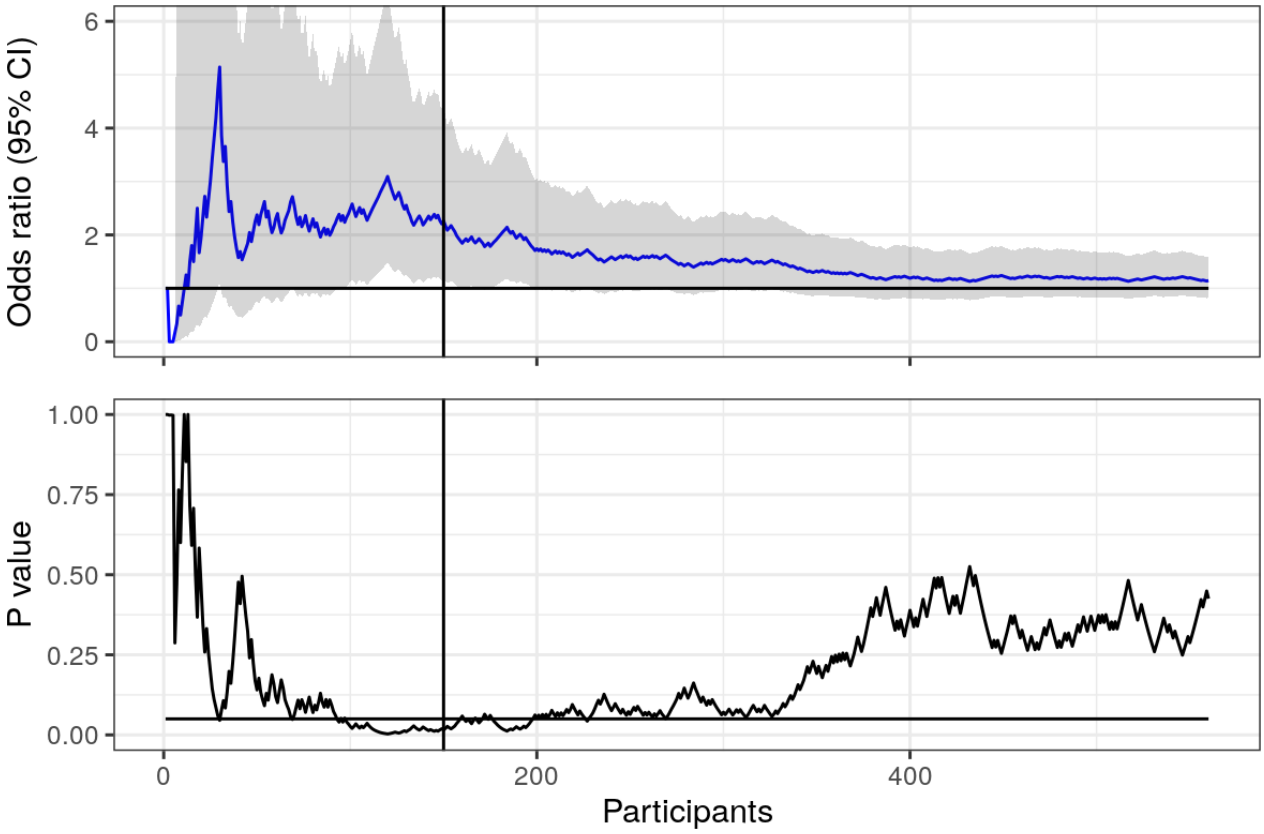


Figure 6. Pressed link in text message: odds ratios and *P* values calculated and plotted over time (number of participants in the study). Horizontal lines represent null value (OR 1) and the .05 statistical significance line. Vertical line represents when interim analyses were conducted.



There was no prespecified sample size for this exploratory outcome, since it was nested within a larger trial. However, think about the number of times you have read reports of trials, and grant and ethics applications, where the power analysis has concluded that recruiting approximately 150 participants will suffice. Given a large enough effect size, this will hold for the calculation; and if researchers are "lucky," it will hold in their experiment. How many reports have you read where statistically significant results have led to a discussion about important results based on 150 participants? What would have happened if they continued recruiting another 50, 100, or 200 participants?

Letting Bayes be the Conductor

One of the core issues underlying the 2 examples given herein is that point estimates and *P* values are very fickle when taking the traditional approach. This fickleness is caused by assuming that data alone is all we care about, and we take no action towards tempering our expectations. That is, after collecting data from 2 participants, 1 from the intervention group and 1 from the control group, the effect is estimated to be the difference between these two. Put in another way, we are susceptible to drawing conclusions from small sample sizes, and we ignore our belief that our interventions will likely have small to modest effects.

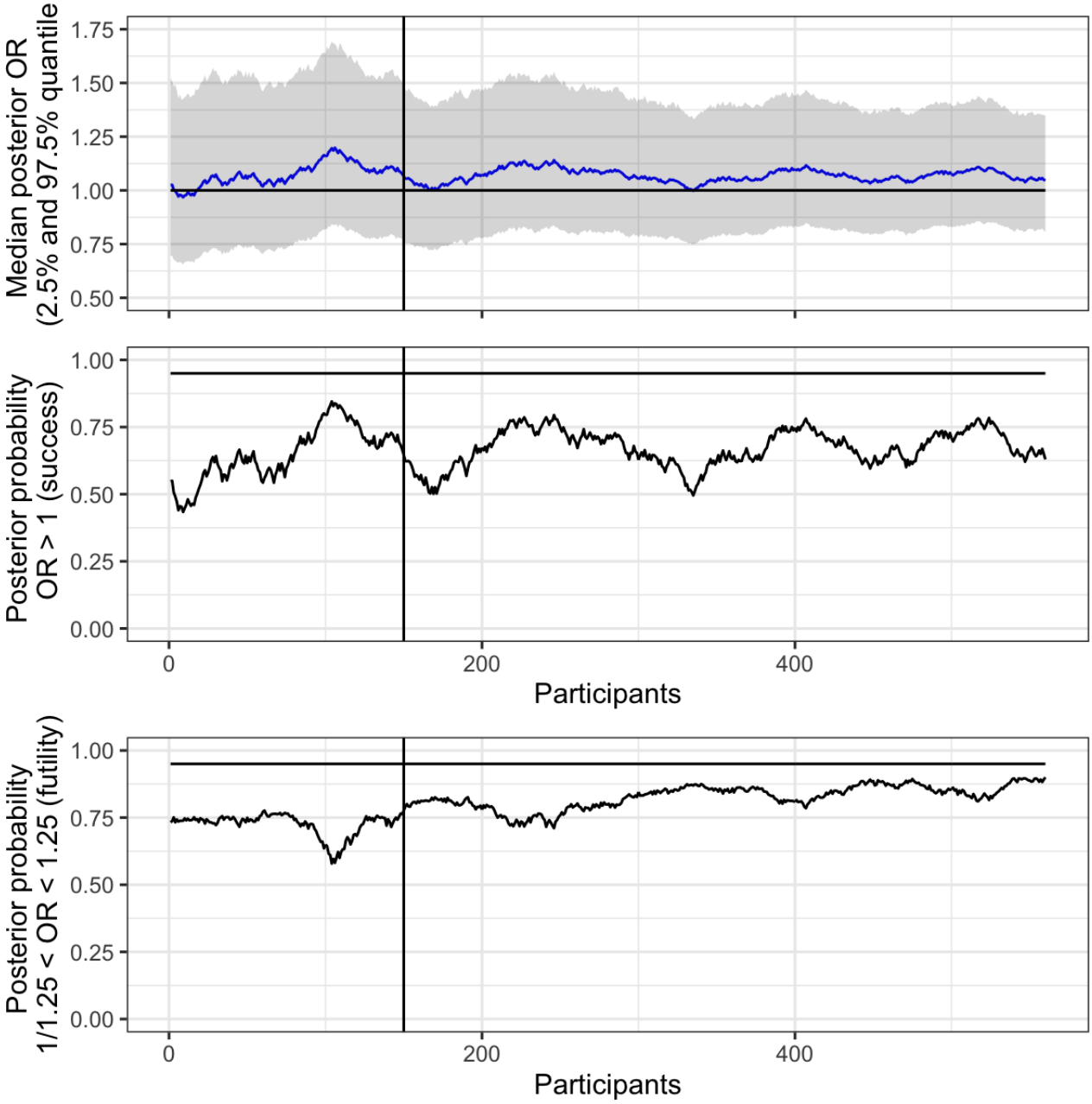
An alternative is to take a Bayesian approach to inference [12-14]. While a full discussion about the details of Bayesian inference cannot possibly fit here, the essentials can be captured as follows: You count what you see (the data), and balance this with what you expect (known as the *prior*). For instance, if you believe that there may be both black and white swans (this is your prior), then you do not make the conclusion that all swans are white after having seen a single white swan. After having seen thousands of white swans you may decide that it is more likely that swans are white, but you do not say that it is impossible for swans to be black.

From a Bayesian perspective, a trial is a series of repeated experiments; and each time we collect data from a participant, we can update our inference about the trial outcomes. This is often referred to as a Bayesian group sequential design [15,16]. We use prespecified criteria to decide if a trial is a success, or if it is futile to continue recruitment (and possibly also if it is unethical [17] to continue due to harm). These criteria can be evaluated as many times as we like. Using our second example from before, where we studied prevalence of pressing on a link in a text message, we may define our success criteria as: *If there is more than a 95% probability that the OR is greater than 1, we end the trial and call it a success*. A criterion for futility may be as follows: *If there is more than a 95% probability that the OR is somewhere between 1/1.25 and 1.25, then we believe that the groups are essentially equal, and there is no need to further investigate the intervention*. These 2 probabilities (for success and futility) can then be calculated using what is commonly referred to as a *skeptical* prior, which encodes a strong a priori belief that the intervention has no effect, and that the data needs to convince us otherwise.

In Figure 7, we have plotted the median OR (top plot), the probability of success (middle plot), and the probability of

futility (lower plot) given the criteria set above. A skeptical prior for the intervention effect (normal distribution with mean 0 and SD 0.2) has been used for both the success and futility criteria. While full details are scarce in this viewpoint, it should be clear from these plots that by tempering our expectations, we have avoided making conclusions early. The *P* value approach (Figure 6) called for a statistically significant effect after 150 participants (vertical line), which was later overturned. We have avoided this by using a Bayesian approach; and it looks like the experiment will be considered futile as new data is collected, as the probability of an OR between 1/1.25 and 1.25 (lower plot) is gradually increasing towards 95% (our predefined futility criteria).

Figure 7. Pressed link in text message: median posterior odds ratios and probability of success and futility plotted over time (number of participants in study). Horizontal line represents null value (OR 1), and 95% probability of success and futility respectively. Vertical line represents interim analysis. A skeptical prior (mean 0, SD 0.2) was used for all inference.



Discussion

The 2 examples herein are not fictive, and they are by no means unique. We invite readers to plot effect estimates and *P* values in a similar fashion as we have and reflect on the robustness of their past conclusions. If such plots were commonplace in scientific papers, would readers' or reviewers' interpretations of the findings change? There is no finger pointing here, we are all as a collective responsible to ensure that the scientific method is sound.

The replication crisis in the social sciences is proof that methods built on dichotomization of evidence are not scientific [18]. It needs to stop. However, if researchers, journals, reviewers, funding agencies, media, and the general public, continue to crave statements of true and false—effect or no effect—then

there is no silver bullet, which will make the line dancing cease [19]. Consulting confidence intervals is veiled hypothesis testing, and reducing the *P* value threshold to .005 [20] is just kicking the can down the road and opening the door for new *P* value–hacking and selective reporting issues [1]. Add to this that *P* values (and confidence intervals) are consistently being misinterpreted [4,21], and even highly respected journals are allowing nonsignificance to be interpreted as an absence of effect [22]. A recent example was the conclusion that lopinavir–ritonavir treatment for COVID-19 "was not beneficial in comparison to standard care" [23], backed by a hazard ratio for clinical improvement of 1.31 with a 95% CI 0.95-1.80, which is not statistically significant, but which cannot rule out a hazard ratio of 1.80.

What is the alternative? Well, as Gelman and Carlin put it [19], "...resist giving clean answers when that is not warranted by the data. Instead, do the work to present statistical conclusions with uncertainty rather than as dichotomies." Doing so is natural from a Bayesian perspective, as posterior distributions can directly describe the relative compatibility of different models given the data (rather than the other way around). In fact, Bayesian inference answers the question that researchers want to ask (but have been told that they cannot): What is the probability that an intervention had a positive effect? Interventions should not be dismissed because the design of an experiment did not allow the *P* value line to be crossed, as we have seen, it may be sheer luck that an experiment stopped exactly when it could show significance.

We recognize the importance for careful planning of trials, including giving estimates on the number of individuals necessary to recruit. However, prespecifying sample sizes based on type I and type II errors is not only ignorant to the fact that it is not possible to know how many individuals are necessary to recruit (it may be considered a random variable itself), but may also be considered unethical as it may lead to over-recruitment, detecting harm and benefit later than necessary [16]. Using multiple looks at the data throughout the trial and making judgments based on null hypothesis tests is not only problematic due to its reliance on fickle point estimates (as demonstrated herein), but also inflates type I errors due to multiplicity if not handled correctly [24]. Instead, a Bayesian group sequential design [15,16] allows for continuous monitoring as data is collected, utilizing target posterior probabilities for success and futility, such that a decision can be made to stop or continue recruitment each time new data is available without concern for multiplicity [25].

It should be noted that all assessments of evidence will fluctuate over time, as we have shown in the enclosed examples. One aspect of this is that smaller samples may not represent the study population well. Another is that changes may occur in the underlying study population as we recruit over time, which means that we may be sampling from different regimes [26-28] in the data (for instance, due to seasonal differences). However, in a Bayesian group sequential design we can use a skeptical prior which will draw back the posterior probability of effect and posterior median, which will automatically correct for too early looks at the data [25]. This goes some way towards protecting from small sample sizes and regimes that may be misrepresentative of the population we wish to study.

Uncertainty is the driving force of science, and uncertainty *in* can never result in certainty *out*. Uncertainty leads in Bayesian methods, and it allows us to more clearly judge our findings in light of it. Convoluted rules for sample size estimation, *P* value spending, or correction for multiplicity are all artifacts from thinking that certainty can be the result of uncertainty. We can increase our understanding of the uncertain through repeating experiments, as was Fisher's intention, which ultimately is the goal of science [29].

Conflicts of Interest

MB owns a private company (Alexit AB), which develops and disseminates digital lifestyle interventions to the general public as well as to public and private sector organizations. Alexit AB had no part in planning, funding, or writing this viewpoint.

References

1. Amrhein V, Greenland S. Remove, rather than redefine, statistical significance. Nat Hum Behav 2017 Sep 25;2(1):4-4. [doi: 10.1038/s41562-017-0224-0]
2. McShane BB, Gal D. Statistical Significance and the Dichotomization of Evidence. Journal of the American Statistical Association 2017 Oct 30;112(519):885-895. [doi: 10.1080/01621459.2017.1289846]
3. Nuzzo R. Scientific method: statistical errors. Nature 2014 Feb 13;506(7487):150-152. [doi: 10.1038/506150a] [Medline: 24522584]
4. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon Statistical Significance. The American Statistician 2019 Mar 20;73(sup1):235-245. [doi: 10.1080/00031305.2018.1527253]
5. Hill AB. The Environment and Disease: Association or Causation? In: Proceedings of the Royal Society of Medicine. 2016 Sep Presented at: Proceedings of the Royal Society of Medicine; May; London p. 295-300. [doi: 10.1177/003591576505800503]
6. Cohen J. The earth is round (p < .05). American Psychologist 1994;49(12):997-1003. [doi: 10.1037/0003-066X.49.12.997]
7. Cohen J. Things I have learned (so far). American Psychologist 1990;45(12):1304-1312. [doi: 10.1037/0003-066x.45.12.1304]
8. Müssener U, Linderoth C, Thomas K, Bendtsen M. mHealth smoking cessation intervention among high school students: 3-month primary outcome findings from a randomized controlled trial. PLoS One 2020;15(3):e0229411 [FREE Full text] [doi: 10.1371/journal.pone.0229411] [Medline: 32142514]

9. Thomas K, Bendtsen M, Linderoth C, Müssener U. mHealth smoking cessation intervention among high-school pupils (NEXit Junior): study protocol for a randomized controlled trial. Trials 2018 Nov 22;19(1). [doi: 10.1186/s13063-018-3028-2]
10. Bendtsen M, McCambridge J. Reducing Alcohol Consumption Among Risky Drinkers in the General Population of Sweden Using an Interactive Mobile Health Intervention: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Apr 18;8(4):e13119 [FREE Full text] [doi: 10.2196/13119] [Medline: 30998221]
11. Halsey LG, Curran-Everett D, Vowler SL, Drummond GB. The fickle P value generates irreproducible results. Nat Methods 2015 Mar;12(3):179-185. [doi: 10.1038/nmeth.3288] [Medline: 25719825]
12. Bendtsen M. A Gentle Introduction to the Comparison Between Null Hypothesis Testing and Bayesian Analysis: Reanalysis of Two Randomized Controlled Trials. J Med Internet Res 2018 Oct 24;20(10):e10873 [FREE Full text] [doi: 10.2196/10873] [Medline: 30148453]
13. Bendtsen M. An Electronic Screening and Brief Intervention for Hazardous and Harmful Drinking Among Swedish University Students: Reanalysis of Findings From a Randomized Controlled Trial Using a Bayesian Framework. J Med Internet Res 2019 Dec 17;21(12):e14420 [FREE Full text] [doi: 10.2196/14420] [Medline: 31845903]
14. Bendtsen M. Electronic Screening for Alcohol Use and Brief Intervention by Email for University Students: Reanalysis of Findings From a Randomized Controlled Trial Using a Bayesian Framework. J Med Internet Res 2019 Nov 07;21(11):e14419 [FREE Full text] [doi: 10.2196/14419] [Medline: 31697242]
15. Gsponer T, Gerber F, Bornkamp B, Ohlssen D, Vandemeulebroecke M, Schmidli H. A practical guide to Bayesian group sequential designs. Pharm Stat 2014;13(1):71-80. [doi: 10.1002/pst.1593] [Medline: 24038922]
16. Berry DA. Bayesian clinical trials. Nat Rev Drug Discov 2006 Jan;5(1):27-36. [doi: 10.1038/nrd1927] [Medline: 16485344]
17. Berry DA. Bayesian Statistics and the Efficiency and Ethics of Clinical Trials. Statist. Sci 2004 Feb;19(1):175-187. [doi: 10.1214/088342304000000044]
18. Camerer CF, Dreber A, Holzmeister F, Ho T, Huber J, Johannesson M, et al. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nat Hum Behav 2018 Sep;2(9):637-644. [doi: 10.1038/s41562-018-0399-z] [Medline: 31346273]
19. Gelman A, Carlin J. Some Natural Solutions to the -Value Communication Problem—and Why They Won't Work. Journal of the American Statistical Association 2017 Oct 30;112(519):899-901. [doi: 10.1080/01621459.2017.1311263]
20. Benjamin D, Berger J, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. Nat Hum Behav 2018 Jan;2(1):6-10. [doi: 10.1038/s41562-017-0189-z] [Medline: 30980045]
21. Gigerenzer G. Mindless statistics. The Journal of Socio-Economics 2004 Nov;33(5):587-606. [doi: 10.1016/j.socec.2004.09.033]
22. Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ 1995 Aug 19;311(7003):485 [FREE Full text] [doi: 10.1136/bmj.311.7003.485] [Medline: 7647644]
23. Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, et al. A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. N Engl J Med 2020 May 07;382(19):1787-1799 [FREE Full text] [doi: 10.1056/NEJMoa2001282] [Medline: 32187464]
24. Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Med 2018 Feb 28;16(1):29 [FREE Full text] [doi: 10.1186/s12916-018-1017-7] [Medline: 29490655]
25. Harrell FE. Continuous learning from data: no multiplicities from computing and using Bayesian posterior probabilities as often as desired. Statistical Thinking. URL: https://www.fharrell.com/post/bayes-seq/ [accessed 2020-07-05]
26. Bendtsen M. Regimes in baseball players' career data. Data Min Knowl Disc 2017 May 27;31(6):1580-1621. [doi: 10.1007/s10618-017-0510-5]
27. Bendtsen M. Regime aware learning. In: Proceedings of the Eighth International Conference on Probabilistic Graphical Models. 2016 Presented at: Eighth International Conference on Probabilistic Graphical Models; 2016; Lugano p. 1-12.
28. Bendtsen M, Peña J. Modelling regimes with Bayesian network mixtures. In: Linko¨ping Electronic Conference Proceedings No. 137. 2017 Presented at: 30th Annual Workshop of the Swedish Artificial Intelligence Society SAIS 2017; 2017; Karlskrona p. 20-29.
29. Mark DB, Lee KL, Harrell FE. Understanding the Role of P Values and Hypothesis Tests in Clinical Research. JAMA Cardiol 2016 Dec 01;1(9):1048-1054. [doi: 10.1001/jamacardio.2016.3312] [Medline: 27732700]

Edited by G Eysenbach; submitted 11.06.20; peer-reviewed by A Bedendo, H Wong; comments to author 03.07.20; revised version received 05.07.20; accepted 27.07.20; published 27.08.20.

Please cite as:
Bendtsen M
The P Value Line Dance: When Does the Music Stop?
J Med Internet Res 2020;22(8):e21345
URL: http://www.jmir.org/2020/8/e21345/
doi:10.2196/21345
PMID:32852275

Viewpoint

Improving the Primary Care Consultation for Diabetes and Depression Through Digital Medical Interview Assistant Systems: Narrative Review

Geronimo Jimenez[1,2], MA; Shilpa Tyagi[3], MBBS; Tarig Osman[1], MBBS; Pier Spinazze[1], MBA, MDiv; Rianne van der Kleij[2], PhD; Niels H Chavannes[2], MD, PhD; Josip Car[1], MD, PhD

[1]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

[2]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands

[3]Saw Swee Hock School of Public Health, National University of Singapore, Singapore

Corresponding Author:
Geronimo Jimenez, MA
Centre for Population Health Sciences
Lee Kong Chian School of Medicine
Nanyang Technological University
11 Mandalay Rd, Clinical Sciences Building, Level 18
Singapore
Phone: 65 69047027
Email: geronimo.jimenez@gmail.com

Abstract

Background: Digital medical interview assistant (DMIA) systems, also known as computer-assisted history taking (CAHT) systems, have the potential to improve the quality of care and the medical consultation by exploring more patient-related aspects without time constraints and, therefore, acquiring more and better-quality information prior to the face-to-face consultation. The consultation in primary care is the broadest in terms of the amount of topics to be covered and, at the same time, the shortest in terms of time spent with the patient.

Objective: Our aim is to explore how DMIA systems may be used specifically in the context of primary care, to improve the consultations for diabetes and depression, as exemplars of chronic conditions.

Methods: A narrative review was conducted focusing on (1) the characteristics of the primary care consultation in general, and for diabetes and depression specifically, and (2) the impact of DMIA and CAHT systems on the medical consultation. Through thematic analysis, we identified the characteristics of the primary care consultation that a DMIA system would be able to improve. Based on the identified primary care consultation tasks and the potential benefits of DMIA systems, we developed a sample questionnaire for diabetes and depression to illustrate how such a system may work.

Results: A DMIA system, prior to the first consultation, could aid in the essential primary care tasks of case finding and screening, diagnosing, and, if needed, timely referral to specialists or urgent care. Similarly, for follow-up consultations, this system could aid with the control and monitoring of these conditions, help check for additional health issues, and update the primary care provider about visits to other providers or further testing. Successfully implementing a DMIA system for these tasks would improve the quality of the data obtained, which means earlier diagnosis and treatment. Such a system would improve the use of face-to-face consultation time, thereby streamlining the interaction and allowing the focus to be the patient's needs, which ultimately would lead to better health outcomes and patient satisfaction. However, for such a system to be successfully incorporated, there are important considerations to be taken into account, such as the language to be used and the challenges for implementing eHealth innovations in primary care and health care in general.

Conclusions: Given the benefits explored here, we foresee that DMIA systems could have an important impact in the primary care consultation for diabetes and depression and, potentially, for other chronic conditions. Earlier case finding and a more accurate diagnosis, due to more and better-quality data, paired with improved monitoring of disease progress should improve the quality of care and keep the management of chronic conditions at the primary care level. A somewhat simple, easily scalable technology could go a long way to improve the health of the millions of people affected with chronic conditions, especially if working in conjunction with already-established health technologies such as electronic medical records and clinical decision support systems.

(J Med Internet Res 2020;22(8):e18109) doi:10.2196/18109

KEYWORDS

digital medical interview assistant, computer-assisted history taking; primary care; chronic conditions

Introduction

Digital medical interview assistant (DMIA) systems, also known as computer-assisted history taking (CAHT) systems, are software programs that allow patients to provide their medical history electronically prior to the consultation, which can be done remotely, via a web-based portal, or on-site in clinic, via tablets or kiosks, before clinical review [1]. DMIA systems have the potential to improve the quality of care and the medical consultation by exploring more patient-related aspects without time constraints and, therefore, acquiring more and better-quality information [2,3].

Primary care is often considered a cornerstone of health care systems [4]. The stronger the country's primary care orientation, the better the health outcomes it obtains (eg, in terms of all-cause mortality, all-cause premature mortality, and cause-specific premature mortality for several conditions), regardless of the differences in health system characteristics, such as gross domestic product per capita, total physicians per 1000 population members, and percentage of elderly people [5]. As such, one of primary care's core functions involves diagnosing and treating chronic conditions [6]. Diabetes and depression, two very prevalent chronic conditions, form part of the majority of nonreferred ambulatory visits to office-based physicians [6,7]. The global prevalence of diabetes reached 8.5% in 2014, affecting 422 million people [8], while depression is the most common mental health disorder and affects more than 264 million people worldwide [9]. Both are multifactorial, chronic conditions that require frequent reassessment and readjustment for their management [10].

The alarming rise of chronic conditions is increasingly straining health systems worldwide [11], and primary care has been widely proposed as one possible solution to tackle this issue [4,5], as the characteristics of chronic conditions make primary care the ideal level of care to manage and treat them. Here, we aim to explore how DMIA systems may be used specifically in the context of primary care, to enhance the consultations for diabetes and depression, as exemplars for chronic conditions. In order to do this, we delineated the particular characteristics of, and the tasks to be performed at, the primary care consultation that could be aided by a DMIA system. This helped identify the key topics a DMIA system should ask about before the primary care consultations for these chronic conditions. In addition, we discuss its potential benefits for the primary care consultation, as well as important considerations for the successful implementation of such a technology in the primary care context.

Methods

A narrative review of the literature [12] was conducted between June and August 2019 and focused on two topics:

1. Characteristics of the primary care consultation in general (ie, scope, structure, communication, time, knowledge, and skills required) and on the primary care medical consultation for diabetes and depression specifically. A Google search was performed using the terms "primary care consultation" for obtaining general information regarding the primary care consultation from a variety of sources, including published work and grey literature. Additionally, a search through the Nanyang Technological University Library was performed using the same terms to locate textbooks regarding the general primary care consultation and for the consultation for diabetes and depression, specifically.
2. The impact of DMIA and CAHT systems on the medical consultation. Searches in PubMed and MEDLINE (Medical Literature Analysis and Retrieval System Online) were performed using the terms "computer-assisted history taking" and "automated history taking."

A purposeful sample of articles was included based on relevance to build the Results and Discussion section. Robust evidence regarding CAHT and DMIA systems and their impact on the medical consultation was limited, as most studies evaluating these systems focused on other aspects, such as acquiring dietary or family history to estimate risk for diabetes, studies evaluating computer-assisted data collection for specific populations or related to screening, studies evaluating the test-retest reliability of computer-based medical histories, or efforts to generally computerize medicine, just to give some examples [13-18]. Relevant insights that were most useful for our purposes came mainly from commentaries, perspectives, and historical pieces [2,3,19], some of which analyzed or identified existing CAHT and DMIA software programs [1,20,21].

Thematic analysis of the extracted data from medical textbooks and literature helped identify higher-order themes and the specific, recurrent elements [22] that primary care consultations contain in relation to diabetes and depression. At the same time, we identified the specific areas where DMIA systems could improve the primary care consultation for these conditions. By combining both sources of information, we mapped the areas of the primary care consultation for diabetes and depression that might be improved by the use of DMIA systems. We did this for both the first presentation or consultation with the primary care provider and the follow-up visits, as the needs and tasks performed at these consultations are different. The information is reported following the structure of a narrative-style literature review [12].

Additionally, we developed a box (see Multimedia Appendix 1) representing the main identified tasks a primary care provider should perform at the initial (ie, screening and diagnosing) and follow-up primary care consultations (ie, general and clinical monitoring, checking for additional health issues, and coordinating); we completed it with the questions that a DMIA system should ask in order to facilitate these tasks, specifically for potential diabetes and depression patients. The box also

includes the corresponding improvements for these primary care consultations.

Results and Discussion

Advantages and Disadvantages of DMIA Systems

No other medical consultation is as broad in terms of topics to be explored (ie, medical and clinical, psychological, and biosocial) [6,10,23] and limited in terms of time as the primary care consultation: mean consultation time is 5 minutes or less [24] or around 9 minutes, with general practitioners feeling that a consultation of less than 10 minutes for primary care was inadequate [25]. A DMIA system could improve primary care consultations in several ways, based on the benefits related to the medical consultation, broadly conceptualized. Textbox 1 and the following paragraphs summarize potential benefits and disadvantages that DMIA systems may bring forth for medical consultations in general, based on both theoretical work and on studies evaluating the capacities of these systems. In the textbox, some articles are perspective or commentary articles, theorizing about the capacities of DMIA systems [1-3,19,26], and others are studies evaluating systems and presenting results [21,27-29].

Textbox 1. Potential benefits and disadvantages of digital medical interview assistant (DMIA) systems for the medical consultation, compared to face-to-face consultations.

Potential benefits of DMIA systems:

- Collect history for health screening or comprehensive clinical consultation [1]
- Collect more complete, accurate, and reliable information [1,3,19,26]
- Potential to increase diagnostic certainty [3]
- Considerably shorten the time spent on history taking, dictation, and documentation [1]
- Streamline office visit; allows for consultation to be focused on identified concerns and problems [3]
- Promote rapport, communication, and decision making [1,2,19]
- Can be integrated with electronic medical records, electronic health records, and online patient diaries, improving access to data [1,3,19,26]
- Enable triage prioritization and improve referral of patients [1]
- Help prepare patient and primary care provider for the consultation [3,19]
- Collect more sensitive information [1]

Potential disadvantages of DMIA systems:

- Technical issues—poor programming and design may result in [21]:
  - Missing relevant information
  - Collection of irrelevant information
  - Erroneous information
- Human-computer interaction-related issues:
  - Perceiving the computer interview as impersonal [1,27]
  - Inability to detect nonverbal behavior [19]
- Duplication of effort—primary care provider attempts to confirm all responses [3]
- Require patient's familiarity with technology and computer literacy [1,3,19]
- Require technical supervision and maintenance [1,19,28]
- Require a variety of factors (ie, legislative, organizational, and physician-level factors) to allow the successful implementation of eHealth innovations [28,29]

First, DMIA systems have the capacity to acquire a more comprehensive set of information than that attainable during a face-to-face consultation. This could provide greater insights into potential risk factors and, possibly, suggest a more accurate differential diagnosis, prior to seeing the patient [2,3,30]. Secondly, it allows for better use of face-to-face consultation time, as previous existing conditions and information about presenting complaints have already been identified before the patient enters the physician's office [2,3]. This allows for a more streamlined consultation, where time that would have been spent on history taking is shifted to discussing management strategies, building rapport, and focusing on the relationship with the patient.

Third, it improves data quality [2,20,21], which in turn improves the diagnosis, as the quality of the diagnostic process is significantly affected by the accuracy of the information made available to the primary care provider. Fourth, DMIA systems include more up-to-date information, including that available from tests and other health care providers, which helps maintain

continuity of information across different health care providers [30].

Additionally, once a diagnosis has been made by the primary care provider and subsequent consultations have been scheduled, DMIA systems can help flag specific aspects that need attention related to the ongoing management or treatment of the condition, and they can collect subsequent patient concerns prior to the follow-up consultation. As a result, the primary care provider in follow-up consultations can focus on guiding the patient regarding self-management aspects and provide further education that might be needed [3]. Alternatively, the information in a DMIA system may flag the need for specialist visits or management from other team members (ie, allied health professionals) in advance of seeing the patient, leading to earlier referrals or timely, urgent treatment, if needed [30].

However, there are also reported disadvantages of DMIA systems, which include possible technical issues, related to problems with programming or design, that may lead to missing or erroneous information [21]. Potential problems could occur as a result of a patient interacting with a computer, such as perceived impersonality of the health care provider–patient interaction or missing body language cues [1,19,27]. This last issue could be overcome by new facial and/or body movement recognition technologies, such as affective computing, where automated analysis of facial expressions can provide accurate depression diagnosis, for example [31]. Also, it has been described that some primary care providers attempt to confirm all the answers, duplicating efforts [3]. Finally, DMIA systems, as with other technologies, require technical supervision and maintenance [1,19,28], which may result in additional resource spending.

Primary Care Consultation Tasks for Diabetes and Depression: Where Could DMIA Systems Make a Difference?

Prior to the Initial Consultation

Overview

Two important tasks of a primary care provider in the first consultation with a new patient are to identify a health condition, by case finding or screening, and diagnose that health condition [10,23,32-34]. The key aspect to these tasks is information gathering. As such, a preconsultation (ie, prior to the first consultation) DMIA system can exhaustively explore and ask questions about the different areas related to screening and diagnosis of a chronic condition prior to the first consultation, without particular time constraints.

Case Finding and Screening

Comprehensive case finding, along the lines of the biopsychosocial model, should include a generic set of questions exploring general health-related aspects, as well as more targeted screening questions, depending on the answers provided in the general questions section. General aspects may refer to lifestyle and health behaviors (ie, exercise, sleep duration, diet, general mood, alcohol and drug use, etc) and social behavior (eg, family, work conditions, and community aspects). Both of these question sets can provide hints for high-risk factors and behaviors, including patient beliefs and perspectives on the illness. Depending on the answers provided to the general screening questions, more targeted specific screening questions could be presented for the risk factors of hypertension, obesity, cancer, and other chronic conditions [35-37]. Answers from these could provide clinically relevant information for an accurate and earlier diagnosis (see Multimedia Appendix 1).

Diagnosing

Based on the responses from the case finding process, using branching logic, further questioning may dive deeper into specific risk factors to attempt a diagnosis. For example, for individuals at risk of diabetes, the system can ask questions related to possible prediabetes and signs and symptoms (see Multimedia Appendix 1) [35,36]; this may include plasma glucose level test results, if available [34]. For depression, the systems could focus on possible depressive and mood disorder diagnosis questions or screening tests, such as the 2-key question approach (see Multimedia Appendix 1) [37].

Urgent Care and Specialist Referral

The primary care provider, as the patient's first contact and care coordinator, should have the ability to decide whether the patient needs emergency care or an urgent specialist referral [35-37]. Therefore, the primary care provider, aided by some of the responses provided by the patient prior to the consultation, could identify severe or uncontrolled cases of chronic conditions that require immediate and urgent attention. In his or her coordinating role, the primary care provider might promptly refer the patient to a specialist or to a hospital, based on the information provided and a quick face-to-face interaction, if needed.

Prior to the Follow-Up Consultation

Overview

Once a diagnosis has been reached, the primary care provider and patient would ideally establish a plan of action to treat and manage the condition. Key tasks of a primary care follow-up consultation include the control or monitoring of the condition [23,35-37], including identifying the potential occurrence of related additional health issues [35,36], as well as the coordination with relevant specialists regarding further workup [10,23]. A DMIA system can support these tasks by checking on the level of control and by monitoring the progress of the condition. It can explore additional health problems that may have arisen as a consequence of the condition and/or by adding additional useful information, such as information from other health care providers and/or lab tests, prior to the patient–health care provider follow-up consultation. Also, this can be done remotely and, hence, more frequently than the face-to-face follow-up consultation.

Control and Monitoring

A set of lifestyle and general questions (ie, exercise, sleep duration, diet, etc) will help check a patient's compliance to lifestyle changes. For diabetes, these may evaluate changes in diet and exercise patterns, among others (see Multimedia Appendix 1). In the case of depression, the questions can relate

to alcohol and drug use, sleeping patterns, and the situation at home and work, among others (see Multimedia Appendix 1).

The monitoring and control of chronic conditions highlight the importance of the patient becoming a partner in the management of their condition [36,38]. Relatedly, the system could also evaluate environmental and social conditions conducive to improved disease control. For example, in diabetes, a DMIA can explore accessibility to healthy food options at home and work and time available to exercise, among others. Regarding depression, cultural stigma regarding mental health and family support could be assessed [37].

After going through these sets of more general monitoring questions, the system can delve deeper into more clinically relevant questions, in order to monitor disease progression and check for medication adherence. It could ask questions about blood glucose levels (for diabetes), recurrent mood and depressed symptoms (for depression), and whether the patient has been taking medications properly.

Checking for Additional Health Issues

In addition to the chronic condition, the patient could have other associated health problems. The patient may not be aware that these other problems may be connected to the underlying chronic condition. In such cases, the DMIA system may be able to check for this prior to the follow-up consultation. In diabetes, it may assess for micro- or macrovascular complications (eg, kidney problems, peripheral vascular disease, foot ulcers, neuropathy, enteropathy, and ophthalmopathy) [35,36]. In depression, it can check for neurovegetative symptoms, difficulties concentrating, and suicidal ideations, among others [37,39].

Coordination

As mentioned above, given the role of coordinator that is taken on by the primary care provider, a DMIA system should ask questions related to the patient seeing other health care professionals, which may provide relevant additional information for the primary care provider. Often, when the patient is referred to specialists or other health care providers for additional treatment and tests, among other items, the information is not always transferred back to the primary care level [40]. Thus, these systems can check whether the consultations with other health care providers have been occurring and if they have been successful. For diabetes, this may relate to visits to the endocrinologist, ophthalmologist, nephrologist, among others [35,36]; for depression, this may relate to visits to the psychotherapist and psychiatrist [32,37,39]. Additionally, the systems may check with the patients whether there have been laboratory or other tests performed, and the primary care office could check with those external sites if the results have not made it back to the primary care level.

Implications of Incorporating DMIA Systems Into the Primary Care Consultation

We described the role that a DMIA system could potentially play in managing consultations on diabetes and depression within the primary care context. The anticipated benefits of incorporating this technology into the primary care consultations for these chronic conditions should mirror those described in the literature: namely, a more comprehensive set of patient information items, better use of face-to-face consultation time, better-quality and more up-to-date data, and more frequent interactions with the patient, among others.

All the information gathered outside of the face-to-face consultation means that the actual time in consultation is better spent. Conversation can be streamlined to address the needs of the patient, such as building rapport, providing education, or responding to concerns, and to improve communication and patient understanding, which ultimately leads to increased patient satisfaction [41]. Relatedly, better-quality data prior to the consultation, as well as support for additional data coming from other levels of care or lab tests, allow for early detection of a chronic condition, earlier and enhanced treatment at the primary care level, better monitoring, and early management of possible complications, all of which translates into better care. In addition, as the DMIA communicates with the patient before and in between consultations, it improves the continuity of care and it gives the patient the feeling of being better cared for, as communication with the primary care provider via the DMIA system occurs more frequently.

Additionally, DMIA systems could improve the treatment and management of chronic conditions, as seen with diabetes and depression, without too much additional effort once the system has been set up. As mentioned above, chronic condition management usually follows a pre-established pattern. Therefore, it would not be difficult to develop a rules-based system to check case finding and diagnosis and appropriate management strategies. Then, once developed, the system could be deployed and repeated throughout, using the same branching logic and platform. Moreover, artificial intelligence could be leveraged and introduced into a DMIA system, which would potentially improve questioning algorithms and language, among other improvements.

Important Implementation Considerations

For such a system to work well, there are several considerations that need to be taken into account. As presented earlier, there are some disadvantages that need to be bypassed or considered when implementing DMIA systems. Some patients may not be able to read the materials in digital form or may not be digitally literate; about 10% of the population chooses not to complete their histories on computers [3], although this figure may vary depending on how technologically acquainted people from different populations are. General practitioners sometimes try to confirm all the answers to the questions, thereby duplicating the work of the computer, which obviously impacts efficiency [3], although as technology and program design improve, there should be more confidence from these physicians in the results obtained by these programs. Additionally, there may be patient data entry and potentially added work for the primary care provider, as more issues might be highlighted and in need of attention. Moreover, bringing a DMIA system into everyday clinical practice requires interdisciplinary collaboration with a team of specialists and primary care physicians, patient advocates, computer scientists, big data analytics, and experience design [2].

Another consideration relates to the language to be used. The ideal DMIA system should provide a *human-like* interaction.

For example, for a consultation ahead of the follow-up, it should phrase the questions in such a way that the patient feels as if they were continuing a previous conversation. Technologies today allow for human-like conversations in the form of a chatbot—also known as a machine conversation system, virtual agent, dialogue system, conversational user interface, and chatterbot—where a computer program interacts with users by using natural language precisely enough to simulate human conversation [42,43]. As such, the system may ask "Have you been able to adequately manage the [chronic condition] we discovered last week? Is there something you'd need extra help with?"

Another essential consideration relates to the factors that influence the adoption of eHealth innovations in health care in general, and primary care in particular. On the one hand, there are factors related to the adaptability, flexibility, and cost of the technology to be implemented [28]. DMIA systems need to be sufficiently flexible and affordable so that they easily enter and adapt to different primary care contexts and into already-existing working practices and systems. On the other hand, the environment in which the technology will be implemented needs to be ready for such a transformation and must include supportive policies, incentives, and leadership, as well as the appropriate resources for implementation [28,29]. In addition, a primary care provider's knowledge and experience with digital technologies needs to be taken into account, as well as their willingness to incorporate a new technology into their everyday practice [28,29]. Only the proper alignment of these factors will enable a technology such as a DMIA system to be successfully implemented and, ultimately, improve care.

Finally, by transforming the primary care consultation in the way we describe here, a DMIA system would standardize the consultation and follow-up of chronic conditions and improve clinical care. We foresee that, given the described benefits, it should become the norm and a regular practice accompanying the primary care consultation if implemented in the appropriate way, by addressing the challenges and factors mentioned above. Moreover, it can become part of a suite of digital health technologies, amplifying its impact. For instance, by directly connecting a DMIA system to an electronic medical record (EMR) system and/or to a clinical decision support system (CDSS), it can have synergistic effects that may transform the way health care is provided in the future, such as providing rapid and remote access to care, improved triage, and more accurate diagnosis, just to name a few benefits. The answers provided to a DMIA system, which are automatically stored in a digital format, can populate whichever data fields are needed for an automated CDSS, providing support across a wide variety of clinical fields and issues [2,3,20]. In addition, clinical data can be standardized across all patients in contact with a health care system and that data can be included in a database to feed and support clinical and population health research [2].

Conclusions

A DMIA system could enhance the primary care consultation and facilitate the management of diabetes and depression, and possibly other chronic conditions, which would hopefully make an impact in primary care. Earlier case finding and a more accurate diagnosis, due to more and better-quality data, paired with improved monitoring of disease progress, should improve the quality of care and keep the management of chronic conditions at the primary care level. A somewhat simple, easily scalable technology could go a long way to improve the health of the millions of people affected with chronic conditions, depending on the context and how successful its implementation is, especially if working in conjunction with already-established health technologies, such as EMRs and CDSSs.

Acknowledgments

This work was supported by the Centre for Population Health Sciences at the Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore. The funder did not have direct involvement in the research, including manuscript preparation, apart from employing some of the authors. The research presented in this article is solely the responsibility of the authors and does not reflect the views of the funder.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Digital medical interview assistant questions for the cases of diabetes and depression.
[DOCX File , 16 KB - jmir_v22i8e18109_app1.docx]

References

1. Pappas Y, Anandan C, Liu J, Car J, Sheikh A, Majeed A. Computer-assisted history-taking systems (CAHTS) in health care: Benefits, risks and potential for further development. Inform Prim Care 2011;19(3):155-160 [FREE Full text] [doi: 10.14236/jhi.v19i3.808] [Medline: 22688224]
2. Zakim D. Development and significance of automated history-taking software for clinical medicine, clinical research and basic medical science. J Intern Med 2016 Sep;280(3):287-299 [FREE Full text] [doi: 10.1111/joim.12509] [Medline: 27071980]

3. Bachman J. Improving care with an automated patient history. Fam Pract Manag 2007;14(7):39-43 [FREE Full text] [Medline: 17696057]
4. Macinko J, Starfield B, Shi L. The contribution of primary care systems to health outcomes within Organization for Economic Cooperation and Development (OECD) countries, 1970-1998. Health Serv Res 2003 Jun;38(3):831-865 [FREE Full text] [doi: 10.1111/1475-6773.00149] [Medline: 12822915]
5. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q 2005;83(3):457-502 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00409.x] [Medline: 16202000]
6. Institute of Medicine (IOM). In: Donaldson MS, Yordy KD, Lohr KN, Vanselow NA, editors. Primary Care: America's Health in a New Era. Washington, DC: National Academy Press; 1996.
7. Finley CR, Chan DS, Garrison S, Korownyk C, Kolber MR, Campbell S, et al. What are the most common conditions in primary care? Systematic review. Can Fam Physician 2018 Nov;64(11):832-840 [FREE Full text] [Medline: 30429181]
8. Diabetes. World Health Organization. 2020 Jun 08. URL: https://www.who.int/news-room/fact-sheets/detail/diabetes [accessed 2019-12-09]
9. Depression. World Health Organization. 2020 Jan 30. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2019-12-09]
10. Rakel RE, Rakel DP. The family physician. In: Textbook of Family Medicine. 9th edition. Philadelphia, PA: Saunders; 2015.
11. Adler-Waxman A. This is the biggest challenge to our health. World Economic Forum. 2017 Dec 07. URL: https://www.weforum.org/agenda/2017/12/healthcare-future-multiple-chronic-disease-ncd/ [accessed 2020-12-18]
12. Ferrari R. Writing narrative style literature reviews. Med Writ 2015 Dec 23;24(4):230-235. [doi: 10.1179/2047480615z.000000000329]
13. Wei I, Pappas Y, Car J, Sheikh A, Majeed A. Computer-assisted versus oral-and-written dietary history taking for diabetes mellitus. Cochrane Database Syst Rev 2011 Dec 07(12):CD008488 [FREE Full text] [doi: 10.1002/14651858.CD008488.pub2] [Medline: 22161430]
14. Pappas Y, Wei I, Car J, Majeed A, Sheikh A. Computer-assisted versus oral-and-written family history taking for identifying people with elevated risk of type 2 diabetes mellitus. Cochrane Database Syst Rev 2011 Dec 07(12):CD008489. [doi: 10.1002/14651858.CD008489.pub2] [Medline: 22161431]
15. Ellsworth MA, Aakre CA, Dziadzko M, Peters SG, Pickering BW, Herasevich V. Early computerization of patient care at Mayo Clinic. Mayo Clin Proc 2016 Jul;91(7):e93-e101. [doi: 10.1016/j.mayocp.2016.04.001] [Medline: 27236428]
16. Fincher D, VanderEnde K, Colbert K, Houry D, Smith LS, Yount KM. Effect of face-to-face interview versus computer-assisted self-interview on disclosure of intimate partner violence among African American women in WIC clinics. J Interpers Violence 2015 Mar;30(5):818-838. [doi: 10.1177/0886260514536280] [Medline: 24923890]
17. Brown JL, Swartzendruber A, DiClemente RJ. Application of audio computer-assisted self-interviews to collect self-reported health data: An overview. Caries Res 2013;47 Suppl 1:40-45 [FREE Full text] [doi: 10.1159/000351827] [Medline: 24107606]
18. Slack WV, Kowaloff HB, Davis RB, Delbanco T, Locke SE, Bleich HL. Test-retest reliability in a computer-based medical history. J Am Med Inform Assoc 2011;18(1):73-76 [FREE Full text] [doi: 10.1136/jamia.2010.005983] [Medline: 21113077]
19. Bachman JW. The patient-computer interview: A neglected tool that can aid the clinician. Mayo Clin Proc 2003 Jan;78(1):67-78. [doi: 10.4065/78.1.67] [Medline: 12528879]
20. Zakim D, Schwab M. Data collection as a barrier to personalized medicine. Trends Pharmacol Sci 2015 Mar;36(2):68-71. [doi: 10.1016/j.tips.2014.11.002] [Medline: 25479798]
21. Zakim D, Fritz C, Braun N, Fritz P, Alscher MD. Computerized history-taking as a tool to manage dyslipidemia. Vasc Health Risk Manag 2010 Nov 12;6:1039-1046 [FREE Full text] [doi: 10.2147/VHRM.S14302] [Medline: 21127700]
22. Nicholson E, Murphy T, Larkin P, Normand C, Guerin S. Protocol for a thematic synthesis to identify key themes and messages from a palliative care research network. BMC Res Notes 2016 Oct 21;9(1):478 [FREE Full text] [doi: 10.1186/s13104-016-2282-1] [Medline: 27769317]
23. Goroll AH, Mulley Jr AG. Primary Care Medicine: Office Evaluation and Management of the Adult Patient. 7th edition. Philadelphia, PA: Wolters Kluwer; 2014.
24. Irving G, Neves AL, Dambha-Miller H, Oishi A, Tagashira H, Verho A, et al. International variations in primary care physician consultation time: A systematic review of 67 countries. BMJ Open 2017 Nov 08;7(10):e017902 [FREE Full text] [doi: 10.1136/bmjopen-2017-017902] [Medline: 29118053]
25. Elmore N, Burt J, Abel G, Maratos FA, Montague J, Campbell J, et al. Investigating the relationship between consultation length and patient experience: A cross-sectional study in primary care. Br J Gen Pract 2016 Dec;66(653):e896-e903 [FREE Full text] [doi: 10.3399/bjgp16X687733] [Medline: 27777231]
26. Zakim D, Braun N, Fritz P, Alscher MD. Underutilization of information and knowledge in everyday medical practice: Evaluation of a computer-based solution. BMC Med Inform Decis Mak 2008 Nov 05;8:50 [FREE Full text] [doi: 10.1186/1472-6947-8-50] [Medline: 18983684]
27. Mazumder SI. Towards Formalized Models of Patients' History Taking [master's thesis]. Stockholm, Sweden: Karolinska Institutet; 2016 May. URL: https://ki.se/sites/default/files/migrate/md._shehabul_islam_towards.pdf [accessed 2019-12-09]

28. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: A systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
29. Swinkels ICS, Huygens MWJ, Schoenmakers TM, Oude Nijeweme-D'Hollosy W, van Velsen L, Vermeulen J, et al. Lessons learned from a living lab on the broad adoption of eHealth in primary health care. J Med Internet Res 2018 Mar 29;20(3):e83 [FREE Full text] [doi: 10.2196/jmir.9110] [Medline: 29599108]
30. Shachak A, Reis S. The impact of electronic medical records on patient-doctor communication during consultation: A narrative literature review. J Eval Clin Pract 2009 Aug;15(4):641-649. [doi: 10.1111/j.1365-2753.2008.01065.x] [Medline: 19522722]
31. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: Harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
32. Ng CWM, How CH, Ng YP. Managing depression in primary care. Singapore Med J 2017 Aug;58(8):459-466 [FREE Full text] [doi: 10.11622/smedj.2017080] [Medline: 28848991]
33. Gilbody SM, Whitty PM, Grimshaw JM, Thomas RE. Improving the detection and management of depression in primary care. Qual Saf Health Care 2003 Apr;12(2):149-155 [FREE Full text] [doi: 10.1136/qhc.12.2.149] [Medline: 12679514]
34. Smith SM. Newly diagnosed type 2 diabetes mellitus. BMJ 2003 Jun 21;326(7403):1371 [FREE Full text] [doi: 10.1136/bmj.326.7403.1371] [Medline: 12816825]
35. Rakel RE, Rakel DP. Diabetes mellitus. In: Textbook of Family Medicine. 9th edition. Philadelphia, PA: Saunders; 2015.
36. Goroll AH, Mulley Jr AG. Endocrinologic problems. In: Primary Care Medicine: Office Evaluation and Management of the Adult Patient. 7th edition. Philadelphia, PA: Wolters Kluwer; 2014.
37. Goroll AH, Mulley Jr AG. Psychiatric and behavioral problems. In: Primary Care Medicine: Office Evaluation and Management of the Adult Patient. 7th edition. Philadelphia, PA: Wolters Kluwer; 2014.
38. Ukoumunne OC, Vaidya B, Frost J, Anderson R, Argyle C, Daly M, et al. A preconsultation web-based tool to generate an agenda for discussion in diabetes outpatient clinics to improve patient outcomes (DIAT): A feasibility study. BMJ Open 2017 Mar 07;7(3):e013519 [FREE Full text] [doi: 10.1136/bmjopen-2016-013519] [Medline: 28270389]
39. Rakel RE, Rakel DP. Anxiety and depression. In: Textbook of Family Medicine. 9th edition. Philadelphia, PA: Saunders; 2015.
40. Gandhi TK, Sittig DF, Franklin M, Sussman AJ, Fairchild DG, Bates DW. Communication breakdown in the outpatient referral process. J Gen Intern Med 2000 Sep;15(9):626-631 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.91119.x] [Medline: 11029676]
41. Fitri Surbakti E, Sari K. The relationship between consultation length and patient satisfaction: A systematic review. In: Proceedings of the 2nd International Conference on Hospital Administration (The 2nd ICHA). 2018 Dec 05 Presented at: 2nd International Conference on Hospital Administration (The 2nd ICHA); November 1-2, 2017; South Jakarta City, Special Capital Region of Jakarta, Indonesia p. 41-49 URL: https://knepublishing.com/index.php/KnE-Life/article/view/3556/7445 [doi: 10.18502/kls.v4i9.3556]
42. Denecke K, Hochreutener SL, Pöpel A, May R. Self-anamnesis with a conversational user interface: Concept and usability study. Methods Inf Med 2018 Nov;57(5-06):243-252. [doi: 10.1055/s-0038-1675822] [Medline: 30875703]
43. Peiris D, Gregor P, Alm N. The effects of simulating human conversational style in a computer-based interview. Interact Comput 2000 Jul;12(6):635-650. [doi: 10.1016/s0953-5438(00)00028-x]

Abbreviations

CAHT: computer-assisted history taking
CDSS: clinical decision support system
DMIA: digital medical interview assistant
EMR: electronic medical record
MEDLINE: Medical Literature Analysis and Retrieval System Online

Edited by G Eysenbach; submitted 04.02.20; peer-reviewed by D Zakim, M Nomali; comments to author 30.03.20; revised version received 10.04.20; accepted 27.04.20; published 28.08.20.

Please cite as:
Jimenez G, Tyagi S, Osman T, Spinazze P, van der Kleij R, Chavannes NH, Car J
Improving the Primary Care Consultation for Diabetes and Depression Through Digital Medical Interview Assistant Systems: Narrative Review
J Med Internet Res 2020;22(8):e18109
URL: http://www.jmir.org/2020/8/e18109/
doi:10.2196/18109
PMID:32663144

Viewpoint

Virtual Consultations and the Role of Technology During the COVID-19 Pandemic for People With Type 2 Diabetes: The UK Perspective

Lauren M Quinn[1], MBChB, BMedSci; Melanie J Davies[1,2], MBChB, MD, FRCP, FRCGP, FMedSci; Michelle Hadjiconstantinou[1,2], PhD, CPsychol

[1]Leicester Diabetes Centre, University of Leicester, Leicester, United Kingdom

[2]Leicester Diabetes Research Centre, University of Leicester, Leicester, United Kingdom

Corresponding Author:
Lauren M Quinn, MBChB, BMedSci
Leicester Diabetes Centre
University of Leicester
Leicester General Hospital
Gwendolen Rd
Leicester, LE5 4PW
United Kingdom
Phone: 44 0116 258 4322
Email: lauren.quinn3@nhs.net

Abstract

The coronavirus disease (COVID-19) pandemic has presented unique challenges for people with diabetes, in addition to their high-risk stratification for infection. Supporting people with diabetes to self-care has been critical to reduce their risk of severe infection. This global pandemic has presented an opportunity to digitalize diabetes care and rapidly implement virtual diabetes clinics, with the aim of optimizing diabetes management and well-being, while keeping patients safe. We performed a rapid review of the literature to evaluate the feasibility and effectiveness of virtual clinics in diabetes care before and during the COVID-19 pandemic and have combined these findings with our own reflections in practice. We identified examples demonstrating safety and feasibility of virtual diabetes clinics, which aligns with our own clinical experience during the pandemic. The advantages of virtual clinics include reduced treatment burden, improved therapeutic alliances, societal and psychological benefits, and in our experience, innovative solutions to overcome the challenges presented by the transition from in-person to virtual care. We have provided three infographics to illustrate lessons learned and key recommendations, including steps to establish a virtual diabetes clinic, a checklist guide for health care professionals conducting virtual clinics, and a patient guide for making the most out of the virtual clinic. It is important to continue adapting to this pandemic and to make technology a sustainable option for the future of diabetes care.

(J Med Internet Res 2020;22(8):e21609) doi:10.2196/21609

KEYWORDS

diabetes; virtual clinic; technology; COVID-19; United Kingdom; pandemic; feasibility; effective; telehealth

Introduction

We are in the midst of the coronavirus disease (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which has resulted in thousands of deaths worldwide [1,2]. Increased age and underlying health conditions, including diabetes, cardiovascular disease, obesity, and hypertension, significantly increase the risk of COVID-19 infection [3,4]. Similarly, disease severity may be worsened, and deaths are overrepresented in people with diabetes [4,5]. Evidence shows poor glycemic control is both associated with and a consequence of COVID-19 infection, the latter demonstrated in older persons with type 2 diabetes [3,5,6]. More recently, evidence has shown that Black, Asian, and minority ethnic groups are more severely affected, with higher death rates observed from COVID-19 infection in this population; high prevalence of diabetes and comorbidities in this subgroup likely contributes to this increased risk [7]. Despite the challenges for people with diabetes, the COVID-19 pandemic has presented a valuable opportunity to digitalize diabetes care. Given the importance of maintaining and improving well-being

and glycemic control during this time, evaluation of novel methods to support self-management remotely is critically important.

The aims of this paper are (1) to explore the evidence for the role of telemedicine to support people with diabetes during the COVID-19 pandemic and beyond; (2) to outline the benefits and challenges presented by virtual diabetes care; (3) to present our experience of virtual consultations in clinical settings during the COVID-19 pandemic; and (4) to share lessons learned to assist researchers, clinicians, and people with diabetes when integrating technology in diabetes care.

Impact of Social Distancing/Shielding on Diabetes Care and Well-Being

With diabetes being classed as a high-risk group by the government, it is important that people with diabetes take care of their health now more than ever. People with diabetes are advised to practice social distancing (eg, working from home or self-isolating), adhere to national recommendations on frequent handwashing, and abstain from nonessential travel to avoid infectious contacts [8]. In some cases, extremely vulnerable people with diabetes are advised to undertake "shielding." Additional guidance, specifically for people with diabetes, focuses on self-management strategies, which help to boost innate immunity for primary prevention [8,9]. The American Association of Diabetes Educators recommends the following seven self-care activities: keeping physically active, healthy diet, following medication regimen, blood glucose monitoring and problem solving, reducing risk of complications, and self-empathy; all of these are endorsed for people with diabetes to reduce their risk and severity of COVID-19 [10].

Interestingly, for people with diabetes, their engagement with these self-care activities has been highly variable during the pandemic, with some being able to focus more time on their diabetes management and undertake increasing amounts of self-care. For example, Bonora et al [11] showed in a small study that glycemic control improved for people with type 1 diabetes during lockdown who self-isolated compared to those who continued working [11]. However, for others, the significant change in lifestyle presented by lockdown has been detrimental to their health and well-being. Barriers specifically for people with diabetes have included difficulty accessing healthy foods because of restricted shopping and bulk buying; inability to access medications; restricting physical activity to once per day in the local area; and being unable to attend face-to-face appointments with their diabetes care providers [8,9]. With this in mind, diabetes health care professionals (HCPs) have valid concerns that glycemic control, quality of life, self-management, and well-being can be significantly jeopardized during social distancing and shielding, posing considerable risks to people with diabetes, both in the short and long term. Lockdown, social distancing, shielding, and the abundance of misinformation in the media also present additional stressors, which may further exacerbate underlying depression and anxiety [12], conditions that are already highly prevalent in the diabetes population [13].

Digitalization of Diabetes Care: Benefits and Challenges

Digital Consultations

Technology has been increasingly integrated into diabetes education and care in modern times, for example, through apps, computer-based or web-based education, and telemedicine [14]. The NHS (National Health Service) long-term plan set out to increase digitalization within NHS programs [15,16]. This included roll-out of virtual or non–face-to-face clinics, with the aim of reducing face-to-face appointments by one third over the next 5 years [16]. This non–face-to-face activity may be synchronous or asynchronous, meaning a direct or indirect line of communication with an HCP, respectively [17,18]. Synchronous activity would be a video or telephone consultation, whereas asynchronous may involve monitoring an email or tracking system and responding to patients' questions through these platforms. The benefits of non–face-to-face appointments are multifactorial and include the opportunity for better care and more connected patient care pathways, as well as cost savings and a reduced environmental impact.

Studies evaluating virtual clinics prior to the COVID-19 pandemic have demonstrated feasibility, accessibility, safety, and effectiveness comparable to in-person consultations [19]. A multicenter mixed methods study evaluating video consultations in diabetes care showed that video consultations were shorter in duration and people with diabetes did relatively more talking than the HCP [20,21]. Although from a management perspective, video consultations were favored, there were significant barriers to uptake from the teams implementing them because of the significant changes introduced to their usual way of working and the care processes, systems, and pathways [21]. A Cochrane review analyzed 21 low-to-high-quality studies comparing telemedicine to usual care in people with diabetes and found that improvement in glycemic control was variable, but low-density lipoprotein (LDL) levels and blood pressure were more effectively lowered by telemedicine approaches compared to usual care [22]. Further studies have evaluated virtual clinics for diabetes care in type 1 and type 2 diabetes and have reported improved biochemical parameters, including glycated hemoglobin (HbA_{1c}) [23]. However, many of these studies combined interventions, such as synchronous and asynchronous programs, rendering it difficult to delineate the efficacy of the individual interventions [23]. Studies focusing on synchronous video consultations have been of short duration with a limited sample size [24]. Additionally, how these findings relate to virtual clinics being implemented during and due to a global pandemic must be considered.

During the COVID-19 pandemic, telemedicine has been widely adopted globally in order to reduce exposure and need for people with diabetes to come into a hospital, while maintaining care standards for people with chronic conditions. For example, virtual consultations have been implemented to triage patients suspected of COVID-19 in primary care and to initiate a hospital visit [25-27]. There have been other examples of successful

virtual care adoption for people with newly diagnosed type 1 diabetes [28], and for people with long-standing type 1 diabetes. A small Italian study showed that for people with type 1 diabetes who were not working during the pandemic and using continuous glucose monitoring, their time in range significantly improved from 54% to 65%, and this was attributed to decreased hyperglycemia [11]. Blood glucose variability and average glucose readings from continuous glucose monitoring also significantly improved in this study. In contrast to patients who continued working, there was no difference in glycemic control [11]. Early reports suggest that virtual clinics are feasible, with some centers increasing virtual clinic consultations from 1% of all consultations prior to COVID-19 to 70% afterwards, and it is technology that has made this rapid transformation possible [29-31].

A recent linguistic ethnographic study has shown that video consultations among people with a long-term condition and their clinician was found to be effective [26]. When patients experienced technical or operational issues with their video equipment or internet connection, they generally found a solution to resolve the problem. Familiarity and experience with technology helped in situations like these. Technical interruptions and delays of connectivity either on patient or clinician device were clearly evident in the study; nevertheless, remote physical examinations were conducted, allowing the patient and/or carer to take an active role in the consultation [26].

Benefits of Virtual Clinics

Patient Safety

In line with social distancing advice and minimizing risk for people with diabetes, virtual clinics bring specialist care to the home of the patient. The virtual clinic is the ideal solution to enable patients to access specialist care but without unnecessarily exposing themselves to a high-risk environment, or come in close contact with staff or other patients. This is particularly pertinent given the new data supporting the number of COVID-19 infections developed while being in hospital [32].

Reduced Burden of Treatment: Improved Accessibility and Overall Experience

Virtual clinics increase accessibility to specialist diabetes care. For example, those shielding can still attend a virtual appointment, those working can more easily take the time out from their day (eg, over a coffee break), and those who would normally be unable to travel would no longer need to. In our experience, we have found that patients who regularly did not attend appointments are now attending their virtual clinic appointments. The virtual clinic may also provide an improved patient experience because people with diabetes are not required to spend additional time traveling to the hospital for regular appointments, which can significantly reduce the burden of treatment that some may experience with their diabetes care. The virtual clinics may provide a solution—video or phone calls may provide a relaxed environment for some people with diabetes, as they are within their own familiar surroundings, which may create a more relaxed atmosphere for their virtual appointment. From our clinic team experience of using virtual consultations with people with diabetes, it is evident that some patients feel comfortable showing the HCP their house environment and exercise regimen, and introduce them to their family members, within professional boundaries. This offers a unique insight into environmental or interpersonal factors, which may influence the self-care activities of people with diabetes. Long-term, these factors may be considered when discussing the care plan and treatment goals of people with diabetes. The opportunities presented in virtual clinics for delivering increasingly person-centered and individualized care cannot be underestimated.

Improved Therapeutic Alliance and Consultation Dynamics

Virtual consultations could significantly shift the locus of control of people with diabetes, which would complement the philosophy around the importance of self-management and self-care in diabetes. Virtual consultations could provide an opportunity for shared care between people with diabetes and the HCP. Multiple members of the diabetes multidisciplinary team (MDT) may come together for the video consultation, to make collaborative decisions with patients. From our experience, this results in a more effective consultation. People with diabetes benefit from MDT input, with reduced time commitment for patients and professionals alike. Physical examination (eg, conducting a diabetes foot examination) can also be performed. This is particularly important in the context of diabetes foot disease, where people with diabetes can be triaged in virtual clinic to determine the need for hospital admission based on the video examination findings. Technical barriers would be expected when setting up virtual consultations; thus, we have shared a checklist that we developed for setting up a virtual clinic (Figure 1).

Figure 1. Checklist for setting up a virtual clinic. GDPR: General Data Protection Regulation; BP: blood pressure; GP: general practitioner.

Societal and Psychosocial Benefits

Virtual clinics reduce the need for travel to and from hospitals, which is a significant benefit since 20% of traffic in the United Kingdom is attributed to health care–related travel [15,16]. Additionally, virtual clinics mean less time is missed from work, which can be a recurrent issue for people with diabetes, so this reduction in treatment burden is highly advantageous.

With social isolation becoming more prevalent than ever during the current climate, it is imperative to acknowledge the positive impact that remote consultations may have on individuals who are required to practice social distancing and shielding. Virtual consultations, whether these are conducted by telephone or video, provide people with diabetes the opportunity to virtually connect with their HCP, mitigating the psychological effects of social isolation.

Overcoming the Challenges of Virtual Care

Elements of the routine diabetes clinic (eg, checking blood pressure, sampling the urine, HbA_{1c}, spot checks, and physical examination) are not possible in the same way in a virtual consultation. However, given the importance of these methods to screen for diabetes complications, alternative strategies can be developed to make them possible. For example, people with diabetes can have their own blood pressure monitor. This may lead to more accurate readings, with the absence of white coat hypertension. Also, urine samples can be delivered locally to general practices and examination can still be performed via video consultation. For example, diabetes foot disease can be screened for and people with diabetes can be triaged to determine need for further assessment.

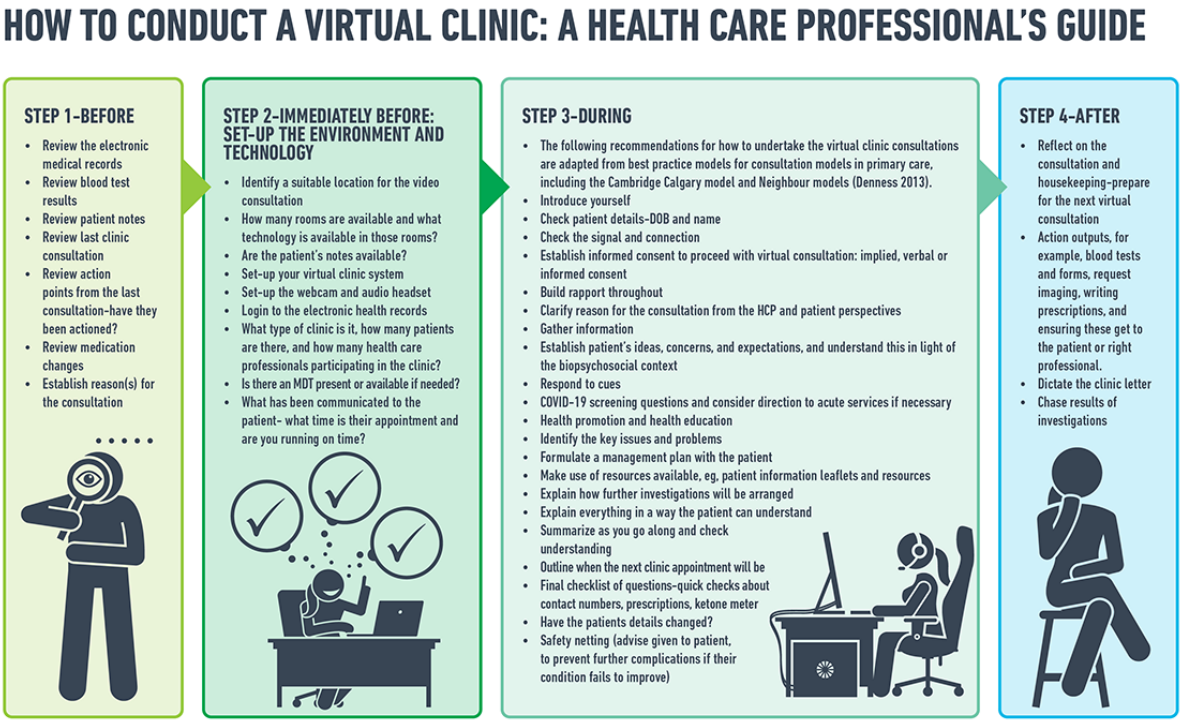
Patient Preference

People with diabetes can choose whether to have a telephone or video consultation, to reflect variability in access to technology and systems for video recording and to ensure access to care services. However, virtual care is a significant change from the usual hospital attendance for their diabetes care, thus, some people with diabetes would still prefer a face-to-face consultation. However, in light of the COVID-19 pandemic, people with diabetes have readily transitioned to virtual methods, and it is hoped that through familiarization with these new consultation methods, this may increase uptake of virtual care in the longer term. There are also concerns that people with diabetes may find it difficult to build rapport with their HCP in a virtual clinic, but in our experience, the opposite is true. People with diabetes have been far more relaxed in the virtual clinic and able to speak more openly with their diabetes team, which is essential to improve their diabetes management. But this is dependent on high-quality video image and sound, highlighting the essential role of technology to optimize communication between people with diabetes and their HCP.

Engaging the Multidisciplinary Team

Engaging professionals and the wider team with virtual care can be a challenge in itself. This is a significant change for the diabetes team as well as the patient, and new processes and pathways have needed to develop rapidly to manage the care of people with diabetes remotely (Figure 2 [33]). For example, the ways in which clinic outcomes are actioned have changed and the wider clinical and nonclinical teams have needed to collaborate to generate these new systems. It is essential that the whole team is committed to the virtual care approach, because this ensures the delivery of safe, high-quality care. It is anticipated that with increased familiarity and established systems, HCPs will be more open to adopt virtual care into their routine practice and overcome the natural aversion to change of how it has always been.

Figure 2. How to conduct a virtual clinic: a health care professional's guide. MDT: multidisciplinary team; DOB: date of birth; HCP: health care professional; COVID-19: coronavirus disease.



The Future of Virtual Clinics

Anecdotally, in our experience virtual clinics have been feasible and accessible, with high patient satisfaction. Virtual clinical consultations offer a different kind of benefit compared to conventional face-to-face appointments, particularly around convenience, logistics, cost-effectiveness, and clinician-patient dynamic/relationship [26]. However, we must acknowledge the pitfalls of these new modes of communication and the challenges that may lie ahead with the clinical quality and safety of appointments.

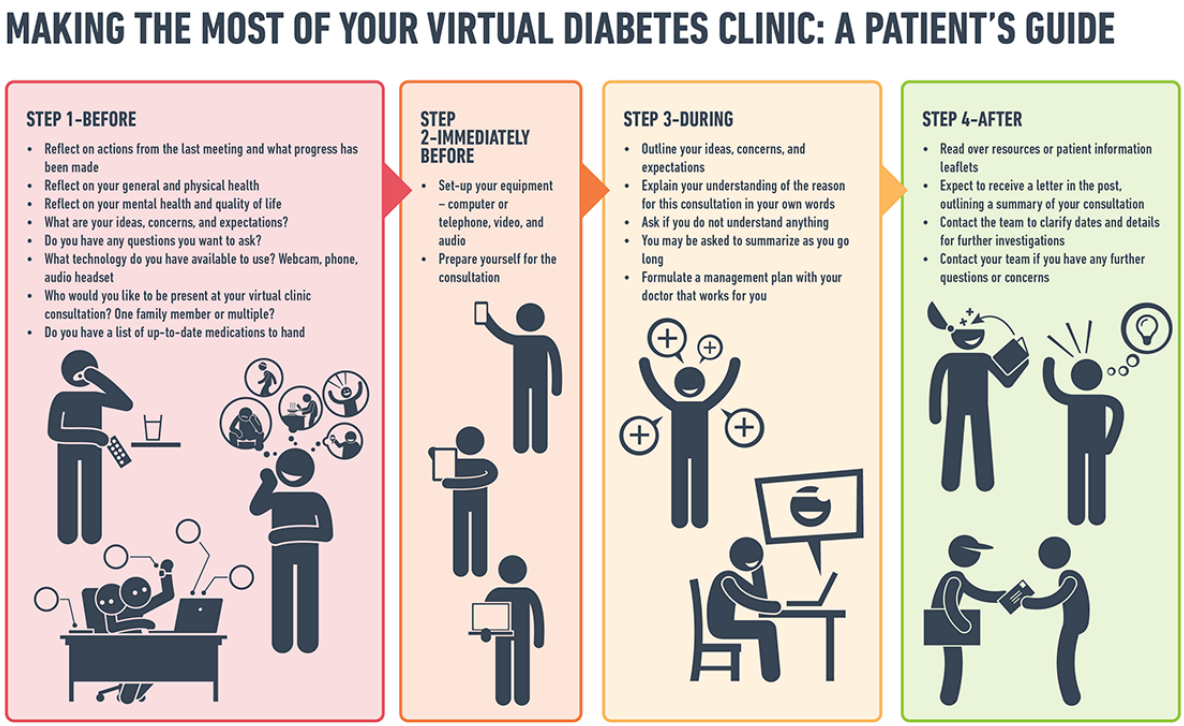
Building on these developments, we are looking to make virtual clinics sustainable for the long term. This is in line with the NHS long-term plan to make better use of data and digital technology in the next 5 years [15,16]. In a time of a pandemic, individualized care is more important than ever and virtual clinics provide a readily accessible solution to facilitate this. Having applied virtual clinics in our setting for the last 2 months,

lessons we have learned include: (a) the importance of integrating multiple members of the MDT into the one virtual consultation; (b) avoiding the checklist approach and instead focusing on an individualized, person-centered consultation; (c) and acknowledging that video consultation may be preferred to telephone because of the additional benefits of human contact, body language, and the opportunity to gain better insight into the lifestyle and livelihood of people with diabetes in order to tailor medical support accordingly. In our experience, virtual clinics may be better suited to individuals with longstanding diabetes and where possible should be performed by a professional with whom they have already built a strong rapport.

Although virtual clinics can be an alternative option, there are key elements that must be considered to make the consultation as efficient as possible. Firstly, even though virtual consultations would require less resources compared to a face-to-face visit, organizational factors prior to the virtual consultation would still be required in order to book and record a clinic appointment. Secondly, logistical and administrative factors must be integrated within the NHS system, a system which for so many years has been based on delivering face-to-face patient services. We are programmed to deliver our outpatient services in a "traditional" way, and therefore would anticipate a colossal challenge adapting this existing pathway to a more digital-focused platform.

By acknowledging the complexity of integrating virtual consultations, we also acknowledge the challenges that may come with technology, in terms of the security and safety of every patient and HCP. Data protection and privacy is of critical importance; the technology, software, and programs used in virtual clinics must be encrypted and adhere to the General Data Protection Regulation and information governance standards to maintain patient confidentiality at all times. We acknowledge that virtual consultations are not for everyone; however, providing options to people enables them to choose an approach that is tailored to their diabetes needs and lifestyle demands, with the aim of reducing the burden of treatment that so many people with diabetes may be experiencing. We aim to not only prepare our patients to attend virtual consultations, by creating a safe environment and respecting their safety (Figure 3), but to also create a digital platform that would integrate within the current NHS system.

Figure 3. Making the most of your virtual diabetes clinic: a patient's guide.



Conclusion

Virtual consultations may become a necessity following this pandemic. The current system pressures due to COVID-19 have led to numerous challenges to the delivery of routine diabetes care and education. Despite the relative lack of data to support virtual care, in the face of adversity, these virtual measures have been imperative to maintain a line of communication with people with diabetes and to support self-management and self-care remotely. With the right infrastructure and systems in place, technology is the key to evolution in diabetes care, and virtual consultations can be effectively embedded into routine diabetes care at the national level in the United Kingdom. At present,

virtual clinics may be an ideal platform to reduce social isolation, encourage self-management remotely and in a less intrusive manner, and reduce burden of treatment. The COVID-19 outbreak will shift the culture of health care across the world and the way we interact within clinical settings will gradually change to ensure that care can be delivered within social distancing rules

Authors' Contributions

LMQ, MJD, and MH made substantial contributions to conception and design, drafted the article and revised it critically for important intellectual content, and approved the version to be published.

Conflicts of Interest

None declared.

References

1. Novel Coronavirus – China. World Health Organization. 2020. URL: https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/ [accessed 2020-07-19]
2. Staying At Home And Away From Others (Social Distancing). Gov.UK. 2020. URL: https://www.gov.uk/government/publications/full-guidance-on-staying-at-home-and-away-from-others [accessed 2020-07-19]
3. Shenoy A, Ismaily M, Bajaj M. Diabetes and covid-19: a global health challenge. BMJ Open Diabetes Res Care 2020 Apr;8(1):e001450 [FREE Full text] [doi: 10.1136/bmjdrc-2020-001450] [Medline: 32345580]
4. Bornstein SR, Rubino F, Khunti K, Mingrone G, Hopkins D, Birkenfeld AL, et al. Practical recommendations for the management of diabetes in patients with COVID-19. The Lancet Diabetes & Endocrinology 2020 Jun;8(6):546-550 [FREE Full text] [doi: 10.1016/S2213-8587(20)30152-2] [Medline: 32334646]
5. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
6. Chen G, Li L, Chen G, Li L. Blood glucose levels in elderly subjects with type 2 diabetes during COVID-19 outbreak: a retrospective study in a single center. medRxiv 2020 [FREE Full text] [doi: 10.1101/2020.03.31.20048579]
7. Ethnic Minority Deaths And Covid-19. The King's Fund. 2020. URL: https://www.kingsfund.org.uk/blog/2020/04/ethnic-minority-deaths-covid-19 [accessed 2020-07-19]
8. Gupta R, Ghosh A, Singh AK, Misra A. Clinical considerations for patients with diabetes in times of COVID-19 epidemic. Diabetes Metab Syndr 2020;14(3):211-212 [FREE Full text] [doi: 10.1016/j.dsx.2020.03.002] [Medline: 32172175]
9. Banerjee M, Chakraborty S, Pal R. Diabetes self-management amid COVID-19 pandemic. Diabetes Metab Syndr 2020;14(4):351-354 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.013] [Medline: 32311652]
10. American Association of Diabetes Educators (AADE) Position Statement. Diabeteseducator.org. 2020. URL: https://www.diabeteseducator.org/docs/default-source/practice/practice-resources/position-statements/aade7-self-care-behaviors-position-statement.pdf?sfvrsn=6 [accessed 2020-07-19]
11. Bonora BM, Boscari F, Avogaro A, Bruttomesso D, Fadini GP. Glycaemic Control Among People with Type 1 Diabetes During Lockdown for the SARS-CoV-2 Outbreak in Italy. Diabetes Ther 2020 May 11;11(6):1-11 [FREE Full text] [doi: 10.1007/s13300-020-00829-7] [Medline: 32395187]
12. Kar P. Covid-19—We must keep faith in our experts. BMJ 2020 Mar 19;368:m1143. [doi: 10.1136/bmj.m1143] [Medline: 32193174]
13. Semenkovich K, Brown ME, Svrakic DM, Lustman PJ. Depression in type 2 diabetes mellitus: prevalence, impact, and treatment. Drugs 2015 Apr;75(6):577-587. [doi: 10.1007/s40265-015-0347-4] [Medline: 25851098]
14. Cahn A, Akirov A, Raz I. Digital health technology and diabetes management. J Diabetes 2018 Jan;10(1):10-17. [doi: 10.1111/1753-0407.12606] [Medline: 28872765]
15. NHS. Chapter 5: Digitally-Enabled Care Will Go Mainstream Across The NHS. Online version of the NHS Long Term Plan. 2020. URL: https://www.longtermplan.nhs.uk/online-version/chapter-5-digitally-enabled-care-will-go-mainstream-across-the-nhs/ [accessed 2020-07-19]
16. Alderwick H, Dixon J. The NHS long term plan. BMJ 2019 Jan 07;364:l84 [FREE Full text] [doi: 10.1136/bmj.l84] [Medline: 30617185]
17. How-To Guide: Non Face-To-Face Clinics. UCLPartners. 2020. URL: https://uclpartners.com/non-face-to-face-clinics-resource/ [accessed 2020-07-19]
18. Bakhai M, Croney L, Waller O, Henshall N, Felstead C. Using Online Consultations In Primary Care: Implementation Toolkit. NHS England. 2020. URL: https://www.england.nhs.uk/wp-content/uploads/2020/01/online-consultations-implementation-toolkit-v1.1-updated.pdf [accessed 2020-07-19]

19. Armstrong N, Hearnshaw H, Powell J, Dale J. Stakeholder perspectives on the development of a virtual clinic for diabetes care: qualitative study. J Med Internet Res 2007 Aug 09;9(3):e23 [FREE Full text] [doi: 10.2196/jmir.9.3.e23] [Medline: 17942385]
20. Greenhalgh T, Vijayaraghavan S, Wherton J, Shaw S, Byrne E, Campbell-Richards D, et al. Virtual online consultations: advantages and limitations (VOCAL) study. BMJ Open 2016 Jan 29;6(1):e009388 [FREE Full text] [doi: 10.1136/bmjopen-2015-009388] [Medline: 26826147]
21. Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P. Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study. Heal Serv Deliv Res 2018;6(21):1-136. [Medline: 29939519]
22. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015 Sep 07;9(9):CD002098 [FREE Full text] [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
23. Dixon RF, Zisser H, Layne JE, Barleen NA, Miller DP, Moloney DP, et al. A Virtual Type 2 Diabetes Clinic Using Continuous Glucose Monitoring and Endocrinology Visits. J Diabetes Sci Technol 2019 Nov 25. [doi: 10.1177/1932296819888662] [Medline: 31762302]
24. Jennings A, Powell J, Armstrong N, Sturt J, Dale J. A virtual clinic for diabetes self-management: pilot study. J Med Internet Res 2009 Mar 30;11(1):e10 [FREE Full text] [doi: 10.2196/jmir.1111] [Medline: 21821504]
25. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]
26. Shaw SE, Seuren LM, Wherton J, Cameron D, A'Court C, Vijayaraghavan S, et al. Video Consultations Between Patients and Clinicians in Diabetes, Cancer, and Heart Failure Services: Linguistic Ethnographic Study of Video-Mediated Interaction. J Med Internet Res 2020 May 11;22(5):e18378 [FREE Full text] [doi: 10.2196/18378] [Medline: 32391799]
27. Using Virtual Consultations In The Fight Against COVID-. The Health Foundation. 2020. URL: https://www.health.org.uk/news-and-comment/newsletter-features/using-virtual-consultations-in-the-fight-against-covid-19 [accessed 2020-07-19]
28. Garg SK, Rodbard D, Hirsch IB, Forlenza GP. Managing New-Onset Type 1 Diabetes During the COVID-19 Pandemic: Challenges and Opportunities. Diabetes Technol Ther 2020 Jun 17;22(6):431-439. [doi: 10.1089/dia.2020.0161] [Medline: 32302499]
29. Jones MS, Goley AL, Alexander BE, Keller SB, Caldwell MM, Buse JB. Inpatient Transition to Virtual Care During COVID-19 Pandemic. Diabetes Technol Ther 2020 Jun 12;22(6):444-448. [doi: 10.1089/dia.2020.0206] [Medline: 32396395]
30. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
31. Javaid M, Haleem A, Vaishya R, Bahl S, Suman R, Vaish A. Industry 4.0 technologies and their applications in fighting COVID-19 pandemic. Diabetes Metab Syndr 2020 Apr 24;14(4):419-422 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.032] [Medline: 32344370]
32. Reducing the Risk of Transmission of COVID-19 in the Hospital Setting. Gov.uk. 2020. URL: https://www.gov.uk/government/publications/wuhan-novel-coronavirus-infection-prevention-and-control/reducing-the-risk-of-transmission-of-covid-19-in-the-hospital-setting [accessed 2020-07-19]
33. Denness C. What are consultation models for? InnovAiT 2013 Sep 06;6(9):592-599. [doi: 10.1177/1755738013475436]

Abbreviations

COVID-19: coronavirus disease
HbA_{1c}: glycated hemoglobin
HCP: health care professional
LDL: low-density lipoprotein
MDT: multidisciplinary team
NHS: National Health Service
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 22.06.20; peer-reviewed by D Rodbard, K Fitzner; comments to author 15.07.20; revised version received 19.07.20; accepted 23.07.20; published 28.08.20.

Please cite as:
Quinn LM, Davies MJ, Hadjiconstantinou M
Virtual Consultations and the Role of Technology During the COVID-19 Pandemic for People With Type 2 Diabetes: The UK Perspective
J Med Internet Res 2020;22(8):e21609
URL: http://www.jmir.org/2020/8/e21609/
doi:10.2196/21609
PMID:32716898

Original Paper

Meeting Kids Where They Are At–A Substance Use and Sexual Risk Prevention Program via Telemedicine for African American Girls: Usability and Acceptability Study

Cristina Lopez[1], PhD; Amanda K Gilmore[2], PhD; Angela Moreland[1], PhD; Carla Kmett Danielson[1], PhD; Ron Acierno[3], PhD

[1]Medical University of South Carolina, Charleston, SC, United States

[2]Georgia State University, Atlanta, GA, United States

[3]University of Texas, Houston, TX, United States

Corresponding Author:
Cristina Lopez, PhD
Medical University of South Carolina
99 Jonathan Lucas Street
MSC 161
Charleston, SC, 29425
United States
Phone: 1 8438761034
Email: lopezcm@musc.edu

Abstract

Background: Rural African American youth lack access to drug and sexual risk–taking *prevention* programs available in more urban areas. Recent data indicate that rural youth now use substances at higher rates and at younger ages than their urban peers.

Objective: This study aims to evaluate the initial usability and acceptability of a low-cost, technology-based approach to delivering effective, culturally tailored, integrated substance use disorder (SUD) and HIV risk behavior prevention programs to African American female youth to inform the use of this intervention via telemedicine for rural youth.

Methods: Effective SUD prevention strategies and emotion regulation skills were integrated into an existing evidence-based HIV risk reduction program culturally tailored for African American female adolescents—Sisters Informing, Healing, Living, and Empowering (SIHLE)—and delivered to 39 African American female youth via group telehealth. The evaluation of the resulting program, 12-session SIHLEplus, was completed by 27 girls who also completed self-report measures that assessed sexual risk behaviors (eg, number of partners and age of sex initiation), substance use, exposure to traumatic events, and emotion regulation.

Results: The descriptive and qualitative results of the pilot study demonstrate the initial usability and acceptability of delivering evidence-based prevention successfully via telehealth to help address health disparities in this vulnerable population.

Conclusions: Although more research is needed, the findings from this study suggest that SIHLEplus has demonstrated initial usability and acceptability.

(J Med Internet Res 2020;22(8):e16725) doi:10.2196/16725

KEYWORDS

adolescents; substance use; sexual risk reduction; telehealth; prevention programs; mobile phone

Introduction

Substance use is common among adolescents, with 18.2%-55.7% of 8th to 12th graders using alcohol, 23.9% using cannabis, and 9% using other illicit drugs annually in 2018 [1]. Rural youth are particularly at risk for substance use [2,3], with youth in rural areas using substances such as alcohol, tobacco, and illicit drugs at higher rates [4,5] and at younger ages [5] than their urban peers. This is particularly problematic because early initiation of substance use is strongly linked with substance use disorders (SUD) in both adolescence and adulthood [6-9]. The risk for negative sequelae related to early substance use initiation in female adolescents (eg, education and employment failure, risky sexual behavior, unintended pregnancy, criminal

justice system involvement, trauma exposure, mental and physical health problems) continues into adulthood [10-12], highlighting the public health impact of adolescent substance use. This is particularly a concern in low-resource areas where recovery programs are extremely limited and difficult to access. Indeed, only 1.2% of the nation's substance abuse treatment facilities are located in small nonurban, nonadjacent counties [13]. These health care disparities in substance use services exacerbate the problems of adolescent substance use, as states with the greatest unmet need for alcohol treatment among youth aged 12-17 years are largely rural [14].

Substance use can impair sexual decision making [15], leading to a range of consequences from sexual risk behaviors, including unwanted pregnancy and sexually transmitted infections (STIs), as well as trauma exposure such as sexual assault [16]. The consequences of sexual risk behaviors disproportionately impact adolescent girls, with youth aged 15 to 24 years representing half of new STIs [17], and girls are particularly vulnerable to the consequences of STIs [18]. HIV infection is particularly problematic for African American youth living in the rural South, with an infection rate over 800% that of Whites in the same region [19]. Therefore, programs are needed to provide integrated substance use and sexual risk–taking prevention to rural African American adolescent girls.

Adolescent Substance Use and Sexual Risk Taking

Evidence-based HIV risk reduction programs exist, which are culturally tailored for African American female adolescents. Specifically, Sisters Informing, Healing, Living, and Empowering (SIHLE) [20] is a group-based prevention program that uses ethnic and gender pride to empower African American female adolescents and provides content on HIV knowledge, communication, condom use skills, and healthy relationships. African American women who participated in SIHLE, compared with those who participated in a control group on exercise and nutrition, reported more consistent condom use, less new sexual partners, and better condom negotiation skills 1 year after the prevention program [20]. However, SIHLE does not address substance misuse despite substance use in itself being a risk factor for sexual risk behaviors. Evidence-based models of substance use prevention, including motivational enhancement therapy [21], are effective at reducing substance use among adolescents, and targeting two related health behavior constructs in an integrated manner is more effective than targeting each separately [22]. Due to the interrelated nature of substance use and sexual risk behaviors, integrated interventions are needed for African American girls who live in rural areas. Integrated interventions targeting multiple health behaviors can also address crosscutting underlying contributors to both behaviors, such as emotion regulation. Adolescent health behavior change is influenced by emotion regulatory processes on behavior and behavior change [23]. Therefore, training in problem-solving and condom use skills to modify adolescent change through skill acquisition, without any modification of the underlying system of self-regulating emotion, is shortsighted. To our knowledge, no studies have examined whether SIHLE or motivational enhancement therapy alone changes emotion regulation. However, neither intervention directly targets emotion dysregulation or provides emotion regulation skills. This is an important limitation given that both sexual risk behavior and drug use are associated with difficulties in emotion regulation among adolescents [24-26], and directly teaching emotion regulation skills is associated with changes in sexual risk behavior among adolescents [27-29]. Thus, targeting emotion regulation in an integrated intervention would not only address a factor associated with substance use and sexual risk behaviors but it may also facilitate behavior change among adolescents.

Need to Develop Telemedicine Approaches to Reach Rural African American Girls

Fortunately, and in contrast to the lack of accessible preventive programs for substance use in rural African American adolescent girls, community-based interventions for sexual risk behavior designed for rural settings exist and show initial efficacy [30,31]. It may be particularly useful, therefore, to leverage the existing, tested infrastructure of such HIV prevention programs and incorporate substance risk reduction strategies alongside those for risky sexual behavior. Leveraging technology to deliver this service via telemedicine technology directly into participants' communities and homes helps overcome traditional barriers to care (eg, stigma, lack of anonymity, and lack of transportation) and lack of resources (eg, no access, not enough providers to meet local needs, and no treatment space) that typically deter rural youth from successful program completion. Telemedicine has been found to be noninferior to in-person delivery of complex treatments [32] and may be an effective means to address health disparities by reducing the significant barriers associated with accessing treatment [33].

This Study

This study examined the usability and acceptability of a telemedicine prevention program for African American rural youth that targets substance use and sexual risk behaviors (SIHLEplus). SIHLEplus integrates SIHLE [20] with motivational enhancement therapy [21] to target both substance use and sexual risk behaviors in an integrated manner using culturally tailored programming for African American adolescent girls and led by a community-accepted professional (eg, stakeholder at a community agency, near peer with established trust). In addition, emotion regulation exercises were integrated to increase the likelihood of implementing sexual risk reduction skills and substance use prevention strategies. It was hypothesized that SIHLEplus would be usable and acceptable among African American girls using a telemedicine delivery modality.

Methods

Overview

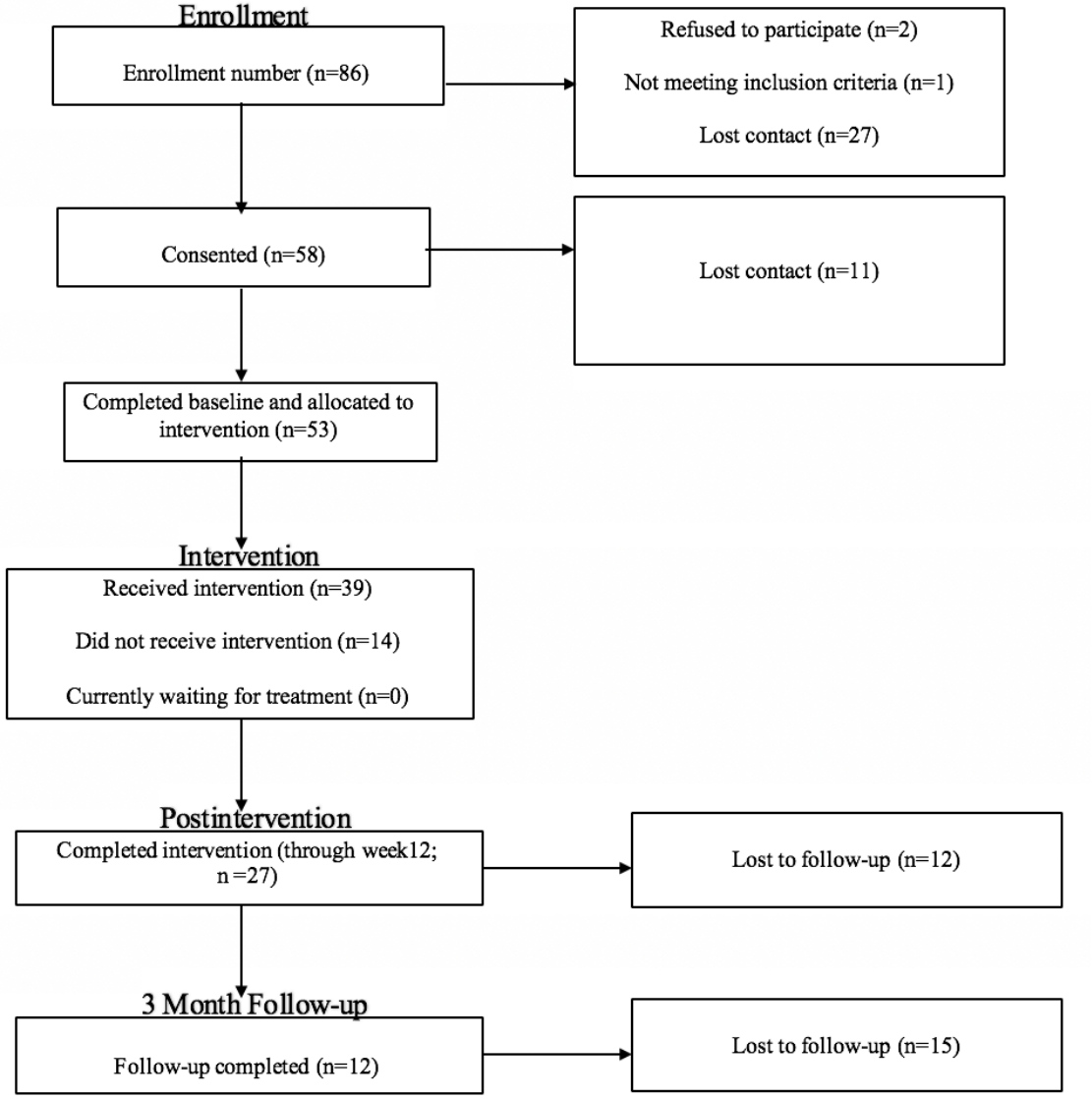
Quantitative and qualitative data were collected from adolescents (aged 13-18 years) recruited from local schools and community events to participate in a preventive group-telehealth intervention, SIHLEplus, focused on prevention of HIV, substance use, and other risky behaviors. Before engagement in SIHLEplus, self-report questionnaires were completed by adolescents assessing substance use, sexual risk, internalizing symptoms, and difficulties with emotion regulation.

Postintervention usability interviews were also conducted with a subsample of the adolescents (n=9) who completed at least 80% (ie, 9 sessions or more) of the 12-session intervention (11 group sessions and 1 parent session) to obtain additional information about engagement factors related to content and format (eg, telehealth).

Participants and Procedures

Participants were recruited through several community events and health fairs, referrals from a multicounty in-school case management program (14/36, 38.9% reported referrals from a school-related professional or paraprofessional), snowball recruitment (18/36, 50% of the sample reported hearing about the study from a friend or family member), and referrals from guidance counselors at several middle and high schools. Given the setting in a tricounty area of a southeastern state (South Carolina), several community events took place in rural counties and more urban areas that attracted both urban and rural populations (eg, Black Expo). Eligibility for participation included identifying as Black or African American and female, aged 13-18 years, and endorsing engagement in a health risk behavior (ie, either sexual activity in the past 12 months or reporting mild substance use activity in the previous 3 months). Reliable access to the internet was also assessed (all potential participants met internet access criteria). These pre-eligibility screening questions were completed before the consent process. After informed consent was obtained from participants and their caregivers, adolescents completed the baseline self-report questionnaires, including the drug abuse screen test (DAST)-10 instrument for substance use assessment. Consistent with other substance use prevention protocols, exclusion criteria included moderate to severe substance use problems as assessed by the DAST-10 [34]. Participants who met this exclusion criterion (n=1) were connected with an appropriate treatment agency to receive an established treatment intervention. Participants who completed the self-report questionnaires at baseline and postintervention (n=27) were adolescents who identified as Black (25/27, 92.6%) and Hispanic (2/27, 7.4%). The participants were 7th graders (2/27, 7.4%), 8th graders (6/27, 22.2%), 9th graders (5/27, 18.5%), 10th graders (4/27, 14.8%), 11th graders (4/27, 14.8%), 12th graders (2/27, 7.4%), and community college students (4/27, 14.8%). All participants identified as female, and the US census designation demonstrated that 27.3% of the sample was from a rural area, 30.3% from an urbanized area, and 42.4% from an urban cluster. Nine postintervention usability interviews (22% rural) were conducted with adolescents who had completed the web-based program. Six intervention groups were conducted over the time of the study, with groups ranging from 7 to 9 members (with the exception of the first group, which contained 2 participants). Please refer to the Consolidated Standards of Reporting Trials diagram for more details on screening and recruitment (Figure 1).

Figure 1. CONSORT diagram.



The Intervention: SIHLEplus

SIHLEplus is an integrated intervention built from existing evidence-based HIV risk reduction program culturally tailored for African American female adolescents: SIHLE [20]. Using the empirically supported ADAPT-ITT (Assessment, Decision, Administration, Production, Topical experts, Integration, Training, Testing) model [35,36] for systematically adapting HIV behavioral programs, evidence-based substance use prevention strategies and emotion regulation skills were integrated into a telehealth-friendly SIHLEplus program. SIHLEplus complements skill acquisition of HIV prevention models with enhancement of emotion regulation skills and coping strategies to increase adolescent behavior change, resulting in the adoption of healthy HIV prevention behaviors (eg, reduced substance use). SIHLEplus consists of 11 weekly 1-hour group sessions with youth participants and 1 20-min individual session with each parent of SIHLEplus participants at some point between weeks 5 and 9 (totaling 12 weeks). All these sessions are delivered via telehealth, with participants logging in to the group virtual room by clicking on a hyperlink sent out by the group facilitator to the participant's email address or via a text weblink. A brief session with parents was included and highlights the importance of parental communication about sex and substance use and parental monitoring, 2 central components in youth externalizing behavior treatment [37,38]. Various education delivery strategies are employed in each session, including lectures, discussions, role-play and cognitive rehearsal, skill acquisition demonstrations, and interactive games (Table 1).

Table 1. Table of contents for the SIHLEplus manual.

Modules, activities
Session One[a]***:*** **Gender, Pride, and Self-Image**
1.1 Program Intro—Introduce ground rules, SIHLEplus motto, and purpose of the group
1.2 Definition of Self-Image and Behavior—Discuss positive characteristics of young Black women
1.3 A Room Full of Sisters—Review poem discussing the uniqueness of Black women
Session Two: **Influence, Relationships, and Role Models**
2.1 Call Me Black Woman—Review poem to revisit pride in being a young Black woman
2.2 Magic Triangle—Introduce the relationship between one's thoughts, emotions, and one's reactions
2.3 Media Masquerade—Participants are shown a video and pictures to start the discussion of how the media can influence one's behaviors and decisions. Emphasis is placed on how the media normalizes sex and drug use
Session Three: **Decision Making and Peer Influences**
3.1 Everyday Decision and Difficult Decision—Participants and group leader discuss the differences between a common decision (ie, what one should wear today) versus a more difficult decision (ie, should I have unprotected sex) and how those decisions can impact one's life on various levels
3.2 What influences our Decision Making—Participants discuss potential influences of their decision making (ie, parents, peers, and media) and how potential influencers play a part in both one's common (everyday) and difficult decisions
3.3 Identifying Situations that Involve Decisions—Participants discuss the benefits of one's reactions and behaviors aligning with one's self-image and the various options one has in how they react to an experience or situation
Session Four: **Personal Values and Challenging Negative Views**
4.1 Values: What Matters Most—Discuss individual values and how those values contribute to one's decision making
4.2 Challenging Different Perspectives and Views—Discuss core beliefs, influences of core beliefs, and how core beliefs can influence our interpretation of and perspective in an experience
4.3 SIHLE Sistas are Special—Participants are reminded of why they are so special and unique. There is discussion about how SIHLE sistas are special because they know what is important to them, but most of all, they know how to care for themselves
Session Five: **STD**[b] **Knowledge**
5.1 STI[c]*Knowledge*—Participants are given facts about STIs that are common among teens. There is further discussion about the myths and facts surrounding the behavioral causes and contributors to STIs
5.2 HIV/AIDS Information and Facts—Participants discuss knowledge about contracting HIV/AIDS, the importance of getting tested, and where testing centers are located
5.3 Are you at Risk—Participants discuss levels of risk (ie, unsafe, safe, and safer) and the behaviors that fall in line with each individual level of risk
Session Six: **Relationships**
6.1 Healthy and Unhealthy Relationships—Participants discuss the differences between healthy and unhealthy relationships
6.2 A Chain Analysis: Freeze—Participants discuss the importance of slowing down the sequence of stressful events in efforts to make a better decision in response to a potentially stressful situation
Session Seven: **Self-Care and Pregnancy Prevention**
7.1 Taking Care of You—Participants revisit the importance of self-care, more specifically self-care in reference to safer sexual behaviors
7.2 LIPSTICK[d]—Participants learn the proper steps to appropriate condom use
7.3 Phenomenal Women—Participants review Phenomenal Women and discuss one's sense of pride and Black beauty
7.4 What's In It For You?—Participants reexamine the outcomes and benefits of choosing to engage in "safe" and "safer" behaviors
Session Eight: **What is drug abuse?**
8.1 What is drug abuse—Drug abuse is defined, and common drugs used among teens are discussed
8.2 Identifying the Causes of Drug Use—Common reasons for drug use among teens are discussed
8.3 Effects of Drug Abuse—Common effects of drug abuse and how the outcomes of drug abuse are initially associated with one's original decision to use drugs are discussed
8.4 Reasons not to Abuse Drugs—Participants are reminded of why some teens use drugs and the benefits of not using drugs
8.5 Jeopardy—Participants are quizzed on the information they have learned thus far

Modules, activities
Session Nine: **Refusal Skills**
9.1 KISS: Know, Indicate, State, and Stand—Participants discuss the importance of refusing unsafe sex and other high-risk behaviors
9.2 Name Game—Participants are taken through a simulated web-based party to discuss the outcomes of engaging in high-risk behaviors at the party
9.3 What Have we Learned?—Participants recap the information learned about the transmission of STDs and HIV
Session Ten: **Anxiety and Anger's Contribution to Risky Decisions**
10.1 Anxiety—Anxiety is defined, and participants discuss how anxiety can impact behaviors and decisions
10.2 Explanation of Anger—Anger is defined, and participants discuss how anger can impact behaviors and decisions
Session Eleven: **Communication**
11.1 Three Ways to Say it—Participants define assertive, aggressive, and passive communication through role-playing. Participants also talk about the importance of assertive communication in refusing unsafe sexual behaviors
11.2 Still I Rise—Participants review poem and discuss one's ability to overcome obstacles
11.3 Graduation! Participants are welcomed into SIHLEplus sisterhood
Parent Module
Parent Module provides support to parents about how to handle the topic of sex and the importance of parent-child communication and parental monitoring

[a]Text in italics is the name of the chapter used in SIHLE.

[b]STD: sexually transmitted disease.

[c]STI: sexually transmitted infection.

[d]LIPSTICK acronym stands for steps for appropriate condom use–the steps are as follows: L–Look at the expiration date; I–Inspect the condom tip to make sure it is latex or polyurethane; P–Pinch the tip of the condom to leave room for ejaculation; S–Slowly roll the condom to the base of the penis; T–Time for some action!; I–Immediately have your partner remove his penis out and away from you after he ejaculates; C–Carefully take off condom and throw it in the trash; K–Keep in mind all of these steps each and every time you have sex!

Modules of SIHLEplus cover the following topics: gender, pride, and self-image (session 1); influence, relationships, and role models (session 2); decision making and peer influences (session 3); personal values and challenging negative views (session 4); sexually transmitted disease (STD) knowledge (session 5); stress and relationships (session 6); self-care and pregnancy prevention (session 7); substance use knowledge (session 8); refusal skills (session 9); influence of anxiety and anger on behavior (session 10); effective communication (session 11); and parent module on parent-child communication and parental monitoring (session 12). Of note, the terminology used to describe a romantic partner utilizes gender-neutral language, and discussion of safe sex practices within same-sex couples is reviewed in the content. All SIHLEplus facilitators (n=3) were formally trained in the delivery of in-person SIHLE. Group facilitator training for this study focused on adaptations in SIHLEplus and coaching on the software platform.

Measures

Usability and Acceptability

Initial usability and acceptability were assessed using recommendations to peers, content engagement, and delivery mode. Specifically, 3 single items were used to assess usability: "I would recommend this program to my friends," "The SIHLEplus content was appropriate and engaging," and "The videoconferencing format that was on my laptop or phone was easy to use." Response choices ranged from strongly agree (1) to strongly disagree (5). Furthermore, CollaboRATE [39], a 3-item assessment of shared decision making in treatment, was used to assess the effectiveness of the SIHLEplus group facilitator. These items included the following: "How much effort was made to help you understand your health issues," "How much effort was made to listen to what matters most to you about your health issues," and "How much effort was made to include what matters most to you in choosing what to do next." Response choices ranged from no effort at all (0) to every effort was made (9).

Baseline Risk Behaviors

Drug, alcohol, trauma, emotional regulation, and sexual risk behaviors were assessed using the DAST-10 [34,40] and the alcohol use disorders identification test (AUDIT) [41,42]; items related to sexual risk and exposure to traumatic events were assessed using the University of California at Los Angeles Posttraumatic Stress Disorder Child/Adolescent Reaction Index (UCLA-RI); and difficulties with emotion regulation were assessed using the emotion regulation questionnaire (ERQ) and difficulty in emotion regulation scale (DERS) at baseline. The DAST-10 [34] is a 10-item self-report instrument that yields an index score measuring the degree of consequences related to drug abuse. Similarly, the AUDIT is a 10-item self-report screening tool to assess alcohol consumption, drinking behaviors, and alcohol-related problems. To assess for sexual risk, questions asked are as follows: "Have you ever willingly had vaginal sex" (yes, no, prefer not to answer) and "How old were you the first time you willingly had vaginal sex?" (prefer not to answer, answer options ranged from 9 to 18 years old). The UCLA-RI [43,44] is a self-report questionnaire to screen for exposure to traumatic events in school-aged children and

adolescents. The 15-item screener includes items that assess for child sexual or physical abuse, witnessing domestic or community violence, natural disasters, unexpected deaths, scary medical procedures, war, and serious accidents and an option to write in a traumatic event in an open text field. The ERQ [45] is a 10-item scale designed to measure participants' tendency to regulate their emotions through cognitive reappraisal (CR) and expressive suppression (ES; eg, emotional avoidance). Each item has response options ranging from 1 (strongly disagree) to 7 (strongly agree). Both total scores and subscales were calculated by summing the respective questions. In addition, the DERS [46] is a 36-item self-report questionnaire that was administered to assess multiple aspects of emotion dysregulation. Response options for each of the DERS items range from 1 (almost never) to 5 (almost always). Both total scores and subscales were calculated by summing the respective questions.

Procedure

The research protocol was approved by the institutional review board (IRB) at the affiliated university, including a written informed consent process of all participants. The use of HIPAA-compliant Adobe Connect for the group telehealth format was used because the functionality allowed for multiple user logins from various types of devices (eg, laptops, tablets, and smartphones). The institution (Medical University of South Carolina) had already used the HIPAA-compliant Adobe Connect platform for telehealth services in a clinical manner, but IRB approval was sought for research use of Adobe Connect [33]. Before completion of the self-report questionnaires, informed consent and parental consent (aged under 18 years) were obtained. Participants (ie, adolescents) were compensated with Walmart gift cards following completion of web-based (Research Electronic Data Capture) self-report questionnaires at baseline (US $20) and postintervention (US $30).

Statistical Analyses

Quantitative Analyses

Quantitative analyses were conducted using SPSS 25. Descriptive statistics were completed to assess satisfaction with SIHLEplus (Table 2) and to describe the baseline characteristics of the sample.

Table 2. Descriptive statistics of usability findings.

Usability items	Indicated agree or strongly agree, n (%)	Value, mean (SD)
I would recommend this program to my friends	9 (100)	1.22 (0.55)
The SIHLE[a]plus content was appropriate and engaging	9 (100)	1.22 (0.43)
The videoconferencing format on my laptop or phone was easy to use	7 (77.8)	2.06 (1.00)
Participant ratings of group facilitator	N/A[b]	8.39 (0.73)

[a]SHILE: Sisters Informing, Healing, Living, and Empowering.

[b]N/A: not applicable.

Qualitative Analyses

Usability interviews were conducted with participants post intervention (ie, who had completed the full intervention) utilizing semistructured questions, with prompts based on the participant's responses. Content addressed usability and acceptability of the intervention and software platform (a list of semistructured questions is shown in Multimedia Appendix 1). Interviews lasted approximately 45 min and were transcribed. NVivo 12.0 software was used for data management and analysis. Analyses of the focus group data consisted of qualitative content analysis [47] informed by grounded theory [48]. Grounded theory explores participants' unique perspectives via the identification of themes or patterns that naturally emerge from the data and the systematic classification of these themes [49]. Specifically, a 3-step inductive approach was utilized, in which the interview responses (ie, raw data) were carefully examined to develop a comprehensive codebook to capture all possible themes emerging from the data [50]. The codebook was then used by 2 independent coders to code and analyze each participant's responses to the interview questions [47,49]. Coders were able to apply more than one code to participant responses if applicable. Interrater discrepancies were discussed and resolved by 2 independent coders. Finally, themes were refined, merged, or subdivided into subthemes. Percentages reported in the findings reflect the percentage of times that the specific themes were mentioned by all participants in the interviews.

Results

Descriptive Statistics

At baseline, 12 (25.5%) participants indicated that they had engaged in consensual vaginal sex during their lifetime. Responses for the age at the first consensual sexual intercourse ranged from 12 to 18 years. The mean age for engaging in consensual sex was 15 (SD 1.81) years. Participant responses indicated that 41.7% of those who had consensual vaginal sex reported the age of initiation of willing vaginal sex at 14 years or below (n=5). Although average scores on the DAST-10 (mean 0.33, SD 0.72) and AUDIT (mean 0.94, SD 1.43) indicated low drug and alcohol use, a total of 9 (19%) participants indicated that they had used drugs. Scores endorsed on the DAST-10 showed that 10.4% of the sample scored 2 or above, indicating problems consistent with a mild SUD. In response to exposure to traumatic events, the sample reported a range of 0 to 11 (out of potential 15), with a mean of 3.14 (SD 2.31) traumatic events experienced in their lifetime. Frequencies demonstrated that more than 50% of the sample reported exposure to 3 or more traumatic events in their lifetime, with 16.7% endorsing 6 or more traumatic events. The ERQ yielded a mean of 27.15 (SD

8.74) for the CR subscale, and the ES subscale yielded a mean of 15.29 (SD 6.00), scores higher than the average CR and ES scores obtained from a community sample of girls aged 10 to 18 years (mean 21.47, SD 3.81 and mean 10.18, SD 2.97, respectively [45]). Emotion regulation as measured by the DERS yielded a total mean score of 95.35 (SD 24.11), a score that is higher than the average of 80.2 (SD 23.4) in a community sample of female adolescents (higher scores indicated more difficulty with emotion regulation [51]).

Initial Usability and Acceptability Findings

Overall, participants rated the usability of SIHLEplus positively. The majority of participants indicated that they would recommend SIHLEplus to their friends, that the content was appropriate and engaging, and that the videoconferencing format was easy to use (Table 2). Furthermore, participants were able to engage with the group facilitator over the videoconferencing format and reported high facilitator engagement and effort (Table 2).

Qualitative Results

Four overarching themes, each with their own subthemes, emerged from the participants' responses during the interviews. Each is described below with representative quotes provided throughout for illustrative purposes.

Program Usage

Discussion of program usage occurred on 15 occasions throughout the 9 interviews (14.7% of all content). Program usage included discussion of place of log-in (46.7% within this theme), including at home, in their bedroom, or in the car; ability to log in to the site (40% within this theme), including difficulties with log-in on first usage, ongoing difficulty with log-in, log-in taking a long time, or always being able to log-in; and technical concerns (13.3% within this theme), including some technical issues that needed repair. Specific comments that participants made about program usage included:

> *I logged in at home, in my room where I usually do my homework.*
>
> *I did have problems a couple of times because of my phone. Sometimes you can't copy and paste the link into the bar. Sometimes I had to type out the whole link.*

Strengths of the Program

Conversation regarding strengths of the program occurred on 54 occasions throughout the interviews (52.9% of all content). The reported strengths of the program included the relationships that were formed with other participants (11.1%), such as feeling comfortable with group members, liking that they were able to build relationships with similar-aged females, and liking the intimacy of the group. In addition, comments about new information learned in the group arose in 31.5% of interviews within this theme, including communication skills, safe sexual practices (condom use), positive relationship skills, STDs, and decision making. Adolescents commented on the use of skills learned in the program (53.7% within this theme), such as communication skills, safe sexual practices (condom use), positive relationship skills, problem-solving skills, meditation, coping skills or how to handle stress, decision making, and keeping sex education conversations ongoing with adolescents. Finally, adolescents liked the web-based platform (3.7% within this theme), reporting that it was more convenient and that they felt more open sharing private information on the web than they would have been in an in-person setting. Specific comments about the strengths of the program included:

> *Liked other girls my age because we all had the same ideas. Well, we were different, but we were all in the same boat and could relate easier.*
>
> *It was pretty comfortable. I mean other than the fact that I knew one person in the group but I wasn't, like, scared to talk about anything.*
>
> *When it came down to the sexual stuff, when done with the STD talk, it had me sit and think, we were always having sex without a condom. I know he was locked up, but we need to practice safe sex, for STD but also to prevent pregnancies.*
>
> *Used some skills to communicate better and some skills with best friend.*
>
> *I also use the meditating when I am too frustrated. I don't know why it helps, when I felt my temperature rising.*

Limitations of the Program

Discussions of program limitations occurred on 5 occasions throughout the 9 interviews (4.9% of all content). Overall, participants noted that they would prefer if the program was offered in person (60% within this theme), that they had to miss some sessions because of other commitments (20% within this theme), and that the group setting made it difficult to share personal information (20%). Some specific comments about the limitations of the program included:

> *It was just a different way of learning, but I prefer to sit down in a room. I think I like that better.*
>
> *I said less things in a group. I'm not up and go get it with other people...depends on how many people I knew before giving out personal information.*
>
> *I didn't like that I missed sometimes due to conflicts in my schedule.*

Sharing Information About the Program

Conversation about sharing information about the program occurred on 28 occasions throughout the 9 interviews (27.5% of all content). Specifically, participants reported that they would recommend the program to a friend (14.3% within this theme) or talk with their parents about the program (17.9% within this theme). Regarding sharing information with others (60.7% within this theme), 24% stated that they had shared information with others, 18% reported that they would feel comfortable sharing information but had not yet, 12% stated that they had not shared because they did not want to, 18% did not share because they did not think it was necessary, and 18% reported that they would not feel comfortable sharing information with others. Specific comments about sharing information with others included:

I showed my best friend how to use a condom. We were having girl talk about sex and how her boyfriend doesn't use condoms. She's on birth control and I told her birth control isn't 100%, but I showed her with paper and pinch the tip. Have it folded like a hat (not like a baby's bottle because you don't want babies). She didn't know you could use a condom backwards.

I mean like, the friends I talk to, don't really engage in the things we learned about so it wasn't necessarily helpful but it was good information to have, if that makes sense.

Discussion

Principal Findings

This study assessed the usability and acceptability of SIHLEplus, a telemedicine prevention program for African American youth that targets substance use and sexual risk behaviors. Findings from this initial usability and acceptability pilot using both quantitative and qualitative results from African American adolescent girls indicated that the content was engaging and easy to use and that they would recommend the program to their friends. Furthermore, recruitment numbers show initial interest and need. Although the sample was not predominately rural, the findings from this suggest that SIHLEplus can be delivered via telemedicine in an acceptable manner and has important implications for rural youth. This is an important finding given that substance use is associated with sexual risk behaviors [15] and rural African American youth are at particular risk for substance use [2,3].

The participants reported several strengths of SIHLEplus. Participants indicated that they were able to form relationships with other participants via the telemedicine platform and that they were able to learn helpful skills. Notably, the adolescents indicated that they were more willing to share private information via telemedicine than they would have in an in-person treatment modality. This was surprising because the telemedicine delivery was only intended to assist with delivery of SIHLEplus to rural youth and not intended to enhance the treatment effects or uptake. However, this is consistent with the literature suggesting that individuals are more likely to report socially unacceptable behaviors via computer-based surveys than in-person surveys [52]. It is possible that the telemedicine platform allowed them to distance themselves from social acceptability pressures that may exist within in-person group settings. Future work is needed to test the efficacy of this SIHLEplus on substance use, sexual risk, and emotion regulation outcomes among rural youth.

Participants, in general, indicated that they would share the information learned in the program with others or that they would recommend the program to their peers. Specifically, about one-fourth of the adolescents indicated that they had already shared the skills they learned in the program, and almost all of the adolescents (94%) indicated that they would recommend the program to their friends. These findings indicate the initial usability and feasibility of SIHLEplus. Furthermore, willingness to talk with peers about these topics indicates that this does not only have the potential to reduce substance use and sexual risk behavior among those who participated, but their peers as well. Given that social norms are a strong predictor of both substance use and sexual behaviors among youth [53], this unexpected finding suggests that future work should examine the social network impacts of SIHLEplus in addition to the individual changes in substance use and sexual risk behavior.

Despite the initial acceptability and usability of SIHLEplus, there were some factors that may need to be considered before conducting a large clinical trial to test its efficacy. First, some participants indicated that they had some difficulty participating in the program because of commitments with after-school activities. Given the noted positive effects of extracurricular activities and endorsing school pride in adolescent academic achievement [54,55], the study team supported participants' requests to be removed from the intervention. It is possible that SIHLEplus would be better adopted in states where sex education was more accepted and potentially have it as part of the school curriculum or as an extra credit activity for some courses. It is also possible that the development of a mobile health app would assuage this concern, given that it would allow adolescents to access the intervention at their convenience. However, the participants did indicate that they enjoyed the relationships formed, so the group telemedicine component may be a crucial piece of the intervention. Future work is needed to determine the feasibility of SIHLEplus given the time constraints of adolescents. Specifically, future work is needed to determine the feasibility of SIHLEplus among *rural* African American girls. It is possible that there may be differential time constraints among rural youth, and this should be considered before conducting a large-scale randomized clinical trial.

The provision of prevention and treatment through telemedicine to high-risk populations may be a useful tool to use in conjunction with treatment for mental health disorders. In the sample in this study, there was particularly high endorsement of exposure to traumatic events. Given the strong association between exposure to traumatic events, substance use, and sexual risk behaviors [15,56-58], it may be worthwhile to consider if prevention programs like this are effective at reducing exposure to subsequent potentially traumatic events. If treatment for posttraumatic stress disorder is needed, it may be possible to address posttraumatic stress symptoms in conjunction with substance use and sexual risk behaviors for rural African American youth. Emotion regulation has also been identified as a risk factor for youth who have been exposed to a traumatic event, so it is not surprising that many youth would benefit from SIHLEplus for externalizing symptoms (sexual risk reduction and substance use) and internalizing symptoms (anxiety related to traumatic stress).

Strengths and Limitations

Strengths of SIHLEplus and this usability and acceptability study include the innovation of this integrated prevention program delivered via telemedicine, allowing for participation from an at-risk group that typically does not have access to prevention programming. This study demonstrated that the internet or smartphone access as a study criterion did not lead

to an ineligibility status for any interested potential participant. Although rural residence was not part of the inclusion criteria for this initial usability step, all participants identified as an ethnic or racial minority and thus are considered underserved populations accessing behavioral health services. Given the false misconception that rural and underserved families may not have access to resources for telemedicine delivery, the findings of this pilot study can be used to justify additional delivery of medical and health information via telemedicine.

This study only assessed the initial usability and acceptability of SIHLEplus; its feasibility and efficacy need to be established before implementation. Therefore, an important next step includes assessing both feasibility and intervention effects among at-risk rural African American adolescent girls. Consistent with approaches used to define the rural HIV epidemic in South Carolina [59], rural areas will be defined using a zip-code approximation version of the rural urban commuting area (RUCA) codes to recruit participants from RUCA codes 4 to 10 (ie, rural areas, *zip-code RUCA approximation* [60]).

Although this study establishes initial usability, more studies are needed to ensure that the program can be delivered to formally defined rural populations. Subsequent feasibility studies with larger samples may look into potential differences in usability themes and engagement from rural- and more urban-residing adolescents. Although rural and urbanized areas and urban clusters were helpful descriptions for the initial stages of usability, future studies could also collect socioeconomic data to ensure generalizability across several rural subpopulations. Furthermore, a large randomized clinical trial assessing the effects of SIHLEplus on sexual risk behaviors, substance use, and emotion regulation is warranted.

Conclusions

The findings from this study provide initial usability and acceptability for an integrated substance use, sexual risk, and emotion regulation program for at-risk African American adolescent girls delivered via telemedicine. Although future work is needed to determine feasibility and efficacy, preliminary findings established the initial usability and acceptability of SIHLEplus.

Acknowledgments

Data collection and manuscript preparation were supported by grants from the National Institute of Child Health And Human Development, the Office of Research on Women's Health (Building Interdisciplinary Research Collaborations in Women's Health K12HD055885; principal investigator [PI]: CL), and the CTSA is a Center for Translational Science Award. The institutions' CTSA is called SCTR, located at Medical University of South Carolina (National Institutes of Health-National Center for Advancing Translational Science Grant Number UL1 TR001450). The manuscript preparation was also partially supported by grants from the National Institute on Drug Abuse (grant K12DA031794-03, PI: Moreland; grant K23DA036566, PI: AKG).

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of semistructured questions for usability interviews.
[DOCX File , 13 KB - jmir_v22i8e16725_app1.docx]

References

1. Schulenberg J, Johnston L, O'Malley P, Bachman J, Miech R, Patrick M. Monitoring the Future National Survey Results on Drug Use, 1975-2018: Volume II, College Students and Adults Ages 19-60. Deep Blue. 2019. URL: https://deepblue.lib.umich.edu/handle/2027.42/150623 [accessed 2020-07-27]
2. Johnston L, O'Malley P, Bachman J, Schulenberg J. Monitoring the Future: National Survey Results on Drug Use 1975–2006. Deep Blue. 1975. URL: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/137787/vol2_2006.pdf?sequence=1 [accessed 2020-07-27]
3. High-Risk Substance Use Among Youth. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/healthyyouth/substance-use/index.htm [accessed 2020-07-14]
4. Lambert D, Gale J, Hartley D. Substance abuse by youth and young adults in rural America. J Rural Health 2008;24(3):221-228. [doi: 10.1111/j.1748-0361.2008.00162.x] [Medline: 18643798]
5. Pruitt L. The Forgotten Fifth: Rural Youth and Substance Abuse. SSRN. 2009. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1279669 [accessed 2020-07-27]
6. Anthony JC, Petronis KR. Early-onset drug use and risk of later drug problems. Drug Alcohol Depend 1995 Nov;40(1):9-15. [doi: 10.1016/0376-8716(95)01194-3] [Medline: 8746919]
7. Clark DB, Bukstein OG. Psychopathology in adolescent alcohol abuse and dependence. Alcohol Health Res World 1998;22(2):117-21, 126 [FREE Full text] [Medline: 15706785]

8. Grant BF, Dawson DA. Age at onset of alcohol use and its association with DSM-IV alcohol abuse and dependence: results from the national longitudinal alcohol epidemiologic survey. J Subst Abuse 1997;9:103-110. [doi: 10.1016/s0899-3289(97)90009-2] [Medline: 9494942]
9. Kilpatrick DG, Acierno R, Saunders B, Resnick HS, Best CL, Schnurr PP. Risk factors for adolescent substance abuse and dependence: data from a national sample. J Consult Clin Psychol 2000 Feb;68(1):19-30. [doi: 10.1037//0022-006x.68.1.19] [Medline: 10710837]
10. Brook JS, Adams RE, Balka EB, Whiteman M, Zhang C, Sugerman R. Illicit drug use and risky sexual behavior among African American and Puerto Rican urban adolescents: the longitudinal links. J Genet Psychol 2004 Jun;165(2):203-220. [Medline: 15259877]
11. Chatterji P. Illicit drug use and educational attainment. Health Econ 2006 May;15(5):489-511. [doi: 10.1002/hec.1085] [Medline: 16389630]
12. DuRant RH, Smith JA, Kreiter SR, Krowchuk DP. The relationship between early age of onset of initial substance use and engaging in multiple health risk behaviors among young adolescents. Arch Pediatr Adolesc Med 1999 Mar;153(3):286-291. [doi: 10.1001/archpedi.153.3.286] [Medline: 10086407]
13. Gale J, Lenardson J. Distribution of Substance Abuse Treatment Facilities Across the Rural—Urban Continuum. Muskie School of Public Service: University of Southern Maine. 2007. URL: http://muskie.usm.maine.edu/Publications/rural/wp35b.pdf [accessed 2020-07-27]
14. National Survey on Drug Use and Health (NSDUH-2008). SAMHSA. 2018. URL: https://www.datafiles.samhsa.gov/study/national-survey-drug-use-and-health-nsduh-2008-nid13602 [accessed 2020-07-14]
15. Scott-Sheldon LA, Carey KB, Cunningham K, Johnson BT, Carey MP, MASH Research Team. Alcohol use predicts sexual decision-making: a systematic review and meta-analysis of the experimental literature. AIDS Behav 2016 Jan;20(Suppl 1):S19-S39 [FREE Full text] [doi: 10.1007/s10461-015-1108-9] [Medline: 26080689]
16. Moreland AD, Walsh K, Hartley C, Hanson R, Danielson CK, Saunders B, et al. Investigating longitudinal associations between sexual assault, substance use, and delinquency among female adolescents: results from a nationally representative sample. J Adolesc Health 2018 Sep;63(3):320-326 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.04.002] [Medline: 30029849]
17. Kann L, McManus T, Harris WA, Shanklin SL, Flint KH, Queen B, et al. Youth risk behavior surveillance - United States, 2017. MMWR Surveill Summ 2018 Jun 15;67(8):1-114 [FREE Full text] [doi: 10.15585/mmwr.ss6708a1] [Medline: 29902162]
18. How STDs Impact Women Differently From Men. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchhstp/newsroom/docs/factsheets/stds-women.pdf?s_CID=tw_STD0170644 [accessed 2020-07-14]
19. Hall HI, Li J, McKenna MT. HIV in predominantly rural areas of the United States. J Rural Health 2005;21(3):245-253. [doi: 10.1111/j.1748-0361.2005.tb00090.x] [Medline: 16092299]
20. DiClemente RJ, Wingood GM, Harrington KF, Lang DL, Davies SL, Hook EW, et al. Efficacy of an HIV prevention intervention for African American adolescent girls: a randomized controlled trial. J Am Med Assoc 2004 Jul 14;292(2):171-179. [doi: 10.1001/jama.292.2.171] [Medline: 15249566]
21. Miller W. Motivational Enhancement Therapy Manual: A Clinical Research Guide for Therapists Treating Individuals With Alcohol Abuse and Dependence. National Institute on Alcohol Abuse and Alcoholism (NIAAA). 1995. URL: https://pubs.niaaa.nih.gov/publications/projectmatch/match02.pdf [accessed 2020-07-27]
22. Hale DR, Fitzgerald-Yau N, Viner RM. A systematic review of effective interventions for reducing multiple health risk behaviors in adolescence. Am J Public Health 2014 May;104(5):e19-e41. [doi: 10.2105/AJPH.2014.301874] [Medline: 24625172]
23. Crockett LJ, Raffaelli M, Shen Y. Linking self-regulation and risk proneness to risky sexual behavior: pathways through peer pressure and early substance use. J Research on Adolescence 2006 Dec;16(4):503-525. [doi: 10.1111/j.1532-7795.2006.00505.x]
24. Tull M, Weiss N, Adams C, Gratz K. The contribution of emotion regulation difficulties to risky sexual behavior within a sample of patients in residential substance abuse treatment. Addict Behav 2012 Oct;37(10):1084-1092 [FREE Full text] [doi: 10.1016/j.addbeh.2012.05.001] [Medline: 22658304]
25. Brown LK, Houck C, Lescano C, Donenberg G, Tolou-Shams M, Mello J. Affect regulation and HIV risk among youth in therapeutic schools. AIDS Behav 2012 Nov;16(8):2272-2278 [FREE Full text] [doi: 10.1007/s10461-012-0220-3] [Medline: 22669595]
26. McKee K, Russell M, Mennis J, Mason M, Neale M. Emotion regulation dynamics predict substance use in high-risk adolescents. Addict Behav 2020 Jul;106:106374. [doi: 10.1016/j.addbeh.2020.106374] [Medline: 32151894]
27. Houck CD, Barker DH, Hadley W, Menefee M, Brown LK. Sexual risk outcomes of an emotion regulation intervention for at-risk early adolescents. Pediatrics 2018 Jun;141(6):- [FREE Full text] [doi: 10.1542/peds.2017-2525] [Medline: 29748192]
28. Houck CD, Hadley W, Barker D, Brown LK, Hancock E, Almy B. An emotion regulation intervention to reduce risk behaviors among at-risk early adolescents. Prev Sci 2016 Jan;17(1):71-82 [FREE Full text] [doi: 10.1007/s11121-015-0597-0] [Medline: 26297499]

29. Hadley W, Houck C, Brown LK, Spitalnick JS, Ferrer M, Barker D. Moving beyond role-play: evaluating the use of virtual reality to teach emotion regulation for the prevention of adolescent risk behavior within a randomized pilot trial. J Pediatr Psychol 2019 May 1;44(4):425-435 [FREE Full text] [doi: 10.1093/jpepsy/jsy092] [Medline: 30551157]
30. Glasser M, Holt N, Hall K, Mueller B, Norem J, Pickering J, et al. Meeting the needs of rural populations through interdisciplinary partnerships. Fam Community Health 2003;26(3):230-245. [doi: 10.1097/00003727-200307000-00008] [Medline: 12829945]
31. Bowen AM, Horvath K, Williams ML. A randomized control trial of Internet-delivered HIV prevention targeting rural MSM. Health Educ Res 2007 Feb;22(1):120-127 [FREE Full text] [doi: 10.1093/her/cyl057] [Medline: 16849391]
32. Acierno R, Knapp R, Tuerk P, Gilmore AK, Lejuez C, Ruggiero K, et al. A non-inferiority trial of prolonged exposure for posttraumatic stress disorder: in person versus home-based telehealth. Behav Res Ther 2017 Feb;89:57-65 [FREE Full text] [doi: 10.1016/j.brat.2016.11.009] [Medline: 27894058]
33. López CM, Qanungo S, Jenkins CM, Acierno R. Technology as a means to address disparities in mental health research: a guide to 'tele-tailoring' your research methods. Prof Psychol Res Pr 2018 Feb;49(1):57-64 [FREE Full text] [doi: 10.1037/pro0000176] [Medline: 30034085]
34. Skinner HA. The drug abuse screening test. Addict Behav 1982 Jan;7(4):363-371. [doi: 10.1016/0306-4603(82)90005-3]
35. Wingood GM, DiClemente RJ. The ADAPT-ITT model: a novel method of adapting evidence-based HIV Interventions. J Acquir Immune Defic Syndr 2008 Mar 1;47(Suppl 1):S40-S46. [doi: 10.1097/QAI.0b013e3181605df1] [Medline: 18301133]
36. Wingood GM, Simpson-Robinson L, Braxton ND, Raiford JL. Design of a faith-based HIV intervention: successful collaboration between a university and a church. Health Promot Pract 2011 Nov;12(6):823-831. [doi: 10.1177/1524839910372039] [Medline: 21511996]
37. Atienzo EE, Walker DM, Campero L, Lamadrid-Figueroa H, Gutiérrez JP. Parent-adolescent communication about sex in Morelos, Mexico: does it impact sexual behaviour? Eur J Contracept Reprod Health Care 2009 Apr;14(2):111-119. [doi: 10.1080/13625180802691848] [Medline: 19340706]
38. Yu S, Clemens R, Yang H, Li X, Stanton B, Deveaux L, et al. Youth and parental perceptions of parental monitoring and parent-adolescent communication, youth depression, and youth risk behaviors. Soc Behav Pers 2006 Jan 1;34(10):1297-1310. [doi: 10.2224/sbp.2006.34.10.1297]
39. Forcino RC, Barr PJ, O'Malley AJ, Arend R, Castaldo MG, Ozanne EM, et al. Using CollaboRATE, a brief patient-reported measure of shared decision making: results from three clinical settings in the United States. Health Expect 2018 Feb;21(1):82-89 [FREE Full text] [doi: 10.1111/hex.12588] [Medline: 28678426]
40. Yudko E, Lozhkina O, Fouts A. A comprehensive review of the psychometric properties of the drug abuse screening test. J Subst Abuse Treat 2007 Mar;32(2):189-198. [doi: 10.1016/j.jsat.2006.08.002] [Medline: 17306727]
41. Saunders J, Aasland O, Babor T, de la Fuente JR, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
42. AUDIT: The Alcohol Use Disorders Identification Test : Guidelines for Use in Primary Health Care. World Health Organization. 2001. URL: https://apps.who.int/iris/handle/10665/67205 [accessed 2020-07-27]
43. Steinberg AM, Brymer MJ, Decker KB, Pynoos RS. The university of California at Los Angeles post-traumatic stress disorder reaction index. Curr Psychiatry Rep 2004 Apr;6(2):96-100. [doi: 10.1007/s11920-004-0048-2] [Medline: 15038911]
44. Steinberg AM, Brymer MJ, Kim S, Briggs EC, Ippen CG, Ostrowski SA, et al. Psychometric properties of the UCLA PTSD reaction index: part I. J Trauma Stress 2013 Feb;26(1):1-9. [doi: 10.1002/jts.21780] [Medline: 23417873]
45. Gullone E, Taffe J. The emotion regulation questionnaire for children and adolescents (ERQ-CA): a psychometric evaluation. Psychol Assess 2012 Jun;24(2):409-417. [doi: 10.1037/a0025777] [Medline: 22023559]
46. Gratz KL, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. J Psychopathol Behav Assess 2004 Mar;26(1):41-54. [doi: 10.1023/b:joba.0000007455.08539.94]
47. Boyatzis R. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks, CA: Sage Publications; 1998.
48. Glaser B, Strauss A. The Discovery of Grounded Theory. New York, USA: Alpine; 1967.
49. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
50. DeCuir-Gunby JT, Marshall PL, McCulloch AW. Developing and using a codebook for the analysis of interview data: an example from a professional development research project. Field Methods 2010 Dec 27;23(2):136-155. [doi: 10.1177/1525822x10388468]
51. Weinberg A, Klonsky ED. Measurement of emotion dysregulation in adolescents. Psychol Assess 2009 Dec;21(4):616-621. [doi: 10.1037/a0016669] [Medline: 19947794]
52. Gnambs T, Kaspar K. Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. Behav Res Methods 2015 Dec;47(4):1237-1259. [doi: 10.3758/s13428-014-0533-4] [Medline: 25410404]

53. Lewis MA, Litt DM, Cronce JM, Blayney JA, Gilmore AK. Underestimating protection and overestimating risk: examining descriptive normative perceptions and their association with drinking and sexual behaviors. J Sex Res 2014;51(1):86-96 [FREE Full text] [doi: 10.1080/00224499.2012.710664] [Medline: 23067203]
54. Cosden M, Morrison G, Gutierrez L, Brown M. The effects of homework programs and after-school activities on school success. Theor Pract 2004 Aug;43(3):220-226. [doi: 10.1207/s15430421tip4303_8]
55. Howie LD, Lukacs SL, Pastor PN, Reuben CA, Mendola P. Participation in activities outside of school hours in relation to problem behavior and social skills in middle childhood. J Sch Health 2010 Mar;80(3):119-125. [doi: 10.1111/j.1746-1561.2009.00475.x] [Medline: 20236413]
56. George WH, Stoner SA. Understanding acute alcohol effects on sexual behavior. Annu Rev Sex Res 2000;11:92-124. [Medline: 11351836]
57. Hendershot CS, George WH. Alcohol and sexuality research in the AIDS era: trends in publication activity, target populations and research design. AIDS Behav 2007 Mar;11(2):217-226 [FREE Full text] [doi: 10.1007/s10461-006-9130-6] [Medline: 16897352]
58. Testa M, Livingston JA. Alcohol consumption and women's vulnerability to sexual victimization: can reducing women's drinking prevent rape? Subst Use Misuse 2009;44(9-10):1349-1376 [FREE Full text] [doi: 10.1080/10826080902961468] [Medline: 19938922]
59. Weissman S, Duffus WA, Vyavaharkar M, Samantapudi AV, Shull KA, Stephens TG, et al. Defining the rural HIV epidemic: correlations of 3 definitions--South Carolina, 2005-2011. J Rural Health 2014;30(3):275-283. [doi: 10.1111/jrh.12057] [Medline: 24329575]
60. Rural-Urban Commuting Area Codes. The Economics of Food, Farming, Natural Resources, and Rural America. URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/ [accessed 2020-04-13]

Abbreviations

AUDIT: alcohol use disorders identification test
CR: cognitive reappraisal
DAST: drug abuse screen test
DERS: difficulty in emotion regulation scale
ERQ: emotion regulation questionnaire
ES: expressive suppression
IRB: institutional review board
PI: principal investigator
RUCA: rural urban commuting area
SIHLE: Sisters Informing, Healing, Living, and Empowering
STD: sexually transmitted disease
STI: sexually transmitted infection
SUD: substance use disorder
UCLA-RI: University of California at Los Angeles Posttraumatic Stress Disorder Child/Adolescent Reaction Index

Edited by G Eysenbach; submitted 17.10.19; peer-reviewed by K Walsh, K Hochstatter; comments to author 13.01.20; revised version received 29.05.20; accepted 03.06.20; published 11.08.20.

Please cite as:
Lopez C, Gilmore AK, Moreland A, Danielson CK, Acierno R
Meeting Kids Where They Are At–A Substance Use and Sexual Risk Prevention Program via Telemedicine for African American Girls: Usability and Acceptability Study
J Med Internet Res 2020;22(8):e16725
URL: http://www.jmir.org/2020/8/e16725/
doi:10.2196/16725
PMID:32780022

Original Paper

Effects of a Brief Electronic Mindfulness-Based Intervention on Relieving Prenatal Depression and Anxiety in Hospitalized High-Risk Pregnant Women: Exploratory Pilot Study

Maren Goetz[1,2*], MD; Claudia Schiele[1*]; Mitho Müller[3], DipPsych; Lina M Matthies[1], MD; Thomas M Deutsch[1], MD; Claudio Spano[3]; Johanna Graf[4], DipPsych; Stephan Zipfel[4], PhD; Armin Bauer[5]; Sara Y Brucker[6], PhD; Markus Wallwiener[1*], PhD; Stephanie Wallwiener[1*], PhD

[1]Department of Obstetrics and Gynecology, University of Heidelberg, Heidelberg, Germany

[2]Department of General Pediatrics, University Children's Hospital, Heidelberg, Germany

[3]Department of Psychology, Ludwig Maximilian University, Munich, Germany

[4]Department of Psychosomatic Medicine and Psychotherapy, University Hospital Tuebingen, Tuebingen, Germany

[5]Research Institute for Women's Health, Department of Women's Health, University Hospital Tuebingen, Tuebingen, Germany

[6]Department of Women's Health, University Hospital Tuebingen, Tuebingen, Germany

[*]these authors contributed equally

Corresponding Author:
Stephanie Wallwiener, PhD
Department of Obstetrics and Gynecology
University of Heidelberg
Im Neuenheimer Feld 440
Heidelberg, 69120
Germany
Phone: 49 6221 5636956
Email: stephanie.wallwiener@googlemail.com

Abstract

Background: Peripartum depression and anxiety disorders are highly prevalent and are correlated with adverse maternal and neonatal outcomes. Antenatal care in Germany does not yet include structured screening and effective low-threshold treatment options for women facing peripartum depression and anxiety disorders. Mindfulness-based interventions (MBIs) are increasingly becoming a focus of interest for the management of such patients. Studies have shown a decrease in pregnancy-related stress and anxiety in expectant mothers following mindfulness programs.

Objective: The aim of this study was to explore the clinical effectiveness of a 1-week electronic course of mindfulness on prenatal depression and anxiety in hospitalized, high-risk pregnant women. We hypothesized that participating in a 1-week electronic MBI (eMBI) could alleviate symptoms of depression and anxiety during the hospital stay.

Methods: A prospective pilot study with an explorative study design was conducted from January to May 2019 in a sample of 68 women hospitalized due to high-risk pregnancies. After enrolling into the study, the participants were given access to an eMBI app on how to deal with stress, anxiety, and symptoms of depression. Psychometric parameters were assessed via electronic questionnaires comprising the Edinburgh Postnatal Depression Scale (EPDS), State-Trait Anxiety Inventory (STAI-S), and abridged version of the Pregnancy-Related Anxiety Questionnaire (PRAQ-R).

Results: We observed a high prevalence of peripartum depression and anxiety among hospitalized high-risk pregnant women: 39% (26/67) of the study participants in the first assessment and 41% (16/39) of the participants in the second assessment achieved EPDS scores above the cutoff value for minor/major depression. The number of participants with anxiety levels above the cutoff value (66% [45/68] of the participants in the first assessment and 67% [26/39] of the participants in the second assessment) was significantly more than that of the participants with anxiety levels below the cutoff value, as measured with the STAI-S. After completing the 1-week electronic course on mindfulness, the participants showed a significant reduction in the mean state anxiety levels (P<.03). Regarding pregnancy-related anxiety, participants who completed more than 50% of the 1-week course showed lower scores in PRAQ-R in the second assessment (P<.05). No significant changes in the EPDS scores were found after completing the intervention.

Conclusions: Peripartum anxiety and depression represent a relevant health issue in hospitalized pregnant patients. Short-term eMBIs could have the potential to reduce anxiety levels and pregnancy-related anxiety. However, we observed that compliance to eMBI seems to be related to lower symptoms of pregnancy-related stress among high-risk patients. eMBIs represent accessible mental health resources at reduced costs and can be adapted for hospitalized patients during pregnancy.

(J Med Internet Res 2020;22(8):e17593) doi:10.2196/17593

KEYWORDS

pregnancy; high-risk pregnancy; hospitalization; preterm labor; anxiety; depression; psychological stress; mindfulness; stress reduction; mobile app

Introduction

Mental disorders are highly prevalent during pregnancy and impose a major burden for the expectant mother, her environment, and the health care system. The prevalence rate of depression has been reported to be 11%-17% in women during the peripartum period, depending on the gestational age [1,2]. For instance, using validated screening instruments, Bennett et al found that depression was prevalent in 11.1% of the women in their third trimester [1]. Apart from depression, many pregnant women show symptoms of pregnancy-related stress and anxiety. Recently, the prevalence rate of antenatal and postnatal anxiety disorders was reported to be approximately 15% [3-5]. The reasons for developing a mental disorder during the peripartum period are still not sufficiently understood. However, it is clear that pregnancy and the puerperium period are times of particular vulnerability and emotional distress. Popular interpretations of what pregnancy should be like, for example, that pregnancy is a happy time when women enjoy the satisfaction of fulfilling a valuable reproductive role in the society, negatively affects those who are already vulnerable to distress and low moods [6,7].

Hormonal changes in pregnant women may play a major role in their emotional well-being, and associations of these hormonal changes with prolactin, steroids, and cortisol levels have been discussed previously [8,9]. If complications such as preterm labor develop during pregnancy, it can be assumed that maternal emotional distress will increase further. Worries about the course of pregnancy and the child's health are common burdens and contribute to negative psychological reactions such as anxiety or emotional lability [10,11]. Hence, it is not surprising that available research on women hospitalized with high-risk pregnancies reports rates of anxiety and depression of up to 40% [12-14].

Peripartum mental disorders have been identified as a potential risk factor for adverse obstetric, fetal, and neonatal outcomes. Studies suggest that there is a link between untreated symptoms of depression, anxiety, or stress and increased rates of birth complications, preterm birth, and fetal or infant growth impairment [15,16]. Considering the high prevalence of mental disorders and their adverse consequences, it is all the more surprising that antenatal care does not yet include structured screening and effective treatment options for women facing this problem. In Germany, the mental health state of pregnant women so far has only been taken into account in regular care by an entry in the maternity record labeled as "mental distress." However, this is based on a subjective assessment of the attending gynecologist and is empirically not proven. Indeed, hospitalized pregnant women are not routinely screened or offered psychological support either. Thus, mental disorders during pregnancy are overlooked in up to 80% of cases, and only 20% of those affected receive appropriate treatment [17]. In addition, affected women are difficult to reach, even with a correct diagnosis, and adequate treatment of peripartum mental disorders is particularly challenging, as drug therapy is often rejected for fear of harming the fetus [18].

Interest in mindfulness-based programs has increased substantially during the last 2 decades [19]. Several potential mechanisms underlying the efficacy of mindfulness-based interventions (MBIs) have been proposed, including exploring internal experiences such as cognitions, emotions, and sensations, affect regulation, decision-making, self-management, and relaxation [20,21]. Recent studies on MBI for pregnant women have shown generally positive effects, including decline in the symptoms of depression, anxiety, pregnancy-related stress, and increased childbirth self-efficacy [22-25].

Although the benefits of MBIs are well supported, less attention has been paid to the potential harm of MBIs. Recent studies and popular media articles suggest that mindfulness or meditation practices might have negative effects such as increased anxiety and unpleasant experiences [26,27]. However, based on the scientific literature, this aspect of MBI is not yet sufficiently understood.

Incorporating mindfulness programs in the prenatal care structure could offer vulnerable pregnant women a stigma-free strategy for addressing these issues [24]. The stigma attached to mental illness represents the major barrier to disclosure and to seeking help in the perinatal period [28]. A recent study of Moore et al showed that many women in the peripartum period were concerned about feeling like or being seen as a "bad mother" if they had a mental illness. They also feared that disclosing symptoms to a health care provider would lead to external stigma. Electronic programs could improve women's disclosure and strengthen treatment uptake and compliance [29].

The general conclusions about the efficacy of electronic MBI (eMBI) programs cannot be drawn according to the current data on mindfulness-based stress reduction during pregnancy. Most of the related research to date lacks methodologically rigorous trials with a randomized-controlled approach [30-32]. A review evaluating the effectiveness of MBI in the perinatal period found that studies tended to focus on healthy rather than the clinical populations [31]. So far, only few studies have included psychiatrically high-risk pregnant women. The results showed

a significant reduction in depression or anxiety scores after the intervention, while mindfulness skills increased. Moreover, the interventions appeared to have a long-term effect on the maternal-fetal attachment domain, which was measured via the maternal-fetal attachment scale [10,33].

Pregnant women engage regularly with digital health technology and they were found to be willing to participate in web-supported perinatal interventions [34]. Indeed, web-based or mobile interventions may represent a promising approach particularly for pregnant women with preterm labor whose mobility is often limited [35]. Nevertheless, studies involving electronic mindfulness programs are sparse. While several meta-analyses have assessed the effectiveness of face-to-face MBIs, evidence supporting the applicability and effectiveness of MBIs when delivered through web-based or mobile devices is clearly lacking [36,37]. A review and meta-analysis of randomized controlled trials (RCTs) estimating the overall effects of electronic MBIs on mental health showed a small but significant beneficial effects on depression, anxiety, well-being, and mindfulness. However, the effect sizes were not significantly related to the study quality [38].

The aim of our study was to investigate the effectiveness of a brief electronic 1-week course of mindfulness on prenatal depression and anxiety in a setting of hospitalized patients with high-risk pregnancies. We hypothesized that attending a 1-week eMBI course can alleviate the symptoms of depression and anxiety during the hospital stay.

Methods

Recruitment of the Participants and Study Design

This pilot study was conducted at the University Women's Hospital in Heidelberg, a perinatal center of the highest level, performing over 2300 deliveries per year. Pregnant women hospitalized due to high-risk pregnancies were asked to participate in the study. The criteria for eligibility included age of 18 years or older, fluency in the German language, a gestational age of ≥24 and ≤34 weeks, and the ability to access to the internet. Women were not eligible to participate if they were expecting multiples. In total, 90 inpatients were asked to participate, of whom 68 women agreed. The reasons for not participating included lack of interest, scheduling conflicts (clash of dates due to lack of time), or severe pregnancy complications. After enrolling in the study, every participant was provided with a tablet and free wireless internet service and access to the eMBI course (Figure 1). The app was designed and developed by an interdisciplinary team of gynecologists, psychologists, and midwives by the Institute for Women's Health Tuebingen, Germany. The study participants took part in a brief 1-week course of mindfulness through an eMBI based on the *mindmom* app, which is currently being examined in a prospective randomized trial (trial registration DRKS 00017210). The RCT aims to examine the clinical effectiveness and cost-effectiveness of an eMBI in a sample of pregnant women during the third trimester of pregnancy who were screened positive for emotional distress according to the Edinburgh Postnatal Depression Scale (EPDS). The screening is administered while attending ambulatory prenatal care either with a registered gynecologist or at one of the study centers. The participants were randomized in 1:1 into the intervention (eMBI) or control (usual treatment) group.

The original intervention was used prenatally in an established concept of 8 weekly 45-min sessions [39]. In this pilot study, the time frame was tightened to a 1-week version, including a total of three 45-min modules on mindfulness (Table 1). The pilot study was carried out to optimize the user-friendliness of the mobile app by making appropriate adjustments. In addition to the questionnaires, semistructured interviews were conducted. Feedback from the participants on user-friendliness and suggestions for improvement (semistructured, qualitative evaluation) were taken into account in the final concept of the content and structure of the mobile app. We chose to use an inpatient population as the convenience sample for a population at heightened risk for mental health conditions and to explore whether eMBIs could represent accessible mental health resources to support hospitalized patients during pregnancy.

Figure 1. Screenshots of the electronic mindfulness-based intervention app. The first screen shows the home screen, the second screen shows the electronic assessment of the patient reported outcomes, and the third screen shows the digital pregnancy counselor.

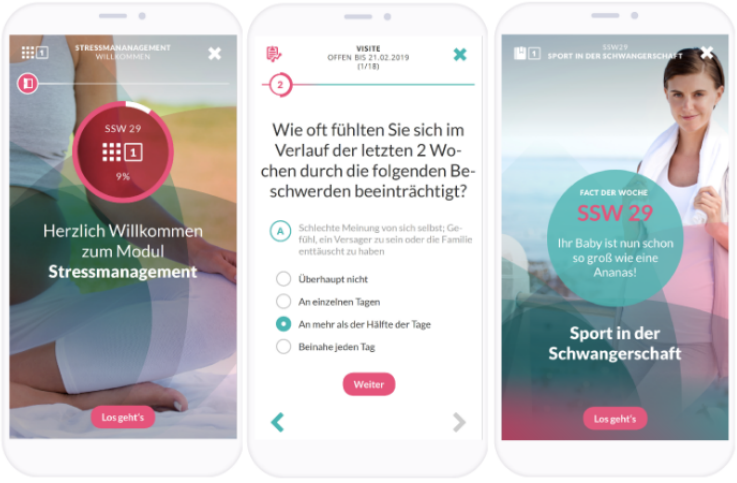


Table 1. Overview of the mindmom app content.

Contents	Module 1 (day 1)	Module 2 (day 3)	Module 3 (day 5)
Topic	Fears and worries about birth and parenting	Coping with stress	Me and my baby
Psychoeducation	Occurrence of pregnancy-related stress, emergence of mental vicious circles, individual sources of strength	Stress and the effects on the body (eg, during pregnancy, birth), favorable conditions for uncomplicated birth process	Positive effects of self-care, caring contact with the baby during pregnancy
Skills	Exit from the vicious circle through your own sources of strength, feel-good place	Encouraging sentences, reward cards	Contact with the baby/positive attitude toward the child, benevolent companions
Mindfulness	Mindful breathing	Mindful body scan	Mindful "loving kindness"

Through mediation of psychoeducational content and cognitive behavioral therapy-related approaches, the app teaches participants how to deal with stress, pregnancy-related anxiety, and symptoms of depression. Thus, it promotes the autonomy of the mother-to-be regarding the upcoming birth and during the initial days and weeks with the newborn. The app contains instructional videos and audio files, interactive worksheets, and a personal "skills box" to collect exercises, videos, and texts, which the participants found helpful. After the 1-week course, participants had the opportunity to continue accessing the exercises. All the participants received scientifically validated information about pregnancy and birth via a pregnancy counselor. The main topics were the physical changes in pregnancy, the birth process, pain relief during birth, bonding between parents and child, breastfeeding, and tips on the formalities related to birth. No financial compensation was offered to the participants. All questionnaires were completed digitally prior to and after the 1-week course. In addition, we gathered medical and sociodemographic data that were double-checked against the hospital records (Table 2). Data were collected via the mindmom app based on electronic patient reported outcomes.

To assess the psychometric data on depression and symptoms of anxiety, the following instruments were used: EPDS, State-Trait Anxiety Inventory (STAI), and Pregnancy-Related Anxiety Questionnaire (PRAQ).

Table 2. Questionnaire assessment.

Visiting schedule, data captured	Questionnaires to be filled
Visit 1 (day 1)	
Sociodemographic data (age, education level, marital status, income, occupation, number of children)	Self-designed[a]
Medical history (gravidity, parity, fertility treatment)	Self-designed
Psychiatric history	Self-designed
Use of the internet	Self-designed
Symptoms of depression and pregnancy-related anxiety	EPDS[b], STAI[c]
Fear of childbirth	PRAQ-R[d]
Visit 2 (day 7)	
Symptoms of depression and pregnancy-related anxiety	EPDS, STAI
Fear of childbirth	PRAQ-R
Feasibility and acceptance of the mindmom app	Self-designed

[a]The questionnaires were not validated as they were self-designed.

[b]EPDS: Edinburgh Postnatal Depression Scale (validated questionnaire).

[c]STAI: State-Trait Anxiety Inventory (validated questionnaire).

[d]PRAQ-R: Pregnancy-Related Anxiety Questionnaire abridged version (validated questionnaire).

EPDS

The EPDS is a 10-item self-rating scale that assesses the depressive symptoms during the peripartum period. It was originally developed by Cox et al in 1987 [40] and translated into German by Bergant et al [41]. The EPDS is used for research purposes and has been proven to be an efficient and effective way of identifying patients at risk for perinatal depression. With a cutoff value of 9 (EPDS>9), the sensitivity of detecting a clinically significant depression is 0.96, the specificity is 1.00, and the positive predictive value is 1.00 [40-42]. The scale reached a good to excellent internal consistency in our sample (Cronbach α=.91 in the first assessment and α=.81 in the second assessment).

STAI

Symptoms of anxiety are assessed with the STAI. The STAI was developed by Spielberger et al [43] and translated into

German by Laux et al [44]. Based on Cattell's theory of anxiety, the STAI consists of 2 scales (STAI-S [STAI-State] and STAI-T [STAI-Trait]), each comprising 20 items to separately assess anxiety as a general trait or as a temporary condition. The 2 scales can be used together or separately. The items are answered on the basis of a 4-point Likert scale; the answers (1-4) are added to the total value, with negative polarized items being reversed. A total value of 20 implies "absolute absence of anxiety" and a maximum score of 80 means "highest level of anxiety" [45]. The STAI was validated for pregnancy by Grant et al [46]. The best cutoff of the STAI for perinatal samples was found to be 40, which matches the original cutoff [43]. The scale reached a good to excellent internal consistency in our sample (STAI-S: Cronbach α=.95 at first and α=.93 at second assessment; STAI-T: Cronbach α=.89 at first and α=.91 at second assessment).

PRAQ

Pregnancy-related anxiety is assessed with the PRAQ. The questionnaire was originally developed by van den Bergh [47] and abridged by Huizink et al (PRAQ-R). It consists of 10 items. The response categories range from "never applicable" to "very strong/very often true" on a 5-point Likert scale. The items are added to a total sum score (maximum score 10), with a higher sum score indicating a higher level of pregnancy-related anxiety. In our sample, the PRAQ-R showed an acceptable to good internal consistency with a Cronbach α of .84 at the first assessment and α of .76 at the second assessment for the whole instrument.

Statistical Analyses

All analyses were conducted using the SPSS software (IBM Corp, v. 24.0.0.0). Since the distributions of the self-report scales did not deviate from the normal distribution (P>.18 in Kolmogorov-Smirnov-test), we chose a parametric analysis strategy for the parametric variables. Due to scale-specific amounts of the missing values, the number of valid cases varies between the analyses. We considered sociodemographic data (eg, age), medical data (eg, gestational age), and the self-report data (eg, EPDS) for this procedure. The missing completely at random test results were not significant (χ^2_{341}=334.4, P=.59), indicating that missing values were at random and that subpopulations were representative for the total sample [48].

First, the descriptive statistics of the sample characteristics in the relevant sociodemographic, medical, and self-reported variables were reported according to the levels measured. Second, the frequency of the cases scoring above the EPDS cutoff (>9) were compared to the expected frequency (11.1%) according to the review of Bennet et al [1] by using the chi-square test. The same analysis was used to determine the frequency of the women scoring above and below the STAI cutoff (>40) for equality. Third, the EPDS, STAI, and PRAQ-R scores were tested for a decline between the first and second measurements by using two-tailed t tests for paired samples. Finally, high versus low treatment compliance was tested for differences regarding the symptom levels by using t tests for independent means. The two-sided critical α-error was set to α=.05. Due to the exploratory nature of the analyses, the α errors were not Bonferroni-adjusted. To estimate the effect sizes, we computed $\omega^2 = \chi^2/N$ for chi-square tests and ω^2 for t tests. ω^2 is a population-based estimator of the explained variance. ω^2=0.01 or ω^2=0.01 is interpreted as small effects, ω^2=0.09 or ω^2=0.06 as medium-sized effects, and ω^2=0.25 or ω^2=0.14 as large effects [49].

Results

Sociodemographic Data, Medical Data, and Self-Reports

In total, 68 hospitalized pregnant women were included in the study, 39 of whom completed the full 1-week electronic mindfulness course. The most common diagnoses of the participants are listed in Table 3. Of the 68 participants, the information for 1 participant regarding the diagnosis for hospitalization was missing; therefore, the information of 67 participants is provided in Table 3.

A total of 29 participants who completed the baseline visit were lost to follow-up. Thus, the overall completion rate was 57% (39/68). The nonparametric sociodemographic characteristics of the study population are summarized in Table 4. The parametric demographic, medical, and self-report data are summarized in Table 5.

Table 3. Cases diagnosed at admission for hospitalization (N=67, multiple answers possible).

Diagnosis at admission	Valid cases, n (%)
Cervical insufficiency	27 (40)
PROM[a]/premature labor	19 (28)
IUGR[b]/oligohydramnios	10 (15)
Vaginal bleeding	8 (12)
Infections	5 (7)
Placental disorder	4 (6)
Fetal malformation	4 (6)
Polyhydramnios	3 (4)
Gestational diabetes	3 (4)
Other	9 (13)

[a]PROM: prelabor rupture of membranes.

[b]IUGR: intrauterine growth restriction.

Table 4. Analysis of the nonparametric sample characteristics.[a]

Nonparametric data	Valid cases, n (%)
Civil status	
Married and living together	49 (73)
Married and living apart	1 (1)
Single	17 (25)
Number of children	
0	40 (59)
1	23 (34)
≥2	5 (7)
Education	
Lower secondary qualification	5 (7)
Higher secondary qualification	22 (32)
University entrance qualification	41 (60)
Occupation	
Unemployed	33 (48)
Part-time	9 (13)
Full-time	26 (38)
Income (1 €=1.15 USD)	
<1500 €	18 (27)
1500-4999 €	39 (59)
>5000 €	9 (14)
Gravidity	
1	30 (44)
2	27 (40)
≥3	11 (16)
Parity	
0	39 (58)
1	23 (34)
≥2	5 (7)
Outpatient psychiatric psychotherapeutic treatment	
Never	53 (78)
Earlier	14 (21)
Current	1 (1)
Inpatient psychiatric/psychotherapeutic treatment	
Never	60 (91)
Earlier	5 (8)
Current	1 (2)
Current mental illness	
No	64 (97)
Yes	2 (3)
Mental illness in family	
No	50 (75)
Yes	17 (25)

Nonparametric data	Valid cases, n (%)
Former medication for states of depression or anxiety[b]	
No	57 (84)
Yes	11 (16)
Current or past prepartum anxiety disorder	
No	62 (91)
Yes	6 (9)
Current or past postpartum anxiety disorder	
No	66 (99)
Yes	1 (1)
Current or past postpartum depression[b]	
No	66 (98)
Yes	1 (1)

[a]The total valid number of cases varied between 66 and 68.

[b]There were no cases with current or past prepartum depression or of those on medications.

Table 5. Analysis of the parametric sample characteristics.

Parametric data	Sample size (n)	Range	Mean (SD)	Standard error
Maternal age (years)	68	22-41	32.07 (4.74)	0.58
Weight before pregnancy (kg)	68	30-138	70.32 (19.77)	2.40
Body height (cm)	68	150-183	165.19 (6.93)	0.84
Current weight (kg)	68	42-146	79.87 (19.11)	2.32
Gestation age (weeks)	63	24-34	30.17 (3.17)	0.40
EPDS[a] score (T1)[b]	67	0-25	8.39 (5.59)	0.68
EPDS score (T2)[c]	39	3-20	8.62 (4.13)	0.66
STAI-S[d] score (T1)	68	28-79	46.66 (11.54)	1.40
STAI-S score (T2)	39	29-65	43.81 (10.09)	1.62
STAI-T[e] score (T1)	68	22-53	38.18 (8.01)	0.97
STAI-T score (T2)	39	24-54	38.43 (8.46)	1.36
PRAQ-R[f] score (T1)	68	10-40	22.83 (7.31)	0.89
PRAQ-R score (T2)	39	10-39	20.69 (6.09)	0.98

[a]EPDS: Edinburgh Postnatal Depression Scale.

[b]T1: first assessment.

[c]T2: second assessment.

[d]STAI-S: State-Trait Anxiety Inventory (State scale).

[e]STAI-T: State-Trait Anxiety Inventory (Trait scale).

[f]PRAQ-R: Pregnancy-Related Anxiety Questionnaire abridged version.

Tests on the Distribution of the Cases Below and Above Cutoffs

Table 6 shows the observed and expected case numbers for the EPDS, the STAI-S, and the STAI-T cutoffs at both assessments. The chi-square tests were highly significant for the EPDS at both assessments (P<.001) with large empirical effects (ω^2=0.778 at first and ω^2=0.908 at second assessment). With 39% (26/67) of the participants at the first assessment and 41% (16/39) of the participants at the second assessment scoring above the cutoff (>9), there were significantly more participants with depressive symptoms in the sample than expected (11.1%). The chi-square tests were also significant for the STAI-S at both assessments (P=.008 at first and P=.04 at second assessment) with medium-sized empirical effects (ω^2=0.105 at first and ω^2=0.111 at second assessment). With 66% (45/68) of the

participants at the first assessment and 67% (26/39) of the participants at the second assessment scoring above the cutoff (>40), there were significantly more state anxious participants in the sample than expected (34/68; 50%). The chi-square tests were not significant for the STAI-T at either assessment (*P*=.22 at first and *P*=.63 at second assessment). With 42.6% (29/68) of the participants at the first assessment and 46.2% (18/39) of the participants at the second assessment scoring above the cutoff (>40), the number of trait anxious participants was not significantly higher in the sample than expected (34/68; 50%). The power to find large effects in this analysis was sufficient for large effects (ω^2=0.25, 1–β=.98 at the first and 1–β=.87 at the second assessment). However, medium-sized (ω^2=0.01, 1–β=.70 at the first and 1–β=.47 at the second assessment) and small effects (ω^2=0.01, 1–β=.13 at the first and 1–β=.10 at the second assessment) cannot sufficiently be ruled out.

Table 6. Chi-square tests on the distribution of the cases below and above the cutoffs.

Assessments, cutoffs	Observed cases (n)	Expected cases (n)	Chi-square (*df*)	*P* value
EPDS[a] (T1)[b]			52.1 (1)	<.001
≤9	41	60		
>9	26	7		
EPDS (T2)[c]			35.4 (1)	<.001
≤9	23	35		
>9	16	4		
STAI-S[d] (T1)			7.1 (1)	.008
<40	23	34		
≥40	45	34		
STAI-S (T2)			4.3 (1)	.04
<40	13	20		
≥40	26	20		
STAI-T[e] (T1)			1.5 (1)	.22
<40	39	34		
≥40	29	34		
STAI-T (T2)			0.2 (1)	.63
<40	21	19.5		
≥40	18	19.5		

[a]EPDS: Edinburgh Postnatal Depression Scale.

[b]T1: first assessment.

[c]T2: second assessment.

[d]STAI-S: State-Trait Anxiety Inventory (State scale).

[e]STAI-T: State-Trait Anxiety Inventory (Trait scale).

Tests on the Changes Between the First and Second Assessments

Table 7 summarizes the descriptive and inferential results of the paired sample *t* tests on the changes between the first and the second assessment in the subsample of 39 participants who were eligible at both assessments. No significant change was found in the EPDS, STAI-T, or PRAQ-R scores (*P*>.20). The power to detect large (ω^2=0.14, 1–β>.99) and medium-sized effects (ω^2=0.06, 1–β=0.86) was sufficient for these analyses. At the same time, with a power of 1–β=0.23, small effects (ω^2=0.01) cannot be ruled out. However, the STAI-S scores significantly declined for these patients between the first and the second assessments (*P*=.03) with a small effect of ω^2=0.051 (ie, 5.1% of the variance of the changes between the first and the second assessment can be explained by the passing of time).

Table 7. Results of the paired sample t tests on the changes between the first and second assessments (n=39).

Assessments	Mean (SD)	Standard error	*t (df)*	*P* value
EPDS[a] score			–0.370 (38)	.71
T1[b]	8.41 (4.77)	0.76		
T2[c]	8.62 (4.13)	0.66		
STAI-S[d] score			2.277 (38)	.03
T1	46.65 (11.35)	1.82		
T2	43.81 (10.09)	1.62		
STAI-T[e] score			0.325 (38)	.75
T1	38.60 (7.39)	1.18		
T2	38.43 (8.46)	1.36		
PRAQ-R[f] score			1.317 (38)	.20
T1	21.63 (6.08)	0.97		
T2	20.69 (6.09)	0.98		

[a]EPDS: Edinburgh Postnatal Depression Scale.

[b]T1: first assessment.

[c]T2: second assessment.

[d]STAI-S: State-Trait Anxiety Inventory (State scale).

[e]STAI-T: State-Trait Anxiety Inventory (Trait scale).

[f]PRAQ-R: Pregnancy-Related Anxiety Questionnaire abridged version.

Differences Between High and Low App Engagement Regarding Symptom Levels

To examine the differences between the high and low app engagement regarding the symptom levels at the second assessment, we defined a group of participants that completed more than 50% of the 1-week course modules. Furthermore, we dummy coded noncompliant as “0” and compliant as “1” for >50% of all the 3 modules. There was no significant difference between high (8.96 [SD 4.30]) and low app engagement (8.30 [SD 4.05]) regarding EPDS scores at the second assessment (t_{37}=–0.49, P=.62). Moreover, there were no significant differences regarding the STAI-S scores (t_{37}=–0.60, P=.55) or the STAI-T scores (t_{37}=–0.60, P=.55) at the second assessment between high (STAI-S: 44.82 [SD 10.56]; STAI-T: 39.26 [SD 8.81]) and low app engagement (STAI-S: 42.85 [SD 9.79]; STAI-T: 37.63 [SD 8.27]). The power to find large effects in these analyses was insufficient for large (ω^2=0.14, 1–β=.68), medium-sized (ω^2=0.06, 1–β=.33), and small effects (ω^2=0.01, 1–β=.09). Consequently, the effects of any size cannot sufficiently be ruled out.

However, a significant, medium-sized difference regarding PRAQ-R scores was observed at the second assessment (t_{37}=2.03, P<.05). If participants completed more than 50% of all the modules, they had significantly lower PRAQ-R scores (18.74 [SD 4.49]) than those with low app engagement (22.54 [SD 6.90]) at the second assessment. Overall, 7.4% (ω^2=0.074) of the PRAQ-R score variance at the second assessment can be explained by the group variable “high vs low app engagement.”

Discussion

Principal Results

In our study, the prevalence of depression and anxiety among hospitalized high-risk pregnant women was found to be high. At the baseline assessment, 39% (26/67) of the study participants achieved EPDS scores above the cutoff value for a minor/major depression and 66% (45/68) of the women had high levels of anxiety as measured with the STAI-S. Therefore, our results reinforce the findings of previous studies on depression and anxiety disorders in perinatal populations demonstrating regularly higher EPDS and STAI scores than the expected ranges in nonpregnant populations [50]. In our study, we used a meta-analysis including 19,284 patients for comparison that reported EPDS scores ≥10 in 11.7% and 11.1% of the women in the second and third trimester, respectively [1]. Regarding anxiety symptoms during pregnancy, recent reviews found pooled prevalence rates of 15%-23% across trimesters [5,51]. The study selection of these reviews was restricted to samples of pregnant women recruited through general obstetric/prenatal units. By contrast, our sample included hospitalized high-risk pregnant patients, in whom prevalence rates are considerably higher.

When compared to recently published studies with inpatient samples, we found similar tendencies for anxiety and depression levels during pregnancy. Regarding antenatal depression assessed via the EPDS, a study in Singapore found that the rate of major depression in a sample of high-risk pregnancies (11%) was higher than that in an unspecified obstetric risk cohort (4.3%) [14]. Dagklis et al presented similar results with high

depression rates of 24.3% and 28% in 2 different high-risk pregnancy unit samples [52,53]. The scores in our sample exceed the aforementioned study results considerably, with up to 41% (16/39) of the participants reaching or exceeding the EPDS cutoff value. In a recent study examining depression, anxiety, and attachment among women hospitalized in an antepartum unit, screening identified over one-third (36%) of the participants to be at risk for depression (EPDS score ≥10) and almost half (47%, 46/98) reported elevated state anxiety (STAI-S ≥40) [54]. Likewise, Barber and Starkey showed that hospitalized pregnant women had significantly higher state anxiety levels, with 47% of the women showing STAI-S scores above the cutoff compared to normative data from nonpregnant samples [55]. These results are in line with our sample, with STAI-S scores >40 for 66% (45/68) of the participants.

In direct comparison, it becomes clear that hospitalized pregnant women represent a vulnerable group, showing rates of anxiety and depression more than three times greater than those reported in nonclinical samples. Despite the high rates of anxiety and depressive symptoms, a surprisingly low number of participants were receiving mental health treatment. In inpatient obstetric settings, where access to individual psychotherapy is often extremely limited, eMBIs could provide an easily accessible support option. Incorporating low-threshold eMBIs not only in hospital care but also in outpatient practices could minimize the stigma of starting mental health treatment.

Measurement Tools

For our sample, we decided to choose a cutoff value of >40 for STAI and a cutoff value >9 for EPDS, which is in line with that reported in recent studies on the use of EPDS and STAI as valid screening tools for depression and anxiety during pregnancy. Tendais et al found optimal cutoffs for STAI-S as 40 and EPDS as 9 during pregnancy [56]. Regarding the accuracy of the EPDS in identifying depression and other mental disorders, Howard et al reported a likelihood ratio of 9.8 for the EPDS [3]. The fact that the EPDS also performs well in screening for depression and anxiety in high-risk pregnant women was confirmed by Thiagayson et al, who recommended further psychiatric assessments for women with a score ≥9, as found in our study [14]. For the STAI, Barnett and Parker recommended cutoffs of high (≥40), moderate (32-33), or low (≤25) anxiety on the basis of the mean trait scores of a sample of 94 primiparae [57]. Grant et al found that a cutoff >40 for both state and trait scales yielded optimal sensitivity (80.95%), specificity (79.75%), and positive predictive value (51.5%) to determine cases of anxiety in the third trimester of pregnancy [46]. In our sample, there were 59% (40/68) primiparae, and the mean gestational age was 30.14 weeks; thus, the values compare well. In all, the EPDS and STAI prove to be effective screening tools that are frequently used in studies including high-risk pregnant women.

Effectiveness of eMBI in Reducing Anxiety

The latest reviews about eMBI during pregnancy suggest that web-based interventions targeted at improving mental health may be beneficial during the peripartum period. However, the findings and their generalizability are limited both by the heterogeneity of the interventions and study designs and by methodological limitations. Pooled results of non-RCTs reporting outcomes on anxiety, depression, and perceived stress showed a significant benefit for the mindfulness group, but this review found no differences between the mindfulness intervention and control groups in RCTs [58]. Another review showed significant reductions in depression, anxiety, and stress by means of preanalyses and postanalyses, each with small to medium effect sizes [31]. However, between-group analyses failed to find any significant postintervention benefits of MBIs in comparison to control conditions.

In our pilot study without a comparison group, a significant reduction in the mean state anxiety levels was found after completing the 1-week eMBI (P<.03). Recent studies including other MBIs during pregnancy found similar effects with significantly lower STAI scores (P<.001) after interventions such as yoga, music therapy, or progressive muscle relaxation [50,59]. Compassion-focused therapy represents another promising approach for effectively treating depression and anxiety in perinatal populations [60]. The equivalence of compassionate mind training on the constructs of self-confidence, inadequate self-criticism, and self-compassion and the superior performance of compassionate mind training in reducing depression and anxiety scores in the latest studies suggest that compassion-focused therapy offers additional benefits beyond the current gold standard of cognitive behavioral therapy [61,62].

App-Related Patient Engagement

As one of our key results, we found a small, yet significant reduction in the mean state anxiety levels after completing the 1-week electronic course of mindfulness (P<.03). Concerning engagement with the app, participants who completed more than 50% of all the 3 modules showed lower scores for the PRAQ-R at the second assessment (P<.05). Our sample population showed an overall completion rate of 57% (39/68). We decided to consider an individual completion rate of more than 50% of the overall module in at least 2 of 3 course modules as compliant, as similar cutoffs have been used before. For instance, previous studies on mindfulness-based programs have considered participants who completed at least 50% of sessions as having fulfilled an adequate minimum amount of the course [30]. Hence, our compliance rate is comparable to that reported in similar studies, which reported even lower compliance rates from 21% to 35% [30,63]. Regarding the nature of this very specific study population, the majority of the dropouts can be explained by hospital discharge before the second assessment. Women who are no longer considered to be at high risk of preterm delivery might not see any reason to continue the eMBI, as they have already experienced symptom relief due to the end of hospitalization. Nonetheless, the overall dropout rate of the course was high (29/68, 43%). The reasons for dropout included discharge from the hospital, actually giving birth prematurely, severe stress and concerns, or adverse pregnancy complications. Studies have shown that although face-to-face mindfulness courses for pregnancy have demonstrated good levels of adherence to the course with completion rates of over 85% [39], adherence after online courses tended to be more difficult. The reasons for low levels of motivation might include rare contact

with the treatment team/researchers or a lack of social support from the other participants [30].

Limitations

As our pilot study design is not an RCT, the changes in the mean state anxiety levels after the 1-week course of mindfulness cannot be clearly attributed to the eMBI. Urech et al [35] assessed the efficacy of a web-based cognitive behavioral stress management training in women with preterm labor and found no significant differences in the psychological parameters between the intervention and the control group (based on distraction). They hypothesized that the psychological well-being improves automatically during the course of pregnancy since participants in both groups showed lower stress-related psychological levels [35]. However, another study suggested that differences in the mean change of the STAI scores between the intervention and control groups were not solely due to feelings and symptoms changing as gestation progresses [50]. Considering the length of the eMBI, most studies report an average duration of 6-8 weeks of intervention [23,39]. In comparison, our 1-week electronic mindfulness course may seem rather short. However, a course duration of 8 weeks would be unsuitable for inpatients. Newham et al already proved that even a single session of yoga reduced both the subjective and physiological measures of state anxiety (STAI-S and cortisol), and this class-induced reduction in anxiety remained at the final session of the intervention [64]. Likewise, studies examining the immediate effects of an intervention reported significantly lowered STAI scores after a single session of complementary therapy-based interventions. The mean post-intervention scores were consistently significantly lower than the baseline intervention scores after all single-session interventions [50]. These facts underscore our results and demonstrate that short-term interventions can be beneficial, especially for those who have the greatest need. However, the small sample size limits the statistical power to detect the effects as well as the generalizability. The large number of statistical tests used in this study increases the global α error; thus, random effects cannot be ruled out sufficiently.

Conclusions

Peripartum anxiety and depression are highly relevant health issues in hospitalized pregnant patients. Despite high prevalence rates and the patients' need for regular mental health support, only a very small proportion receives adequate treatment. In this pilot study, we could show that our 1-week course of mindfulness may be an effective brief intervention that could have the potential to positively influence childbirth anxiety and pregnancy-related stress among high-risk patients. eMBIs may provide an easily accessible and effective means to support pregnant women in managing anxiety and stress during their inpatient stay and beyond. RCTs with confirmatory analyses in larger samples are needed to confirm the effectiveness of mindfulness in promoting perinatal mental health and to study the potential for harm in mindfulness-based programs.

Conflicts of Interest

None declared.

References

1. Bennett HA, Einarson A, Taddio A, Koren G, Einarson TR. Prevalence of depression during pregnancy: systematic review. Obstet Gynecol 2004 Apr;103(4):698-709. [doi: 10.1097/01.AOG.0000116689.75396.5f] [Medline: 15051562]
2. Underwood L, Waldie K, D'Souza S, Peterson ER, Morton S. A review of longitudinal studies on antenatal and postnatal depression. Arch Womens Ment Health 2016 Dec;19(5):711-720. [doi: 10.1007/s00737-016-0629-1] [Medline: 27085795]
3. Howard LM, Ryan EG, Trevillion K, Anderson F, Bick D, Bye A, et al. Accuracy of the Whooley questions and the Edinburgh Postnatal Depression Scale in identifying depression and other mental disorders in early pregnancy. Br J Psychiatry 2018 Jan;212(1):50-56 [FREE Full text] [doi: 10.1192/bjp.2017.9] [Medline: 29433610]
4. Wallwiener S, Goetz M, Lanfer A, Gillessen A, Suling M, Feisst M, et al. Epidemiology of mental disorders during pregnancy and link to birth outcome: a large-scale retrospective observational database study including 38,000 pregnancies. Arch Gynecol Obstet 2019 Mar;299(3):755-763. [doi: 10.1007/s00404-019-05075-2] [Medline: 30734864]
5. Dennis C, Falah-Hassani K, Shiri R. Prevalence of antenatal and postnatal anxiety: systematic review and meta-analysis. Br J Psychiatry 2017 May;210(5):315-323. [doi: 10.1192/bjp.bp.116.187179] [Medline: 28302701]
6. Staneva A, Bogossian F, Pritchard M, Wittkowski A. The effects of maternal depression, anxiety, and perceived stress during pregnancy on preterm birth: A systematic review. Women Birth 2015 Sep;28(3):179-193. [doi: 10.1016/j.wombi.2015.02.003] [Medline: 25765470]
7. Gawley L, Einarson A, Bowen A. Stigma and attitudes towards antenatal depression and antidepressant use during pregnancy in healthcare students. Adv Health Sci Educ Theory Pract 2011 Dec;16(5):669-679. [doi: 10.1007/s10459-011-9289-0] [Medline: 21431361]
8. Kramer MS, Lydon J, Séguin L, Goulet L, Kahn SR, McNamara H, et al. Stress pathways to spontaneous preterm birth: the role of stressors, psychological distress, and stress hormones. Am J Epidemiol 2009 Jun 01;169(11):1319-1326. [doi: 10.1093/aje/kwp061] [Medline: 19363098]
9. Szpunar MJ, Parry BL. A systematic review of cortisol, thyroid-stimulating hormone, and prolactin in peripartum women with major depression. Arch Womens Ment Health 2018 Apr;21(2):149-161 [FREE Full text] [doi: 10.1007/s00737-017-0787-9] [Medline: 29022126]

10. Weidner K, Bittner A, Junge-Hoffmeister J, Zimmermann K, Siedentopf F, Richter J, et al. A psychosomatic intervention in pregnant in-patient women with prenatal somatic risks. J Psychosomatic Obstet Gynec 2010 Jun 29;31(3):188-198. [doi: 10.3109/0167482x.2010.497233]
11. Sciscione AC. Maternal activity restriction and the prevention of preterm birth. Am J Obstet Gynecol 2010 Mar;202(3):232.e1-232.e5. [doi: 10.1016/j.ajog.2009.07.005] [Medline: 19766979]
12. Kingston D, Janes-Kelley S, Tyrrell J, Clark L, Hamza D, Holmes P, et al. An integrated web-based mental health intervention of assessment-referral-care to reduce stress, anxiety, and depression in hospitalized pregnant women with medically high-risk pregnancies: a feasibility study protocol of hospital-based implementation. JMIR Res Protoc 2015 Jan 16;4(1):e9 [FREE Full text] [doi: 10.2196/resprot.4037] [Medline: 25595167]
13. Byatt N, Hicks-Courant K, Davidson A, Levesque R, Mick E, Allison J, et al. Depression and anxiety among high-risk obstetric inpatients. Gen Hosp Psychiatry 2014 Nov;36(6):644-649 [FREE Full text] [doi: 10.1016/j.genhosppsych.2014.07.011] [Medline: 25149040]
14. Thiagayson P, Krishnaswamy G, Lim ML, Sung SC, Haley CL, Fung DSS, et al. Depression and anxiety in Singaporean high-risk pregnancies - prevalence and screening. Gen Hosp Psychiatry 2013;35(2):112-116 [FREE Full text] [doi: 10.1016/j.genhosppsych.2012.11.006] [Medline: 23265951]
15. Szegda K, Markenson G, Bertone-Johnson ER, Chasan-Taber L. Depression during pregnancy: a risk factor for adverse neonatal outcomes? A critical review of the literature. J Matern Fetal Neonatal Med 2014 Jun;27(9):960-967 [FREE Full text] [doi: 10.3109/14767058.2013.845157] [Medline: 24044422]
16. Grigoriadis S, VonderPorten EH, Mamisashvili L, Tomlinson G, Dennis C, Koren G, et al. The Impact of Maternal Depression During Pregnancy on Perinatal Outcomes. J. Clin. Psychiatry 2013 Apr 15;74(04):e321-e341. [doi: 10.4088/jcp.12r07968]
17. Marcus SM, Heringhausen JE. Depression in Childbearing Women: When Depression Complicates Pregnancy. Primary Care: Clinics in Office Practice 2009 Mar;36(1):151-165. [doi: 10.1016/j.pop.2008.10.011]
18. Vitale SG, Laganà AS, Muscatello MRA, La Rosa VL, Currò V, Pandolfo G, et al. Psychopharmacotherapy in Pregnancy and Breastfeeding. Obstet Gynecol Surv 2016 Dec;71(12):721-733. [doi: 10.1097/OGX.0000000000000369] [Medline: 28005135]
19. Pan W, Gau M, Lee T, Jou H, Liu C, Wen T. Mindfulness-based programme on the psychological health of pregnant women. Women and Birth 2019 Feb;32(1):e102-e109 [FREE Full text] [doi: 10.1016/j.wombi.2018.04.018]
20. Brown KW, Ryan RM, Creswell JD. Mindfulness: Theoretical Foundations and Evidence for its Salutary Effects. Psychological Inquiry 2007 Oct 19;18(4):211-237. [doi: 10.1080/10478400701598298]
21. Baer RA. Mindfulness Training as a Clinical Intervention: A Conceptual and Empirical Review. Clinical Psychology: Science and Practice 2006 May 11;10(2):125-143. [doi: 10.1093/clipsy.bpg015]
22. Khoury B, Lecomte T, Fortin G, Masse M, Therien P, Bouchard V, et al. Mindfulness-based therapy: a comprehensive meta-analysis. Clin Psychol Rev 2013 Aug;33(6):763-771. [doi: 10.1016/j.cpr.2013.05.005] [Medline: 23796855]
23. Dunn C, Hanieh E, Roberts R, Powrie R. Mindful pregnancy and childbirth: effects of a mindfulness-based intervention on women's psychological distress and well-being in the perinatal period. Arch Womens Ment Health 2012 Apr;15(2):139-143. [doi: 10.1007/s00737-012-0264-4] [Medline: 22382281]
24. Duncan LG, Cohn MA, Chao MT, Cook JG, Riccobono J, Bardacke N. Benefits of preparing for childbirth with mindfulness training: a randomized controlled trial with active comparison. BMC Pregnancy Childbirth 2017 May 12;17(1):140 [FREE Full text] [doi: 10.1186/s12884-017-1319-3] [Medline: 28499376]
25. Byrne J, Hauck Y, Fisher C, Bayes S, Schutze R. Effectiveness of a Mindfulness-Based Childbirth Education Pilot Study on Maternal Self-Efficacy and Fear of Childbirth. Journal of Midwifery & Women's Health 2013 Dec 10;59(2):192-197. [doi: 10.1111/jmwh.12075]
26. Schlosser M, Sparby T, Vörös S, Jones R, Marchant NL. Unpleasant meditation-related experiences in regular meditators: Prevalence, predictors, and conceptual considerations. PLoS One 2019;14(5):e0216643 [FREE Full text] [doi: 10.1371/journal.pone.0216643] [Medline: 31071152]
27. Baer R, Crane C, Miller E, Kuyken W. Doing no harm in mindfulness-based programs: Conceptual issues and empirical findings. Clin Psychol Rev 2019 Jul;71:101-114 [FREE Full text] [doi: 10.1016/j.cpr.2019.01.001] [Medline: 30638824]
28. Maloni JA, Przeworski A, Damato EG. Web recruitment and internet use and preferences reported by women with postpartum depression after pregnancy complications. Arch Psychiatr Nurs 2013 Apr;27(2):90-95. [doi: 10.1016/j.apnu.2012.12.001] [Medline: 23540519]
29. Moore D, Ayers S, Drey N. A Thematic Analysis of Stigma and Disclosure for Perinatal Depression on an Online Forum. JMIR Ment Health 2016 May 19;3(2):e18 [FREE Full text] [doi: 10.2196/mental.5611] [Medline: 27197516]
30. Krusche A, Dymond M, Murphy SE, Crane C. Mindfulness for pregnancy: A randomised controlled study of online mindfulness during pregnancy. Midwifery 2018 Oct;65:51-57. [doi: 10.1016/j.midw.2018.07.005] [Medline: 30099285]
31. Lever Taylor B, Cavanagh K, Strauss C. The Effectiveness of Mindfulness-Based Interventions in the Perinatal Period: A Systematic Review and Meta-Analysis. PLoS One 2016;11(5):e0155720 [FREE Full text] [doi: 10.1371/journal.pone.0155720] [Medline: 27182732]
32. Hall HG, Beattie J, Lau R, East C, Anne Biro M. Mindfulness and perinatal mental health: A systematic review. Women Birth 2016 Feb;29(1):62-71. [doi: 10.1016/j.wombi.2015.08.006] [Medline: 26346905]

33. Muzik M, Hamilton SE, Lisa Rosenblum K, Waxler E, Hadi Z. Mindfulness yoga during pregnancy for psychiatrically at-risk women: preliminary results from a pilot feasibility study. Complement Ther Clin Pract 2012 Nov;18(4):235-240 [FREE Full text] [doi: 10.1016/j.ctcp.2012.06.006] [Medline: 23059438]
34. Peragallo UR, Berger AA, Ivins AA, Beckham AJ, Thorp JM, Nicholson WK. Internet Use and Access Among Pregnant Women via Computer and Mobile Phone: Implications for Delivery of Perinatal Care. JMIR Mhealth Uhealth 2015 Mar 30;3(1):e25 [FREE Full text] [doi: 10.2196/mhealth.3347] [Medline: 25835744]
35. Urech C, Scherer S, Emmenegger M, Gaab J, Tschudin S, Hoesli I, et al. Efficacy of an internet-based cognitive behavioral stress management training in women with idiopathic preterm labor: A randomized controlled intervention study. J Psychosom Res 2017 Dec;103:140-146 [FREE Full text] [doi: 10.1016/j.jpsychores.2017.10.014] [Medline: 29167041]
36. Lyzwinski LN, Caffery L, Bambling M, Edirippulige S. A Systematic Review of Electronic Mindfulness-Based Therapeutic Interventions for Weight, Weight-Related Behaviors, and Psychological Stress. Telemed J E Health 2018 Mar;24(3):173-184. [doi: 10.1089/tmj.2017.0117] [Medline: 28885896]
37. Plaza I, Demarzo MMP, Herrera-Mercadal P, García-Campayo J. Mindfulness-based mobile applications: literature review and analysis of current features. JMIR Mhealth Uhealth 2013;1(2):e24 [FREE Full text] [doi: 10.2196/mhealth.2733] [Medline: 25099314]
38. Spijkerman MPJ, Pots WTM, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
39. Dimidjian S, Goodman SH, Felder JN, Gallop R, Brown AP, Beck A. Staying well during pregnancy and the postpartum: A pilot randomized trial of mindfulness-based cognitive therapy for the prevention of depressive relapse/recurrence. J Consult Clin Psychol 2016 Mar;84(2):134-145 [FREE Full text] [doi: 10.1037/ccp0000068] [Medline: 26654212]
40. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [Medline: 3651732]
41. Bergant AM, Nguyen T, Heim K, Ulmer H, Dapunt O. [German language version and validation of the Edinburgh postnatal depression scale]. Dtsch Med Wochenschr 1998 Jan 16;123(3):35-40. [doi: 10.1055/s-2007-1023895] [Medline: 9472218]
42. Wisner KL, Sit DKY, McShea MC, Rizzo DM, Zoretich RA, Hughes CL, et al. Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings. JAMA Psychiatry 2013 May;70(5):490-498 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.87] [Medline: 23487258]
43. Spielberger C, Gorsuch R, Lushene R. Manual for the State-Trait Anxiety Inventory. Palo Alto, California: Consulting Psychologists Press; 1970.
44. Laux L, Glanzmann P, Schaffner P, Spielberger CD. Das State-Trait-Angstinventar (STAI). Weinheim: Beltz Testgesellschaft; 1981.
45. Reck C, Zimmer K, Dubber S, Zipser B, Schlehe B, Gawlik S. The influence of general anxiety and childbirth-specific anxiety on birth outcome. Arch Womens Ment Health 2013 Oct;16(5):363-369. [doi: 10.1007/s00737-013-0344-0] [Medline: 23558948]
46. Grant K, McMahon C, Austin M. Maternal anxiety during the transition to parenthood: a prospective study. J Affect Disord 2008 May;108(1-2):101-111. [doi: 10.1016/j.jad.2007.10.002] [Medline: 18001841]
47. Van den Bergh B. The influence of maternal emotion during pregnancy on fetal and neonatal behavior. Journal of Prenatal & Perinatal Psychology & Health. 1990. URL: https://psycnet.apa.org/record/1991-18227-001 [accessed 2020-05-09]
48. Little RJA. A Test of Missing Completely at Random for Multivariate Data with Missing Values. J American Stat Assoc 1988 Dec;83(404):1198-1202. [doi: 10.1080/01621459.1988.10478722]
49. Cohen J. Statistical Power Analysis for the Behavioral Sciences (rev. ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1988.
50. Newham JJ, Westwood M, Aplin JD, Wittkowski A. State–trait anxiety inventory (STAI) scores during pregnancy following intervention with complementary therapies. Journal of Affective Disorders 2012 Dec;142(1-3):22-30 [FREE Full text] [doi: 10.1016/j.jad.2012.04.027]
51. Sinesi A, Maxwell M, O'Carroll R, Cheyne H. Anxiety scales used in pregnancy: systematic review. BJPsych Open 2019 Jan;5(1):e5 [FREE Full text] [doi: 10.1192/bjo.2018.75] [Medline: 30762504]
52. Dagklis T, Papazisis G, Tsakiridis I, Chouliara F, Mamopoulos A, Rousso D. Prevalence of antenatal depression and associated factors among pregnant women hospitalized in a high-risk pregnancy unit in Greece. Soc Psychiatry Psychiatr Epidemiol 2016 Jul;51(7):1025-1031. [doi: 10.1007/s00127-016-1230-7] [Medline: 27138949]
53. Dagklis T, Tsakiridis I, Chouliara F, Mamopoulos A, Rousso D, Athanasiadis A, et al. Antenatal depression among women hospitalized due to threatened preterm labor in a high-risk pregnancy unit in Greece. J Matern Fetal Neonatal Med 2018 Apr;31(7):919-925. [doi: 10.1080/14767058.2017.1301926] [Medline: 28277908]
54. Nagle-Yang S, Phillips M, Albaugh A, Zhao L, Amin J, Ahmad E, et al. Depression, anxiety, and attachment among women hospitalized on an antepartum unit. Int J Psychiatry Med 2019 Nov;54(6):395-407. [doi: 10.1177/0091217419837071] [Medline: 30924701]
55. Barber CC, Starkey NJ. Predictors of anxiety among pregnant New Zealand women hospitalised for complications and a community comparison group. Midwifery 2015 Sep;31(9):888-896. [doi: 10.1016/j.midw.2015.04.017] [Medline: 25987104]

56. Tendais I, Costa R, Conde A, Figueiredo B. Screening for depression and anxiety disorders from pregnancy to postpartum with the EPDS and STAI. Span J Psychol 2014;17:E7. [doi: 10.1017/sjp.2014.7] [Medline: 25012783]
57. Barnett B, Parker G. Professional and non-professional intervention for highly anxious primiparous mothers. Br J Psychiatry 1985 Mar;146:287-293. [doi: 10.1192/bjp.146.3.287] [Medline: 3986438]
58. Dhillon A, Sparkes E, Duarte RV. Mindfulness-Based Interventions During Pregnancy: a Systematic Review and Meta-analysis. Mindfulness (N Y) 2017;8(6):1421-1437 [FREE Full text] [doi: 10.1007/s12671-017-0726-x] [Medline: 29201244]
59. Garcia-Gonzalez J, Ventura-Miranda MI, Requena-Mullor M, Parron-Carreño T, Alarcon-Rodriguez R. State-trait anxiety levels during pregnancy and foetal parameters following intervention with music therapy. J Affect Disord 2018 May;232:17-22. [doi: 10.1016/j.jad.2018.02.008] [Medline: 29471206]
60. Van Dam NT, Sheppard S, Forsyth J, Earleywine M. Self-compassion is a better predictor than mindfulness of symptom severity and quality of life in mixed anxiety and depression. Journal of Anxiety Disorders 2011:123-130 [FREE Full text] [doi: 10.1016/j.janxdis.2010.08.011]
61. Kelman AR, Stanley ML, Barrera AZ, Cree M, Heineberg Y, Gilbert P. Comparing Brief Internet-Based Compassionate Mind Training and Cognitive Behavioral Therapy for Perinatal Women: Study Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2016 Apr 15;5(2):e65 [FREE Full text] [doi: 10.2196/resprot.5332] [Medline: 27084301]
62. Kelman AR, Evare BS, Barrera AZ, Muñoz RF, Gilbert P. A proof-of-concept pilot randomized comparative trial of brief Internet-based compassionate mind training and cognitive-behavioral therapy for perinatal and intending to become pregnant women. Clin Psychol Psychother 2018 Feb 23:608-619. [doi: 10.1002/cpp.2185] [Medline: 29473698]
63. Homko CJ, Deeb LC, Rohrbacher K, Mulla W, Mastrogiannis D, Gaughan J, et al. Impact of a telemedicine system with automated reminders on outcomes in women with gestational diabetes mellitus. Diabetes Technol Ther 2012 Jul;14(7):624-629 [FREE Full text] [doi: 10.1089/dia.2012.0010] [Medline: 22512287]
64. Newham JJ, Wittkowski A, Hurley J, Aplin JD, Westwood M. Effects of antenatal yoga on maternal anxiety and depression: a randomized controlled trial. Depress Anxiety 2014 Aug;31(8):631-640. [doi: 10.1002/da.22268] [Medline: 24788589]

Abbreviations

eMBI: electronic mindfulness-based intervention
EPDS: Edinburgh Postnatal Depression Scale
MBI: mindfulness-based intervention
RCT: randomized controlled trial
PRAQ-R: Pregnancy-Related Anxiety Questionnaire abridged version
STAI-S: State-Trait Anxiety Inventory (State scale)
STAI-T: State-Trait Anxiety Inventory (Trait scale)

Edited by G Eysenbach; submitted 23.12.19; peer-reviewed by N Farič, J D'Arcey, S Guendelman; comments to author 10.03.20; revised version received 03.06.20; accepted 15.06.20; published 11.08.20.

Please cite as:
Goetz M, Schiele C, Müller M, Matthies LM, Deutsch TM, Spano C, Graf J, Zipfel S, Bauer A, Brucker SY, Wallwiener M, Wallwiener S
Effects of a Brief Electronic Mindfulness-Based Intervention on Relieving Prenatal Depression and Anxiety in Hospitalized High-Risk Pregnant Women: Exploratory Pilot Study
J Med Internet Res 2020;22(8):e17593
URL: https://www.jmir.org/2020/8/e17593
doi:10.2196/17593
PMID:32780023

Original Paper

Understanding the Feasibility, Acceptability, and Efficacy of a Clinical Pharmacist-led Mobile Approach (BPTrack) to Hypertension Management: Mixed Methods Pilot Study

Lorraine R Buis[1], PhD; Dana N Roberson[1], BA; Reema Kadri[1], MLIS; Nicole G Rockey[2,3], PharmD, BCACP; Melissa A Plegue[1], MA; Shivang U Danak[1], MD; Timothy C Guetterman[1], PhD; Melanie G Johnson[3], BS; Hae Mi Choe[2,3], PharmD; Caroline R Richardson[1], MD

[1]Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States

[2]Pharmacy Innovations and Partnerships, University of Michigan Medical Group, Ann Arbor, MI, United States

[3]College of Pharmacy, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Lorraine R Buis, PhD
Department of Family Medicine
University of Michigan
1018 Fuller Street
Ann Arbor, MI, 48104
United States
Phone: 1 734 998 7120
Email: buisl@umich.edu

Abstract

Background: Hypertension is a prevalent and costly burden in the United States. Clinical pharmacists within care teams provide effective management of hypertension, as does home blood pressure monitoring; however, concerns about data quality and latency are widespread. One approach to close the gap between clinical pharmacist intervention and home blood pressure monitoring is the use of mobile health (mHealth) technology.

Objective: We sought to investigate the feasibility, acceptability, and preliminary effectiveness of BPTrack, a clinical pharmacist-led intervention that incorporates patient- and clinician-facing apps to make electronically collected, patient-generated data available to providers in real time for hypertension management. The patient app also included customizable daily medication reminders and educational messages. Additionally, this study sought to understand barriers to adoption and areas for improvement identified by key stakeholders, so more widespread use of such interventions may be achieved.

Methods: We conducted a mixed methods pilot study of BPTrack, to improve blood pressure control in patients with uncontrolled hypertension through a 12-week pre-post intervention. All patients were recruited from a primary care setting where they worked with a clinical pharmacist for hypertension management. Participants completed a baseline visit, then spent 12 weeks utilizing BPTrack before returning to the clinic for follow-up. Collected data from patient participants included surveys pre- and postintervention, clinical measures (for establishing effectiveness, with the primary outcome being a change in blood pressure and the secondary outcome being a change in medication adherence), utilization of the BPTrack app, interviews at follow-up, and chart review. We also conducted interviews with key stakeholders.

Results: A total of 15 patient participants were included (13 remained through follow-up for an 86.7% retention rate) in a single group, pre-post assessment pilot study. Data supported the hypothesis that BPTrack was feasible and acceptable for use by patient and provider participants and was effective at reducing patient blood pressure. At the 12-week follow-up, patients exhibited significant reductions in both systolic blood pressure (baseline mean 137.3 mm Hg, SD 11.1 mm Hg; follow-up mean 131.0 mm Hg, SD 9.9 mm Hg; P=.02) and diastolic blood pressure (baseline mean 89.4 mm Hg, SD 7.7 mm Hg; follow-up mean 82.5 mm Hg, SD 8.2 mm Hg; P<.001). On average, patients uploaded at least one blood pressure measurement on 75% (SD 25%) of study days. No improvements in medication adherence were noted. Interview data revealed areas of improvement and refinement for the patient experience. Furthermore, stakeholders require integration into the electronic health record and a modified clinical workflow for BPTrack to be truly useful; however, both patients and stakeholders perceived benefits of BPTrack when used within the context of a clinical relationship.

Conclusions: Results demonstrate that a pharmacist-led mHealth intervention promoting home blood pressure monitoring and clinical pharmacist management of hypertension can be effective at reducing blood pressure in primary care patients with uncontrolled hypertension. Our data also support the feasibility and acceptability of these types of interventions for patients and providers.

Trial Registration: ClinicalTrials.gov NCT02898584; https://clinicaltrials.gov/ct2/show/NCT02898584

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.8059

(J Med Internet Res 2020;22(8):e19882) doi:10.2196/19882

KEYWORDS

cell phone; mobile phone; hypertension; blood pressure; medication adherence; telemedicine; pharmacists

Introduction

Hypertension is a prevalent and costly burden in the United States, affecting about 116.4 million adults ≥20 years of age, and resulting in approximately $55.9 billion in estimated annual direct and indirect costs from 2014-2015. Poor rates of control exacerbate these concerns; of those affected, only about 50% of people achieve blood pressure control, and another 20% remain unaware of their condition [1]. Hypertension is a key risk factor for heart disease and stroke, which are the first and fifth leading causes of death in the US, respectively [2]. Thus, the identification of strategies to manage hypertension is vital to public health in the US. In 2017, the American College of Cardiology/American Heart Association implemented new hypertension guidelines defining hypertension as ≥130/80 mm Hg, lowering the threshold from the ≥140/90 mm Hg defined by the seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC7). As a result, the prevalence of hypertension among US adults increased from 31.9% to 45.6% [3], making this an even more salient health problem to address.

Clinical pharmacists, who assist patients in managing chronic conditions in primary care clinics [4,5], provide effective management of hypertension [6,7]. Another strategy for hypertension management is home blood pressure monitoring [8-12], although concerns about data quality and latency are widespread in instances where patients maintain paper-based logs for self-monitoring [13-15]. One approach to close the gap between clinical pharmacist intervention and home blood pressure monitoring is the use of mobile health (mHealth) technology. Since about 96% of American adults have a cell phone, 81% have a smartphone, and the rate of smartphone adoption is growing [16], mHealth interventions may be a viable technique to increase the efficacy of clinical pharmacist care and home blood pressure monitoring.

The current mHealth landscape is limited by applications that do not support bidirectional patient-provider communication or automatic transmission of electronic data from home blood pressure monitors in real-time. However, a bidirectional intervention that allows for the immediate upload of electronic data has the potential to increase the number of hypertensive patients a clinical pharmacist could assist, as well as improve the quality of blood pressure management for patients.

The goal of this study was to investigate the feasibility, acceptability, and preliminary effectiveness of BPTrack, a clinical pharmacist-led intervention that makes electronically collected data available to the pharmacist in real-time for hypertension management. The bidirectional intervention supports both home blood pressure monitoring and medication adherence for patients with uncontrolled hypertension. The study also aimed to understand barriers to adoption and areas for improvement identified by key stakeholders so that more widespread use of such interventions may be achieved, and further research can occur.

Methods

The BPTrack study protocol has been described elsewhere [17]; however, key elements are summarized below. All methods used in this study were approved by the University of Michigan Institutional Review Board (HUM00105772).

Study Design

We conducted a one-group design, pre-post pilot study of BPTrack, a clinical pharmacist-led mHealth intervention, intending to improve blood pressure control in patients with uncontrolled hypertension through a 12-week pre-post intervention. All patients were recruited from a primary care setting where they received treatment from a clinical pharmacist at the recruiting clinic site for hypertension management. Participants completed a baseline visit at the recruiting clinic, then spent 12 weeks utilizing the intervention at home before returning for a follow-up visit.

Recruitment

Patient participants were recruited from a Family Medicine clinic associated with a large Midwestern academic medical center. The clinic site serves a majority blue-collar, underserved African American and Hispanic population, and hypertension is a commonly treated chronic disease at the site. Recruitment occurred through two primary procedures: recruitment flyers distributed by clinic staff to potential patient participants and targeted recruitment letters. Further details of these recruitment methods are described elsewhere [17]. All patient participants received $25 cash and were allowed to keep the Bluetooth-enabled blood pressure monitor (Welch Allyn Remote Monitoring Upper Arm Blood Pressure Device RPM-BP100), worth approximately $100, to incentivize study completion. Data from the trial portion of the study were collected between December 2016 and September 2017.

Stakeholder participants were recruited through the purposive sampling of individuals affiliated with the BPTrack program

or health care providers for enrolled patients. Solicitation letters or emails were sent to medical assistants, physicians, a nurse, and the current and former director of the Family Medicine clinic site, as well as the study pharmacist who managed blood pressure care for patient-participants during the study.

Eligibility Screening and Consent

All potential patient participants were screened for eligibility through a phone interview. Once deemed eligible, candidates were scheduled for the baseline visit, where written informed consent was obtained by research staff and baseline data collection commenced.

Inclusion Criteria

To be eligible for participation, patients had to be English speakers, ≥18 years of age, possess a smartphone compatible with the mobile intervention, have a diagnosis and history of uncontrolled hypertension (systolic blood pressure >140 mm Hg and/or diastolic blood pressure >90 mm Hg with repeated measurements), under the care of a primary care physician at the recruiting clinic, and taking at least one antihypertensive medication.

Exclusion Criteria

Exclusion criteria included age >65 years or previously established with a cardiologist or clinical pharmacist for hypertension management. Exclusions also applied to those who were pregnant, had existing medical conditions that would make blood pressure control difficult or required frequent hospitalization. Disqualifying conditions included resistant hypertension, steroid-dependent asthma or emphysema, cirrhosis or hepatic failure, stage C or D chronic heart failure, stage IV or V chronic kidney disease, and terminal cancer or ongoing chemotherapeutic or radiation therapy. Patients were also excluded if they had other serious medical conditions that would inhibit their ability to self-monitor their blood pressure, such as stroke or dementia.

BPTrack Intervention

The BPTrack intervention consisted of two different mobile applications, one for the patient participant and one for the clinical pharmacist, developed by Tactio Health Group and customized from their TactioRPM Platform. BPTrack was the fully automated patient-facing smartphone app for iOS and Android, and BPTrack Pharm was the mobile application for iPad used by the clinical pharmacist. Together, these applications allowed real-time electronic home blood pressure monitoring and medication adherence tracking so that the clinical pharmacist had access to reliable and timely data. As the apps are not publicly available, participants were granted access by study staff after trial enrollment. The apps were provided free to users for use during the trial. Both applications have been described previously [17].

Patients were asked to measure their blood pressure using the provided blood pressure cuff and sync or manually enter the readings into the app. A written manual provided instructions on how to prepare for and properly obtain blood pressure measurements. Patients were encouraged in the written manual and as part of onboarding to reach out to study staff using the study hotline with any questions, comments, or to report any adverse events. The study manual also included safety instructions related to symptoms or repeat measurements indicating hypotensive or hypertensive emergencies.

The clinician-facing app provided the pharmacist with a dashboard view of all enrolled patients and a summary of their recent blood pressure readings, as well as individual page views for each patient, with full access to patient-generated blood pressure data. The clinical pharmacist was instructed to use their best clinical judgment in the interpretation of the BPTrack Pharm data and the appropriate clinical follow-up.

Data Collection

Data were collected in a variety of ways to determine the feasibility and acceptability of this mHealth intervention. Data collection methods are described elsewhere [17] and summarized below.

Patient Surveys

Patients completed investigator-developed surveys at baseline and 12 weeks. The baseline survey collected demographics, health status, hypertension history, self-reported medication adherence and use, and other characteristics. The 12-week follow-up survey also collected self-reported medication adherence and use, as well as perceptions of feasibility, acceptability, and effectiveness of the BPTrack intervention. The baseline and 12-week follow-up survey both took approximately 10 to 15 minutes to complete.

Clinical Measures

Blood pressure and medication adherence, as measured by pill counts, were assessed by research staff in the clinic at baseline and 12 weeks. Blood pressure measurements were gathered using manual blood pressure cuffs by trained research staff. Staff also educated patients on best practices for measuring blood pressure, including sitting upright with feet flat on the ground, keeping the measuring arm at heart height, and not having moved for 5 minutes.

Patient Utilization of the BPTrack App

Patient utilization of the BPTrack app was documented in a variety of ways. Home blood pressure readings were extracted from the BPTrack Secured Cloud and analyzed for blood pressure trends, as well as compliance with self-monitoring protocols. As previously described, patients were asked to take their blood pressure three times per sitting, twice a day. Any text logs sent from within the app from the participants to the pharmacist were extracted as well.

Patient Interviews

To more fully understand patient participant perceptions of the BPTrack program, we conducted semistructured interviews at the 12-week follow-up. Semistructured interviews lasted 2-32 minutes.

Patient Participant Chart Review and Abstraction

We conducted a chart review to abstract data to document patients' health care utilization during the study period. Data abstracted from patient charts included the visit date, diagnoses,

type (phone, in-person, or secure messaging) and reason for the visit, location, whether the visit was with the pharmacist or another provider, medication changes, and blood pressure measurements taken during the encounter.

Stakeholder Interviews

Finally, we invited key stakeholders (physicians, pharmacist, clinic medical director, clinic nurses) to participate in semistructured interviews. These interviews focused on hypertension management, the use of mHealth for hypertension management, how BPTrack was (or could be) used within the clinic, perceived effects of BPTrack, barriers to BPTrack use and implementation, and suggestions for improvement of the program.

Statistical Analysis

Descriptive statistics were compiled for patient characteristics, perceptions of BPTrack, self-reported medication adherence, blood pressure, pill counts, and app and health care utilization. Categorical data were displayed as frequencies and percentages, and chi-square tests were used for comparison. Systolic and diastolic blood pressures were expressed as mean (SD), and pre- and postintervention blood pressure means were compared using 2-tailed paired-samples *t* tests. The effect of clinical pharmacist contact, other health care utilization, and app utilization with changes in blood pressure levels were assessed using Pearson correlations. Available patient characteristics were assessed regardless of study completion; however, only participants who completed the study were included in the analysis of pre- and postmeasures.

Qualitative thematic analysis of patient and stakeholder interviews was performed [18]. We used the concepts of feasibility, acceptability, and effectiveness as a guiding framework for analysis. Two coders analyzed the patient interviews by coding all transcripts, reviewing line-by-line, and resolved initial disagreements to develop an initial codebook. We revised the codebook when clarifications were needed or new categories arose. Finally, we examined code patterns to identify themes related to feasibility, acceptability, and effectiveness. As a validity check, we searched for disconfirming evidence for each theme to challenge themes against the data.

Results

We enrolled 16 patients in the BPTrack pilot study; however, one patient was immediately withdrawn after informed consent, but before collecting study measures, as the home blood pressure cuff did not fit the patient's arm properly. Improper fit of the blood pressure cuff would have led to inaccurate home blood pressure readings. The remaining 15 patients had a mean age of 52.2 years (SD 6.0), and were predominantly male (53.3%; 8/15), married (66.7%; 10/15), employed (73.3%; 11/15), and had a high school or less education (46.7%; 7/15), an annual household income <$50,000 (46.7%; 7/14), and private insurance (60.0%; 9/15). On average, participants had been living with hypertension for 12.1 years (SD 11.2), and the majority were on two antihypertensive medications (53.3%; 8/15). See Table 1 for a complete list of participant demographics. On average, participants had a mean systolic blood pressure of 137.3 mm Hg (SD 11.1) and mean diastolic blood pressure of 89.4 mm Hg (SD 7.7) at baseline. Out of 15 enrolled participants, we lost 2 to follow-up, as we were unable to reach them to complete final data collection assessments (86.7% retention).

Table 1. Participant demographics (N=15).

Characteristic	Value
Age, mean (SD)	52.2 (6.0)
Gender, n (%)	
Female	7 (46.7)
Male	8 (53.3)
Income, n (%)	N=14
<50K	7 (50.0)
50K-100K	3 (21.4)
100K+	4 (28.6)
Race, n (%)	
White	10 (66.7)
Black	3 (20.0)
Other	2 (13.3)
Marital status, n (%)	
Married/living as married	10 (66.7)
Divorced/never married	5 (33.3)
Insurance, n (%)	
Private	9 (60.0)
Medicaid/Medicare	6 (40.0)
Education, n (%)	
High school or less	7 (46.7)
Some college/2-year degree	4 (26.7)
Bachelor's degree +	4 (26.7)
Employment, n (%)	
Employed/self-employed	11 (73.3)
Retired/on disability	4 (26.7)
Baseline health , n (%)	
Very Good	2 (13.3)
Good	8 (53.3)
Fair	4 (26.7)
Poor	1 (6.7)

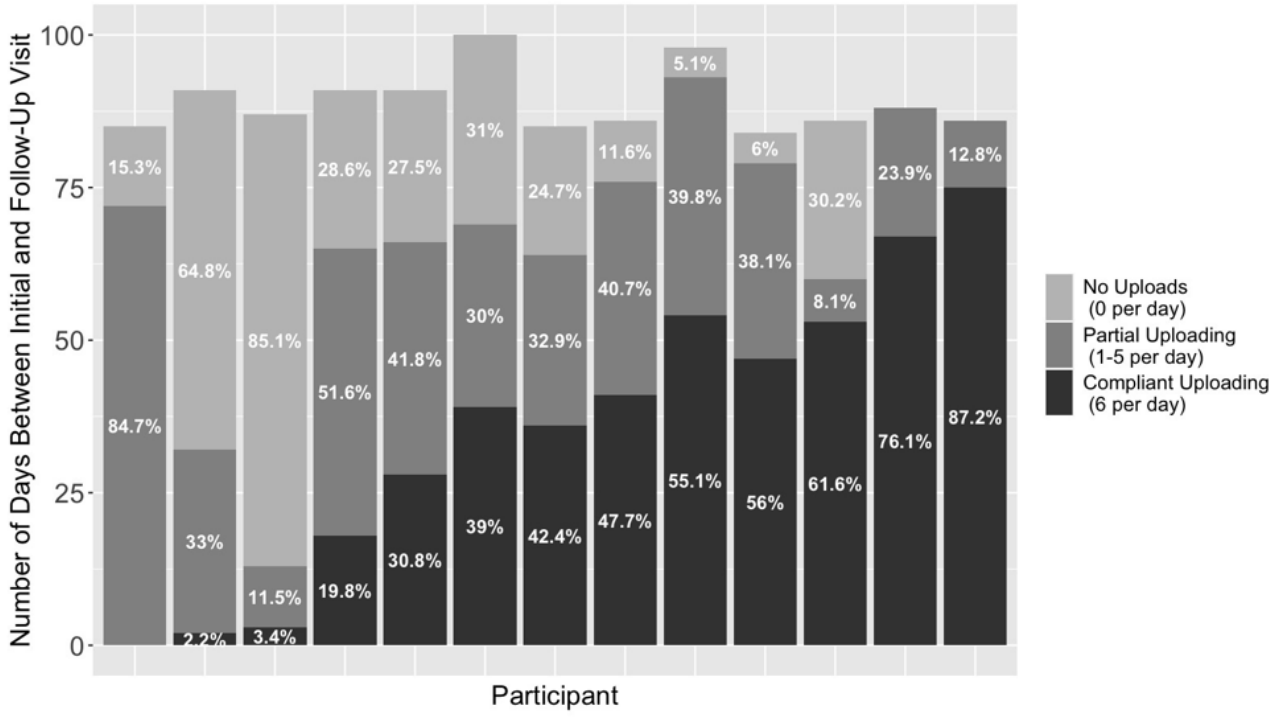
Feasibility

Across multiple sources of data, we found the use of BPTrack to be feasible within a primary care clinic.

Participant Utilization of BPTrack

The BPTrack protocol asked participants to measure their blood pressure two times per day, with three measurements at each sitting, for a total of 6 daily measurements. All participants were given at least 12 weeks (84 days) between their initial and follow-up visits to upload data, and most uploaded well into the 12th week of the study period. Because not all follow-up visits occurred on day 84, participants had varying numbers of days in the study and were encouraged to continue to measure their blood pressure until the end-of-study visit. Use of BPTrack varied widely across the 13 participants who completed the study; the majority of people (n=11) uploaded at least one blood pressure measurement on 60 or more days, ranging between 13 and 93 days across participants (Figure 1). The average number of days uploading data was 66.4 (SD 22.2).

Figure 1. Patient adherence to blood pressure monitoring.



The number of measurements uploaded per day by any given person ranged between 1 and 12, with the average being 4.6 (SD 1.5). The most common number of measurements were 3 (27.7% of all days) and 6 (40% of all days). In terms of adherence to the monitoring protocol, we examined total compliance (6 measurements per day) as well as partial compliance (1-5 measurements per day) for all 13 completers. The number of days of total compliance for each participant ranged from 0 to 75, and the number of days where participants logged at least one reading ranged from 13 to 93 days. On average, participants uploaded at least one measurement on 75% (SD 25%) of study days. Also, on average, participants were fully compliant 40.1% (SD 28%) of the days between baseline and follow-up.

Participant Perceptions of Feasibility

Findings from patient surveys consistently showed that the use of BPTrack was feasible. The majority of participants agreed or strongly agreed that the program was easy to use (92.3%; 12/13), that learning to use BPTrack was easy (92.3%; 12/13), and that they tried to use BPTrack every day (92.3%; 12/13). Moreover, most patients did not find it challenging to incorporate the information from the application into their blood pressure management (92.3%; 12/13), and most patients did not find the integration of the blood pressure monitor and their cell phone to be confusing (84.6%; 11/13). Only 1 participant (7.7%; 1/13) indicated that using BPTrack took up too much of their time.

Findings from our semistructured patient interviews confirmed that patients viewed BPTrack overall as feasible for use within a clinical relationship. Generally, patients were able to utilize the BPTrack functions as designed, reporting, “I had no problems with the program” and “I just liked that it kept me informed. I could just open it up and look at it.”

Although patients saw BPTrack as a whole as feasible for use within primary care, two themes arose related to improving app function and adherence challenges. Patients recommended changes to the blood pressure cuffs, sync, and message functions to improve ease of use. While many patients reported that the blood pressure cuffs “worked well” and were “very sturdy,” one patient requested a larger cuff size and stated that the cuff tended to overinflate. Additional concerns expressed by multiple patients included trouble syncing or pairing their blood pressure cuff:

> *The… phone would unsync from the blood pressure. (Um-hm) And it would take several tries. Even though it said it paired, it did not pair. […] So I tried, you know, 5, 6, 7 times.*

A few patients requested improvements to messaging. One suggestion was to allow custom messages:

> *Um, I had my own way of managing all that, you know. ‘Cause I have other things I like to remind myself of, and… to have yet another… set a notifications. And I turned it off. I, uh… I know you can’t do it in the app, but I can have, on Android I, you can, you know, force the notifications to stop. (Right) And I did. (Okay) So I, I turned ‘em all off.*

Another patient pointed out that direct communication from the application to the medical record would be beneficial:

> *I’m not a reporting type person. So… it would be better for my health care, if…. It all just went straight to the medical record, and my doctor could just pull up my blood pressure reading.*

Finally, several participants noted they had trouble adhering to the blood pressure monitoring protocol, which is not surprising given the BPTrack utilization statistics reported above. One concern was simply remembering to take blood pressure:

> *… sometimes it was… it was easy to, to forget to do it in the morning.*
>
> *I did have to keep the monitor, like on a table, so that I would remember to do it, (Right) on a regular basis. I didn't, if I didn't see it, I wouldn't think about it, so.*

Patients also expressed concerns about keeping the routine for the long term:

> *I don't know that I could take my blood pressure 3 times… every morning, and 3 times every night, for the rest of my life…"*
>
> *It was kinda hard for me to sit down and take my blood pressure at 2 different times. But, so I did it all, l generally at the same time most days.*

Stakeholder Perceptions of Feasibility

Despite our best efforts to recruit multiple stakeholders for semistructured interviews after the BPTrack pilot was complete, we only enrolled 1 pharmacist and 1 primary care physician. The stakeholders' recognized the potential advantages of the BPTrack program. In particular, the physician indicated that "we have so many resources" and that BPTrack could "be part of the clinic workflow" with some minor adjustments. The physician even noted that something like BPTrack could "open up time for other stuff. I mean, we spend a lot of time… follow-ups, nurse visits for [blood pressure checks]." However, despite the feasibility and benefit of integrating BPTrack into routine practice, the stakeholders also indicated a few notable reservations, including the accuracy of readings, patient adherence, and a desire for data to transfer directly from BPTrack into the electronic health record (EHR).

In many ways, the physician's and pharmacist's perceptions of feasibility were similar. The physician explained, "I think it's good to have, but I think we're gonna see that it's not… for everybody." The physician indicated that better clinical decisions could be made when an accurate home blood pressure logs were available, either through a program such as BPTrack, or otherwise. Similarly, the pharmacist pointed out the utility of collecting the patient log in a mobile device: "It's a very portable record because they almost always have their phone […] and they could share it with whoever they're seeing." However, they both raised concerns about accuracy. The physician pointed out the importance of using the cuff correctly: "Make sure the blood pressure cuff is… they're doing it right at home." Moreover, the pharmacist also noted concern with home blood pressure monitors in general: "I wish that they were more accurate; more consistent."

Ensuring ease of use and minimal burden are critical to improving adherence. Regarding the application specifically, the physician explained:

> *It has to be very simplistic, especially for our older patients. So, I worry that if they have to type in their readings, that they're gonna put… you know, wrong numbers.*

The physician and pharmacist echoed patients' concerns about the long-term sustainability of frequent monitoring. In terms of the desired frequency of home blood pressure testing, the physician and pharmacist suggested monitoring three times per week. The physician explained the potential burden of maintaining the twice-daily regimen long-term and suggested that "Ideally, 3 times a week… All at the same time [of day]," was more maintainable and useful.

> *"I think… like anything else it… when it starts, the patient might be very… motivated […] maybe at the beginning it's really exciting, but after time people just… not really interested in using it.*

Similarly, the pharmacist said: "I think once a day is awesome… if that's asking too much, I think 3 days a week is usually sufficient."

Finally, a concern was noted that the data from the application did not go directly into the EHR. The provider had to manually enter blood pressures and calculate averages, which was less than ideal. The pharmacist recommended:

> *I would love to be able to do that functionality, you know, just import that and have it… average them out for me.*

The pharmacist went on to add:

> *The parts that we still need to work on are… getting that data into [the EHR]… Relying on a provider to log into a secondary system… long term is going to be very difficult.*

Lowering the provider-side burden of managing the data will be necessary, mainly through integration with the EHR.

Acceptability

Participant Perceptions of Acceptability

Patients self-reported a high degree of acceptability towards the BPTrack program with the majority indicating that they are satisfied with the program (92.3%; 12/13), the program was easy to learn (92.3%; 12/13) and use (92.3%; 12/13), and would like to continue using the program (84.6%; 11/13). Fully 100% (13/13) of participants who completed the follow-up assessment indicated that they liked being able to keep track of their blood pressures visually, they liked knowing someone was watching over their blood pressure in between clinic visits and would recommend BPTrack to others. Lending support for our pharmacist-led approach, only 23.1% (3/13) of participants agreed or strongly agreed that they would prefer that their doctor oversee their blood pressure. Only 7.7% (1/13) agreed that they had concerns about the privacy of their data.

Findings from our semistructured patient interviews confirmed that patients viewed BPTrack as acceptable for use within a clinical relationship. Thematic analysis revealed that participants were pleased with the program:

> *It was good. I enjoyed bein' able to see… day to day what my blood pressure was and… you know, seein' how I was improving, as I was takin' the medication and stuff.*

Another patient appreciated the convenience of checking blood pressure outside of a clinic:

> *Uh, the thing that I liked the most about the program was that it was in an environment that wasn't at a doctor's office… I could,… check my blood pressure in a usual state, either work, or home, or whatever. And I think that it gave a more accurate representation of what my blood pressure was like than to do it at a doctor's office where I'm always… uh, stressed out.*

Patients indicated they would continue to use the application if offered: "Uh, just that I, I would, if it was offered, you know, full time, I would, I would sign up for it." Participants reported positive feedback regarding timeline and personalized blood pressure feedback:

> *It was… interesting. And… very informative, I thought, because I'm new to high blood pressure. And… um… it was neat to see the trends.*

> *So without this app, I wouldn't know… what my blood pressure was yesterday or the day before that…*

A patient participant appreciated the timeliness of the information: "Before I… usually didn't see my blood pressure till uh… the next, uh, visit, uh, to my doctor." Participants also commonly reported the reassurance of having a health care professional monitor their incoming data:

> *And so I really… liked that part of it, where… it kept it recorded for me. And it was nice to know that it was… being read by a professional. You know, somebody that could help me if it did go out a whack.*

Because of the increased monitoring by a health care professional, several participants reported that they had a more positive view of their medication regimen:

> *I just feel like I'm … uh, having, taking my medication based more on… on real blood pressure readings than a one or two-time visit to the doctor.*

Despite positive views on the BPTrack program, participants noted a few areas for improvement. In particular, the daily medication reminders, as implemented, were not well-liked by all. Some complaints concerned the frequency of reminders:

> *The reminders were really nice at the beginning. But then… got annoying because, like I already know to take my blood pressure medicine at the same time every day. So the repetitiveness of it just got annoying.*

Others suggested tailoring reminders and information beyond medication reminders:

> *I don't know the messages were kinda you know, generic.*

> *There's a lot of little things that might get somebody to adjust their… their medication. But they only, um they only seem to address forgetting.*

Stakeholder Perceptions of Acceptability

Both the physician and pharmacist indicated their acceptance of incorporating the BPTrack application into practice. The physician appreciated access to blood pressure readings:

> *(It) would be a benefit because… we could do a lot more pharmacotherapy management like that [and] if we can have that information available to us, then it should help us take better control of our patients.*

In addition, the physician elaborated that extracting home blood pressure readings verbally from patients during office visits was time-consuming and often not representative of actual readings.

The pharmacist also appreciated access to readings, commenting that when logs are available, it was appealing to provide the patient with their data: "it's a very cool, um… system to record blood pressures, so that patients have a record." However, the pharmacist expressed reservations about timely reporting of elevated readings:

> *I wasn't checking the app every single day […] I didn't get enough, I guess, warning or that there was a problem.*

Preliminary Effectiveness

Effect of BPTrack on Blood Pressure and Medication Adherence

At 12 weeks follow-up, patients exhibited significant reductions in both systolic blood pressure (baseline mean 137.3 mm Hg, SD 11.1 mm Hg; follow-up mean 131.0 mm Hg, SD 9.9 mm Hg; P=.02) and diastolic blood pressure (baseline mean 89.4 mm Hg, SD 7.7 mm Hg; follow-up mean 82.5 mm Hg, SD 8.2 mm Hg; P<.001). Regarding medication adherence, at 12-week follow-up, the effect of BPTrack as measured by the Adherence to Refills and Medications scale was negligible and not significant (baseline 23.7 points, follow-up 23.1 points, P=.45). During this trial, 3 of our 13 participants had hypertension medications either removed from their treatment plan or had dosages lowered. Changes in systolic and diastolic blood pressure were not significantly associated with app utilization measures. There was a significant correlation between the number of hypertension-related encounters (r=0.77, P=.002) and the number of encounters with the clinical pharmacist (r=0.65, P=.02) and change in systolic blood pressure. In both instances, individuals with more encounters saw a greater reduction in their systolic blood pressure. There was no correlation with the number of encounters that were not hypertension-related (r=0.47, P=.11), nor was there an association with change in diastolic blood pressure and any of the health care utilization measures.

Participant Perceptions of Effectiveness

The majority of patients reported that using BPTrack was a benefit to their overall health (92.3%; 12/13), that BPTrack helped them to get their blood pressure under control (69.2%; 9/13), and helped them remember to take their medications (61.5%; 8/13).

Findings from our semistructured patient interviews confirmed that patients perceived BPTrack as useful in a variety of ways. Specifically, patients indicated that BPTrack helped raise their awareness of their hypertension and helped them to make behavior changes, such as eating healthfully, managing medications, and reducing stress.

Patients discussed how awareness helped them to consider what they were eating and make adjustments:

> *When I saw the different times of days and, and how that, how it varied across the day, um, you know, which got me thinkin' about what I, maybe what I was eating throughout the day and then the types of food.*

A specific example was sodium intake:

> *It created an awareness within me to really pay attention to my diet, my salt intake, the types of food and, and become more involved with my overall health.*

Despite some of the suggestions to go beyond medication reminders, some patients reported the program was helpful for medication management:

> *Helped me remember to take my medication for one thing.*
>
> *Um, just heavy monitored. [The pharmacist] was able to make some adjustments, and we were able to delete some medication and increase some medication and get everything to a nice manageable level.*

Other patients used the data to recognize the causes of stress by taking readings more frequently than office visits:

> *It's very telling, I thought. I was able to… uh… distinguish a difference in my blood pressure, uh, based on my stress at work. (Um-hm) You know, when, when my work was more stressful, my blood pressure was definitely hanging out higher area, than it was, uh, when my work was less stressful… I never would have found out if I was taking my blood pressure manually or waited until I came to a doctor to have it taken.*

Another patient explained being able to act more quickly to reduce stress:

> *It gave me the information [Systolic Blood Pressure] was 140, so I need to relax. It will be better after I get some rest. It kept me alerted as to what I should be doing. As opposed to not knowing at all what the blood pressure was. 'Cause I felt the same the whole time through. I could never tell if my blood pressure was up or not.*

Stakeholder Perceptions of Effectiveness

The physician and pharmacist both thought that BPTrack would have definite benefits when integrated into routine practice; however, both noted that solutions such as BPTrack cannot stand alone and need to be used in conjunction with standard care. While the physician felt it could be helpful, a concern was over-reliance on BPTrack to manage hypertensive patients:

> *Using this BP method, I worry that they won't be seen for long periods of time. Or that we'll… start throwing too many meds without having an adequate follow-up.*

The physician also explained that regular visits are necessary and help the patient to feel more open to talking:

> *I still think patients should come in and be seen and, even if everything's fine […] just a regular physical, making sure they're taking their meds appropriately, that they're not having any side effects, sometimes these things you won't, they won't tell you over the phone." […] They'll be more… likely to talk about their side effects or issues that they're struggling with in person.*

The pharmacist focused on how the application lent itself to improved patient education:

> *I liked that it graphs the numbers out for them, that it color-codes them. I think it helps patients learn, um, what's good and what's not good in terms of blood pressure control.*

Healthcare Utilization

Throughout the 12-week intervention, the 15 participants had 122 points of contact with their primary care clinic or the emergency department (mean 8 points of contact; range 4-20 points of contact). Of these points of contact, one involved a patient who presented to the emergency room experiencing symptoms associated with high blood pressure due to medication nonadherence and was treated accordingly. Most of the points of contact (78.7%; 96/122) included a hypertension-related focus (provider-initiated follow-ups to monitor BP, medications, symptoms, side effects, etc). These hypertension-related points of contact were conducted predominantly via phone (64.6%; 62/96), with in-person (33.3%; 32/96) and email (2.1%; 2/96) contacts occurring less frequently. Many of these points of contact are attributed to the increased blood pressure management oversight by our study pharmacist, as 76.0% (73/96) were contacts with our study pharmacist. Moreover, during the 12-week intervention, a total of 20 hypertension medication adjustments were made across our 15 participants (range 0-5 medication changes per participant), 70% (14/20) of which were made by the study pharmacist. Reasons for medication changes included adding or removing medications and dose adjustments.

The number of contacts with the study pharmacist (r=0.65, P=.02), number of hypertension-related encounters (r=0.77, P=.002), and number of encounters resulting in hypertension medication changes (r=0.68, P=.01) were all positively correlated with a change in systolic blood pressure. Given the high level of collinearity between clinical pharmacists and hypertension-related encounters and med changes as well as the small sample size adjusted analyses were not feasible.

Discussion

Our pharmacist-led, mHealth supported approach to hypertension management shows great promise for helping to reduce blood pressures among uncontrolled hypertensive patients within primary care, as we have found this approach to be feasible, acceptable, and effective among patient and stakeholder participants.

Feasibility

Like other recent digital health interventions for hypertension that use home blood pressure monitors, patients found BPTrack relatively easy to learn and use [19]. Complaints noted by patient participants largely related to issues that could be easily addressed through refining the app itself, as well as usage protocol. These issues include refinements to the BPTrack interface, documentation and protocols for the app and blood pressure cuff syncing, and the in-app messaging. Despite the burden placed on patient participants for self-monitoring, adherence to a minimum of daily self-monitoring was surprisingly high in this pilot, with participants self-monitoring their blood pressure at least once a day on 75% of the days in the study; however, criticism from patient participant concerning the self-monitoring protocol, which required three separate measurements twice daily, were common. Participants found the self-monitoring protocol to be laborious and time-consuming. Our decision to include such a rigid monitoring protocol was made because these were the monitoring guidelines in place at the University of Michigan at the time the study took place and followed guidance commonly recommended to patients by other institutions [20,21]. Interviews with health care providers revealed that despite those monitoring guidelines, they did not see the need to have that much data available, and the protocol would be likely laborious to patients. In reality, both providers agreed independently that self-monitoring blood pressure three times per week was sufficient. Furthermore, both providers noted concerns about the accuracy of home blood pressure cuffs, and regular calibration or comparison against clinic blood pressure cuffs was important to ensure the data was useful, with the pharmacist also noting that ensuring proper fit was essential. These are issues that have been established in the literature with clinicians expressing concern about home blood pressure monitoring and the use of inaccurate devices, adherence to protocols, or patient ability to interpret data [14]. As the use of remote blood pressure monitoring becomes more common within a clinical context, health systems should look to professional organizations such as the American Medical Association, for recommendations on how to engage patients in a way that maximizes data quality and patient care [22].

Providers were also united in their belief that tools such as BPTrack must integrate with other health information technology (IT) systems, namely the EHR, to be truly useful. As noted in several other studies, when systems contain patient data but do not integrate with the EHR, their usability is severely limited due to incompatibility with workflow [23-25]. In an ideal world, future research would seek to evaluate the efficacy of BPTrack or a BPTrack-like intervention, that is seamlessly integrated with the EHR to promote continuity of care and to leverage the collected data within the full care team.

As noted, the two participating providers suggested several concerns about the long-term feasibility of incorporating BPTrack or similar interventions into routine clinical practice. Our current clinical environment, and existing clinical workflows, are not designed to utilize digital interventions such as these. Most primary care practices in the United States do not have embedded clinical pharmacists, let alone the technological infrastructure needed to support this type of intervention.

In addition to reimagining workflow to support the expansion of digital health, well-designed protocols to address the inherent ethical concerns need to be established by health systems. For example, the pharmacist in the present pilot study expressed concerns about response times to concerning blood pressure values and a lack of a consistent alert system embedded in the EHR. Although the data was available to view, the pharmacist was unable to allocate time to review blood pressure values consistently. Integration into the workflow will become increasingly essential as mHealth interventions are scaled up. Also of concern is identifying the correct health care professional to review and respond to incoming patient-generated data. In the present study, we utilized a clinical pharmacist, but perhaps a medical assistant or personnel hired and trained to interact with this type of data specifically would be better suited [26].

Acceptability

Qualitative and quantitative data reveal that patients found BPTrack to be acceptable for use. As noted, participants had high degrees of satisfaction with the program and would like to continue using the program. In particular, participants report that they found it valuable to see their blood pressure data, both daily as it was measured, and as a longitudinal trend. These observations are consistent with constructs such as self-regulation theory (on which this intervention is built) [17], which suggests that individuals engage in a dynamic feedback loop where they synthesize information about past behavior and integrate that information into goals and motivation to change future behaviors. Self-monitoring and self-reflection are key components of this dynamic feedback loop and are directly supported by the BPTrack intervention [27].

In addition to general thoughts on the acceptability of BPTrack, participants expressed their sense of satisfaction and safety in knowing their blood pressure measurements were being monitored by a health care professional who could make changes to their medication regimens. For some, this was noted as increasing satisfaction with the actual treatment plan, which may have downstream consequences for medication adherence [28]. For reasons of logistics, cost, reimbursement, and liability, many mHealth apps circumvent the health care system by focusing solely on consumer-facing apps; however, this pilot study demonstrates that individual patients have a real appetite to engage with their health care team through tools such as BPTrack rather than basic patient portals or EHRs. Likewise, on the whole, the two providers we spoke with agreed that there were benefits to engaging with patients through BPTrack for issues such as guiding treatment, although they were quick to point out concerns with an overreliance on the tool. Concerns were largely focused on implementation concerns (how to fit BPTrack into existing clinical workflow and health IT systems) as well as ethical considerations (issues related to liability and the responsibility to patients and their data). Regardless, expanded use of digital tools is likely to feature prominently in the future of health care, and implementation factors (including

developing proper protocols to address ethical concerns) have been noted as the biggest hurdle facing expansion [26].

Effectiveness

After 12 weeks, patients exhibited significant reductions in both systolic and diastolic blood pressure, both of which are considered clinically meaningful. Our finding of efficacy in this pilot study is supported by literature demonstrating the benefit of pharmacist-led [6,29-32] and self-monitoring interventions [32,33] for hypertension management. Not only were blood pressure outcomes improved at the end of 12 weeks, but patients' perceptions of the effectiveness of BPTrack were noted in qualitative interviews.

Despite improvements in blood pressure and the presence of medication reminders within the BPTrack app, it is particularly interesting to note that our patients exhibited no significant improvements in medication adherence, which contributes to the already mixed literature on the effectiveness of mobile medication reminders [34-37]. Our adherence measures may not be sensitive enough to identify changes, which is a potential limitation of this study. It is also possible that medication adherence did not change, as we noted, but that participants engaged in other behavior changes that were not assessed as a result of their participation in BPTrack (increased physical activity or improvements in diet). Without knowing the mechanism through which participants managed to lower their blood pressure, it is difficult to identify how BPTrack led to improved blood pressure outcomes at 12 weeks, which is typical of black box interventions such as this.

Limitations

This study was not without limitations. Given that the BPTrack intervention was available only to participants with smartphones, there are inherent digital-divide concerns. It should be noted though that smartphone adoption is exceedingly high (about 81% of American adults) and increasing; however, smartphone adoption is still lagging among individuals 65+ years of age (53%), with less than a high school education (66%), with annual household incomes <$30,000 (71%), and who live in rural areas (71%) [16]. Moreover, in the present study, our participants were quite heterogeneous in terms of socioeconomic status. Future work should seek to identify how this type of intervention fares with more disparate groups. Members of the study team are conducting a large randomized controlled trial of a home-based blood pressure self-monitoring intervention among uncontrolled hypertensive African Americans recruited from urban emergency departments and community settings (NCT02955537). Although this is not a pharmacist-led intervention, it does target a population that typically suffers from great health disparities, and who are more apt to be negatively impacted by the digital divide. This study was also limited by the research design, which was a simple one-group, pre-post design with no control group, as well as a short-term follow-up (12 weeks).

Moreover, the small sample size of both patients and stakeholders limits our findings. Future work should seek to look at longer-term implications, with larger sample sizes of patients and stakeholders, of a digital health intervention embedded within clinical care to determine whether the positive effects might be sustained over time. Finally, as with all packaged interventions, it is not clear which intervention components may have individually, or as a set, contributed to the positive reductions in patient participant blood pressure. Given that pharmacist-led interventions are efficacious [38-40], we do not know what the value add of the BPTrack app was. As we all desire to keep health care costs in check, future work should seek to understand whether the technology component added benefit above and beyond pharmacist-led interventions for managing hypertension that did not include technology, as well as to conduct cost-effectiveness studies of interventions like BPTrack, as well as pharmacist-led interventions without technology.

Conclusions

Our results demonstrate that a pharmacist-led mHealth intervention that promotes home blood pressure monitoring and clinical pharmacist management of hypertension can be effective at reducing systolic and diastolic blood pressure in primary care patients with uncontrolled hypertension. Our data also support the idea that these types of interventions are feasible and acceptable to patients and providers and are effective at improving health outcomes. There is a need for more robust trials of digital health interventions integrated into routine clinical care to more robustly determine potential effectiveness, as well as guided investigations to more fully understand how to implement these types of interventions into clinical practice thoughtfully.

Acknowledgments

This study was funded by the Michigan Institute for Clinical & Health Research (UL1TR000433). The authors wish to thank: Eric Caillibot, André Piccolomini, Christopher Tracton, Sébastien Filion, Charles Coutu-Nadeau, Michel Nadeau, and the rest of the Tactio Health Group for technical assistance and support, as well as Lauren Nichols, Mary Rapai, Judy Connelly, Lilly Pritula, Rania Ajilat, and Maggie Riley from the University of Michigan Department of Family Medicine for administrative support.

Conflicts of Interest

None declared.

References

1. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, et al. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. Circulation 2014 Jan 21;129(3):e28-e292 [FREE Full text] [doi: 10.1161/01.cir.0000441139.02102.80] [Medline: 24352519]
2. Kochanek K, Murphy S, Xu J, Arias E. Deaths: Final data for 2017. 2019. URL: https://stacks.cdc.gov/view/cdc/79486 [accessed 2020-04-20]
3. Muntner P, Carey RM, Gidding S, Jones DW, Taler SJ, Wright JT, et al. Potential US Population Impact of the 2017 ACC/AHA High Blood Pressure Guideline. Circulation 2018 Jan 09;137(2):109-118 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.117.032582] [Medline: 29133599]
4. Choe HM, Farris KB, Stevenson JG, Townsend K, Diez HL, Remington TL, et al. Patient-centered medical home: developing, expanding, and sustaining a role for pharmacists. Am J Health Syst Pharm 2012 Jun 15;69(12):1063-1071. [doi: 10.2146/ajhp110470] [Medline: 22644984]
5. Nigro SC, Garwood CL, Berlie H, Irons B, Longyhore D, McFarland MS, et al. Clinical pharmacists as key members of the patient-centered medical home: an opinion statement of the Ambulatory Care Practice and Research Network of the American College of Clinical Pharmacy. Pharmacotherapy 2014 Jan;34(1):96-108. [doi: 10.1002/phar.1357] [Medline: 24122857]
6. Margolis KL, Asche SE, Bergdall AR, Dehmer SP, Groen SE, Kadrmas HM, et al. Effect of home blood pressure telemonitoring and pharmacist management on blood pressure control: a cluster randomized clinical trial. JAMA 2013 Jul 3;310(1):46-56 [FREE Full text] [doi: 10.1001/jama.2013.6549] [Medline: 23821088]
7. Santschi V, Chiolero A, Colosimo AL, Platt RW, Taffé P, Burnier M, et al. Improving blood pressure control through pharmacist interventions: a meta-analysis of randomized controlled trials. J Am Heart Assoc 2014 Apr 10;3(2):e000718 [FREE Full text] [doi: 10.1161/JAHA.113.000718] [Medline: 24721801]
8. Agarwal R, Bills JE, Hecht TJW, Light RP. Role of home blood pressure monitoring in overcoming therapeutic inertia and improving hypertension control: a systematic review and meta-analysis. Hypertension 2011 Jan;57(1):29-38 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.110.160911] [Medline: 21115879]
9. Glynn LG, Murphy AW, Smith SM, Schroeder K, Fahey T. Interventions used to improve control of blood pressure in patients with hypertension. Cochrane Database Syst Rev 2010 Mar 17(3):CD005182. [doi: 10.1002/14651858.CD005182.pub4] [Medline: 20238338]
10. Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007;14(3):269-277 [FREE Full text] [doi: 10.1197/jamia.M2270] [Medline: 17329725]
11. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. J Med Internet Res 2010;12(2):e21 [FREE Full text] [doi: 10.2196/jmir.1357] [Medline: 20554500]
12. Patil SJ, Wareg NK, Hodges KL, Smith JB, Kaiser MS, LeFevre ML. Home Blood Pressure Monitoring in Cases of Clinical Uncertainty to Differentiate Appropriate Inaction From Therapeutic Inertia. Ann Fam Med 2020 Jan;18(1):50-58. [doi: 10.1370/afm.2491] [Medline: 31937533]
13. Burke LE, Sereika SM, Music E, Warziski M, Styn MA, Stone A. Using instrumented paper diaries to document self-monitoring patterns in weight loss. Contemp Clin Trials 2008 Mar;29(2):182-193 [FREE Full text] [doi: 10.1016/j.cct.2007.07.004] [Medline: 17702667]
14. Liyanage-Don N, Fung D, Phillips E, Kronish IM. Implementing Home Blood Pressure Monitoring into Clinical Practice. Curr Hypertens Rep 2019 Feb 12;21(2):14. [doi: 10.1007/s11906-019-0916-0] [Medline: 30747350]
15. Parati G, Omboni S. Role of home blood pressure telemonitoring in hypertension management: an update. Blood Press Monit 2010 Dec;15(6):285-295. [doi: 10.1097/MBP.0b013e328340c5e4] [Medline: 21084882]
16. Pew Research Center. Mobile Fact Sheet. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-04-20]
17. Buis LR, Roberson DN, Kadri R, Rockey NG, Plegue MA, Choe HM, et al. Utilizing Consumer Health Informatics to Support Management of Hypertension by Clinical Pharmacists in Primary Care: Study Protocol. JMIR Res Protoc 2017 Oct 10;6(10):e193 [FREE Full text] [doi: 10.2196/resprot.8059] [Medline: 29017994]
18. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
19. Ciemins EL, Arora A, Coombs NC, Holloway B, Mullette EJ, Garland R, et al. Improving Blood Pressure Control Using Smart Technology. Telemed J E Health 2018 Dec;24(3):222-228. [doi: 10.1089/tmj.2017.0028] [Medline: 28930497]
20. American Heart Association. Monitoring Your Blood Pressure at Home. Heart.org. 2017. URL: https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings/monitoring-your-blood-pressure-at-home [accessed 2020-04-20]
21. Mayo Clinic. Get the most out of home blood pressure monitoring. Mayo Clinic. 2019. URL: https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/in-depth/high-blood-pressure/art-20047889 [accessed 2020-04-20]
22. Murakami L, Rakotz M. Self-measured Blood Pressure Monitoring Program: Engaging Patients in Self-measurement. American Medical Association and the Johns Hopkins University School of Medicine. 2015. URL: https://www.ama-assn.org/

sites/ama-assn.org/files/corp/media-browser/public/about-ama/iho-bp-engaging-patients-in-self-measurment_0.pdf [accessed 2020-04-20]
23. Cox CE, Jones DM, Reagan W, Key MD, Chow V, McFarlin J, et al. Palliative Care Planner: A Pilot Study to Evaluate Acceptability and Usability of an Electronic Health Records System-integrated, Needs-targeted App Platform. Ann Am Thorac Soc 2018 Jan;15(1):59-68 [FREE Full text] [doi: 10.1513/AnnalsATS.201706-500OC] [Medline: 29121480]
24. Kumar RB, Goren ND, Stark DE, Wall DP, Longhurst CA. Automated integration of continuous glucose monitor data in the electronic health record using consumer technology. J Am Med Inform Assoc 2016 May;23(3):532-537 [FREE Full text] [doi: 10.1093/jamia/ocv206] [Medline: 27018263]
25. Peeples MM, Iyer AK, Cohen JL. Integration of a mobile-integrated therapy with electronic health records: lessons learned. J Diabetes Sci Technol 2013 May;7(3):602-611 [FREE Full text] [Medline: 23759392]
26. Buis L. Implementation: The Next Giant Hurdle to Clinical Transformation With Digital Health. J Med Internet Res 2019 Nov 20;21(11):e16259 [FREE Full text] [doi: 10.2196/16259] [Medline: 31746763]
27. Carver C. Self-regulation of actioneffect. In: Baumeister RF, Vohs KD. editors. Handbook of Self-regulation: Research, Theory, and Application. New York, NY: Guilford Press; 2004:1572309911-1572309948.
28. Barbosa CD, Balp M, Kulich K, Germain N, Rofail D. A literature review to explore the link between treatment satisfaction and adherence, compliance, and persistence. Patient Prefer Adherence 2012;6:39-48 [FREE Full text] [doi: 10.2147/PPA.S24752] [Medline: 22272068]
29. Bajorek B, Lemay KS, Magin P, Roberts C, Krass I, Armour CL. Implementation and evaluation of a pharmacist-led hypertension management service in primary care: outcomes and methodological challenges. Pharm Pract (Granada) 2016;14(2):723 [FREE Full text] [doi: 10.18549/PharmPract.2016.02.723] [Medline: 27382427]
30. Cheema E, Sutcliffe P, Singer DRJ. The impact of interventions by pharmacists in community pharmacies on control of hypertension: a systematic review and meta-analysis of randomized controlled trials. Br J Clin Pharmacol 2014 Dec;78(6):1238-1247 [FREE Full text] [doi: 10.1111/bcp.12452] [Medline: 24966032]
31. Clay PG. Science snippets. Pharmacist-led blood pressure intervention demonstrates positive results. J Am Pharm Assoc (2003) 2013;53(5):560. [doi: 10.1331/JAPhA.2013.13534] [Medline: 24030135]
32. Magid DJ, Olson KL, Billups SJ, Wagner NM, Lyons EE, Kroner BA. A pharmacist-led, American Heart Association Heart360 Web-enabled home blood pressure monitoring program. Circ Cardiovasc Qual Outcomes 2013 Mar 1;6(2):157-163 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.112.968172] [Medline: 23463811]
33. Kaambwa B, Bryan S, Jowett S, Mant J, Bray EP, Hobbs FDR, et al. Telemonitoring and self-management in the control of hypertension (TASMINH2): a cost-effectiveness analysis. Eur J Prev Cardiol 2014 Dec;21(12):1517-1530. [doi: 10.1177/2047487313501886] [Medline: 23990660]
34. Buis L, Hirzel L, Dawood RM, Dawood KL, Nichols LP, Artinian NT, et al. Text Messaging to Improve Hypertension Medication Adherence in African Americans From Primary Care and Emergency Department Settings: Results From Two Randomized Feasibility Studies. JMIR Mhealth Uhealth 2017 Feb 01;5(2):e9 [FREE Full text] [doi: 10.2196/mhealth.6630] [Medline: 28148474]
35. Conway CM, Kelechi TJ. Digital Health for Medication Adherence in Adult Diabetes or Hypertension: An Integrative Review. JMIR Diabetes 2017 Aug 16;2(2):e20 [FREE Full text] [doi: 10.2196/diabetes.8030] [Medline: 30291093]
36. Gandapur Y, Kianoush S, Kelli HM, Misra S, Urrea B, Blaha MJ, et al. The role of mHealth for improving medication adherence in patients with cardiovascular disease: a systematic review. Eur Heart J Qual Care Clin Outcomes 2016 Oct 01;2(4):237-244 [FREE Full text] [doi: 10.1093/ehjqcco/qcw018] [Medline: 29474713]
37. Kim JY, Wineinger NE, Steinhubl SR. The Influence of Wireless Self-Monitoring Program on the Relationship Between Patient Activation and Health Behaviors, Medication Adherence, and Blood Pressure Levels in Hypertensive Patients: A Substudy of a Randomized Controlled Trial. J Med Internet Res 2016 Jun 22;18(6):e116 [FREE Full text] [doi: 10.2196/jmir.5429] [Medline: 27334418]
38. Morgado M, Rolo S, Castelo-Branco M. Pharmacist intervention program to enhance hypertension control: a randomised controlled trial. Int J Clin Pharm 2011 Feb;33(1):132-140 [FREE Full text] [doi: 10.1007/s11096-010-9474-x] [Medline: 21365405]
39. Morgado MP, Morgado SR, Mendes LC, Pereira LJ, Castelo-Branco M. Pharmacist interventions to enhance blood pressure control and adherence to antihypertensive therapy: Review and meta-analysis. Am J Health Syst Pharm 2011 Feb 01;68(3):241-253. [doi: 10.2146/ajhp090656] [Medline: 21258029]
40. Stewart K, George J, Mc Namara KP, Jackson SL, Peterson GM, Bereznicki LR, et al. A multifaceted pharmacist intervention to improve antihypertensive adherence: a cluster-randomized, controlled trial (HAPPy trial). J Clin Pharm Ther 2014 Oct;39(5):527-534. [doi: 10.1111/jcpt.12185] [Medline: 24943987]

Abbreviations

EHR: electronic health record
IT: information technology
mHealth: mobile health

Edited by G Eysenbach; submitted 08.05.20; peer-reviewed by S Omboni, L Schwiebert; comments to author 01.06.20; revised version received 08.06.20; accepted 13.06.20; published 11.08.20.

Please cite as:
Buis LR, Roberson DN, Kadri R, Rockey NG, Plegue MA, Danak SU, Guetterman TC, Johnson MG, Choe HM, Richardson CR Understanding the Feasibility, Acceptability, and Efficacy of a Clinical Pharmacist-led Mobile Approach (BPTrack) to Hypertension Management: Mixed Methods Pilot Study
J Med Internet Res 2020;22(8):e19882
URL: https://www.jmir.org/2020/8/e19882
doi:10.2196/19882
PMID:32780026

Original Paper

Recruiting Adolescents With Chronic Fatigue Syndrome/Myalgic Encephalomyelitis to Internet-Delivered Therapy: Internal Pilot Within a Randomized Controlled Trial

Emma Anderson[1], DHealthPsych; Roxanne Parslow[1], PhD; William Hollingworth[2], PhD; Nicola Mills[2], PhD; Lucy Beasant[1], PhD; Daisy Gaunt[2,3], MSc; Chris Metcalfe[2,3], PhD; David Kessler[2], MD; John Macleod[2], PhD; Susan Pywell[4]; Kieren Pitts[5], PhD; Simon Price[6], PhD; Paul Stallard[7], PhD; Hans Knoop[8], PhD; Elise Van de Putte[9], MD, PhD; Sanne Nijhof[9], MD, PhD; Gijs Bleijenberg[10], PhD; Esther Crawley[1], PhD

[1]Centre for Academic Child Health, Bristol Medical School: Population Health Sciences, University of Bristol, Bristol, United Kingdom

[2]Bristol Medical School: Population Health Sciences, University of Bristol, Bristol, United Kingdom

[3]Bristol Trials Centre, University of Bristol, Bristol, United Kingdom

[4]Advanced Computing Research Centre, University of Bristol, Bristol, United Kingdom

[5]Research IT, University of Bristol, Bristol, United Kingdom

[6]Department of Computer Science, University of Bristol, Bristol, United Kingdom

[7]Department for Health, University of Bath, Bath, United Kingdom

[8]Department of Medical Psychology, Amsterdam University Medical Centres, Amsterdam Public Health Research Institute, University of Amsterdam, Amsterdam, Netherlands

[9]Department of Paediatrics, Wilhelmina Children's Hospital, University Medical Centre Utrecht, Utrecht, Netherlands

[10]Radboud University Medical Center, Nijmegen, Netherlands

Corresponding Author:
Emma Anderson, DHealthPsych
Centre for Academic Child Health
Bristol Medical School: Population Health Sciences
University of Bristol
1-5 Whiteladies Road
Bristol
United Kingdom
Phone: 44 117 4283102
Email: emma.anderson@bristol.ac.uk

Abstract

Background: Chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) in adolescents is common and disabling. Teenagers in the United Kingdom are more likely to recover if they access specialist care, but most do not have access to a local specialist CFS/ME service. Delivering treatment remotely via the internet could improve access to treatment.

Objective: This study aims to assess (1) the feasibility of recruitment and retention into a trial of internet-delivered specialist treatment for adolescents with CFS/ME and (2) the acceptability of trial processes and 2 web-based treatments (to inform continuation to full trial).

Methods: This study is an internal pilot for the initial 12 months of a full randomized controlled trial (RCT), with integrated qualitative methods (analysis of recruitment consultations and participant and clinician interviews). Recruitment and treatment were delivered remotely from a specialist pediatric CFS/ME treatment service within a hospital in South West United Kingdom. Adolescents (aged 11-17 years) from across the United Kingdom with a diagnosis of CFS/ME and no access to local specialist treatment were referred by their general practitioner to the treatment center. Eligibility assessment and recruitment were conducted via remote methods (telephone and on the web), and participants were randomized (via a computer-automated system) to 1 of 2 web-based treatments. The trial intervention was Fatigue in Teenagers on the InterNET in the National Health Service, a web-based modular CFS/ME-specific cognitive behavioral therapy program (designed to be used by young people and their parents or caregivers) supported by individualized clinical psychologist electronic consultations (regular, scheduled therapeutic message exchanges between participants and therapist within the platform). The comparator was Skype-delivered activity management

with a CFS/ME clinician (mainly a physiotherapist or occupational therapist). Both treatments were intended to last for up to 6 months. The primary outcomes were (1) the number of participants recruited (per out-of-area referrals received between November 1, 2016, to October 31, 2017) and the proportion providing 6-month outcome data (web-based self-report questionnaire assessing functioning) and (2) the qualitative outcomes indicating the acceptability of trial processes and treatments.

Results: A total of 89 out of 150 (59.3% of potentially eligible referrals) young people and their parents or caregivers were recruited, with 75 out of 89 (84.2%) providing 6-month outcome data. Overall, web-based treatment was acceptable; however, participants and clinicians described both the advantages and disadvantages of remote methods. No serious adverse events were reported.

Conclusions: Recruiting young people (and their parents or caregivers) into an RCT of web-based treatment via remote methods is feasible and acceptable. Delivering specialist treatment at home via the internet is feasible and acceptable, although some families prefer to travel across the United Kingdom for face-to-face treatment.

Trial Registration: ISRCTN 18020851; http://www.isrctn.com/ISRCTN18020851

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-018-2500-3

(J Med Internet Res 2020;22(8):e17768) doi:10.2196/17768

KEYWORDS

pediatrics; chronic fatigue syndrome; myalgic encephalomyelitis; cognitive behavioral therapy; eHealth; online systems; e-therapy; e-counseling; pilot projects; qualitative research

Introduction

Between 1% and 2.4% of children and teenagers are estimated to have chronic fatigue syndrome or myalgic encephalomyelitis (CFS/ME) [1,2]. Affected children and teenagers can have severe disability [3,4], and most of them have no access to specialist treatment in their locality. This situation forces children and teenagers to either remain untreated or travel long distances to access specialist care, which may exacerbate symptoms.

One solution is to provide specialist treatment remotely to enable all families in the United Kingdom to access treatment and remove the need for travel. A Dutch trial of a web-based CFS/ME-specific cognitive behavioral therapy (CBT) called Fatigue in Teenagers on the InterNET (FITNET) showed promising results [5]. Two-thirds of young people offered the intervention recovered at 6 months compared with just 8% in the control (usual care) arm.

However, the United Kingdom has a different system of health care provision, and further testing is required to investigate whether FITNET is effective and cost-effective in the National Health Service (NHS). In contrast to the Dutch study, the Fatigue in Teenagers on the InterNET in the National Health Service (FITNET-NHS) trial uses an active specialist treatment comparison group (activity management [AM] via Skype), rather than treatment as usual or waiting list control, as specifically recommended for this trial by the funders.

This pilot study aimed to assess the feasibility of the remote recruitment of adolescents (and their parents/caregivers) into a randomized controlled trial (RCT) of a UK-adapted version of the Dutch CBT program—FITNET-NHS—compared with a version of usual care—AM (delivered via Skype), and to assess the acceptability of the 2 web-based interventions.

Methods

Ethics

The pilot and full RCT protocol and all associated documents were reviewed and approved by the South West-Frenchay Research Ethics Committee (reference 16/SW/0268).

Design and Recruitment

This study presents the findings of the initial 12-month internal pilot phase of the FITNET-NHS RCT, which used entirely remote methods to recruit, randomize, and treat adolescents who were referred to a specialist pediatric CFS/ME treatment service within a hospital in South West United Kingdom between November 1, 2016, to October 31, 2017.

The eligibility criteria included the following: (1) young people aged between 11 and 17 years, with (2) a diagnosis of CFS/ME, and (3) no access to local specialist pediatric CFS/ME treatment (defined as more than 1 hour's journey to their closest specialist treatment center or >6 months' waiting list). Exclusions were as follows: (1) patients whose fatigue was due to another cause or was not disabling, (2) patients who would be unable to complete video calls or web-based modules (eg, due to developmental problems, lack of literacy, or lack of internet access), or (3) patients who were pregnant at the time of assessment.

The young person's parent/caregiver was asked to provide consent for the study. Additionally, participants aged between 11 and 15 years provided assent, whereas those aged between 16 and 17 years provided their consent to participate in the study.

The South West United Kingdom specialist pediatric CFS/ME treatment center has always accepted out-of-area (as well as local) referrals of children and young people from general practitioners (GPs) across the United Kingdom for both diagnosis and treatment of CFS/ME. This meant that many families had to travel long distances to the center for treatment.

The FITNET-NHS trial was launched with a message to GPs in the United Kingdom that out-of-area referrals could now receive treatment without travel.

A detailed description of the study methods is presented elsewhere [6]. In brief, all referrals from GPs were screened by administrative staff, and those potentially eligible for FITNET-NHS were contacted by research nurses. The research nurses conducted an initial brief telephone discussion with the families of potentially eligible adolescents and provided information about the trial (including emailing patient information leaflets). For interested families, a second call by a research nurse was arranged to conduct full eligibility assessment, recruitment discussion, and take consent (via a web-based form). After consenting, participants were randomly allocated (individually via an automated web randomization service, set up and managed by Bristol Randomised Trials Collaboration to maintain allocation concealment) on a 1:1 ratio (using minimization to facilitate balance by age and gender) to 1 of the 2 interventions. Due to the nature of the interventions, it was not practical to blind the participant, family, or clinical service to treatment allocation. Participants and parents/caregivers completed a web-based (emailed) self-report set of baseline questionnaires, and all were followed up with (emailed) web-based questionnaires at 3, 6, and 12 months after recruitment.. A subsample was invited to be interviewed (qualitative recruitment details are provided under *Integrated Qualitative Methods* below). The 6-month web-based self-report Short Form Health Survey Physical Function Subscale (SF-36-PFS) 10-item questionnaire [7] is the primary outcome for the full trial. The SF-36-PFS measures disability, an important outcome for children with CFS/ME [8], and is sufficiently sensitive in this patient group.

Outcome Measures

The primary outcomes for this pilot study were the number of eligible adolescents (and their parents/caregivers) recruited compared with the number referred before October 31, 2017, and the proportion of those providing 6-month web-based self-report outcome data (completion of SF-36-PFS). Key qualitative outcomes were acceptability of trial processes and treatments, identified through thematic analysis of interview data.

Intervention Descriptions

FITNET-NHS

This is a web-based modular CFS/ME-specific CBT program designed to be used by young people and their parents or caregivers. It is supported by individualized clinical psychologist e-consultations (with messages sent separately to the young person and their parent or caregiver) delivered within the program itself. Each participant and their parent or caregiver are set up on the platform by the therapist once allocated to this treatment. The participant and parent or caregiver then each set up an independent password-protected log-in. There are up to 19 chapters for young people to work through, which are unlocked by the psychologist on completion of the previous chapter. Some chapters are optional and are only unlocked if the psychologist thinks they are relevant for the young person (eg, a chapter looking more closely at mood problems). The earlier chapters include explanations of the links between thoughts, feelings, and behavior that form patterns contributing to the maintenance of CFS/ME symptoms. The chapters include questions for young people to complete, designed to identify unhelpful patterns and help with problem-solving. Young people are encouraged to monitor their activity, establish a manageable baseline, and build on this gradually. There are diaries included in the program for young people to record their sleep, activity levels, and helpful thoughts, which they can then discuss with therapists. Parents or caregivers can read the content of the chapters but not the answers to the questions. Therapists can view participants' question responses and diaries and they use tailored e-consultations (via message function within the platform) to help the young person through the course. Therapists request that the patient responds before an arranged appointment date within which the therapist will deliver their next detailed and tailored message. Treatment was designed to last approximately 6 months.

AM Via Skype

This is the comparison treatment, delivered by a CFS/ME clinician (usually a physiotherapist/occupational therapist). AM is a standard behavioral treatment offered within the specialist service and recommended by the National Institute for Health and Care Excellence [9]. It involves assessment of the young person's activity level and begins with establishing a manageable baseline of activity to be maintained daily (usually reduction of activity) from which to build gradually and safely at a pace that the patient can manage. The AM intervention offered within the trial is protocolized and explicitly prohibits detailed engagement with cognitions, keeping it as a behavioral treatment. A total of 3 AM sessions were offered as a version of usual care for comparison with the FITNET-NHS intervention. However, in response to feedback, this was increased to 6 sessions from November 2017 onward. Although standard care is usually delivered face-to-face, Skype delivery of at least some of the AM sessions is becoming more routine within the service. For this trial, every aspect was delivered remotely. Treatment was designed to last approximately 6 months.

Integrated Qualitative Methodology

We undertook one-off in-depth interviews with participants and their parents to understand their experiences and views of trial processes. The interviews assessed the provision and acceptability of patient information, treatment preferences, and acceptability of both the content and the delivery of treatments. Participants were purposively selected for maximum variation (intervention, age, and gender) [10]. Families were given a choice of being interviewed over Skype or telephone, together or alone.

Trial staff (recruiters and therapists) for both treatment arms were also interviewed to ascertain their views on the provision of trial information, how the trial treatment compares with standard care, the feasibility of delivering the intervention to children, their view of treatment effectiveness, and potential changes to the interventions offered. Interviews followed a checklist of topics to ensure that key areas were explored but

were sufficiently flexible to allow new issues of importance to the participants to emerge. All interviews were audio recorded with consent using encryption software, transcribed verbatim, and anonymized.

We additionally audio recorded (with consent) the recruitment consultations (second call from research nurses) to identify areas for improvement to optimize recruitment and informed consent in the trial via research nurse training (results to be presented in a separate publication) [11,12].

Qualitative Data Analysis

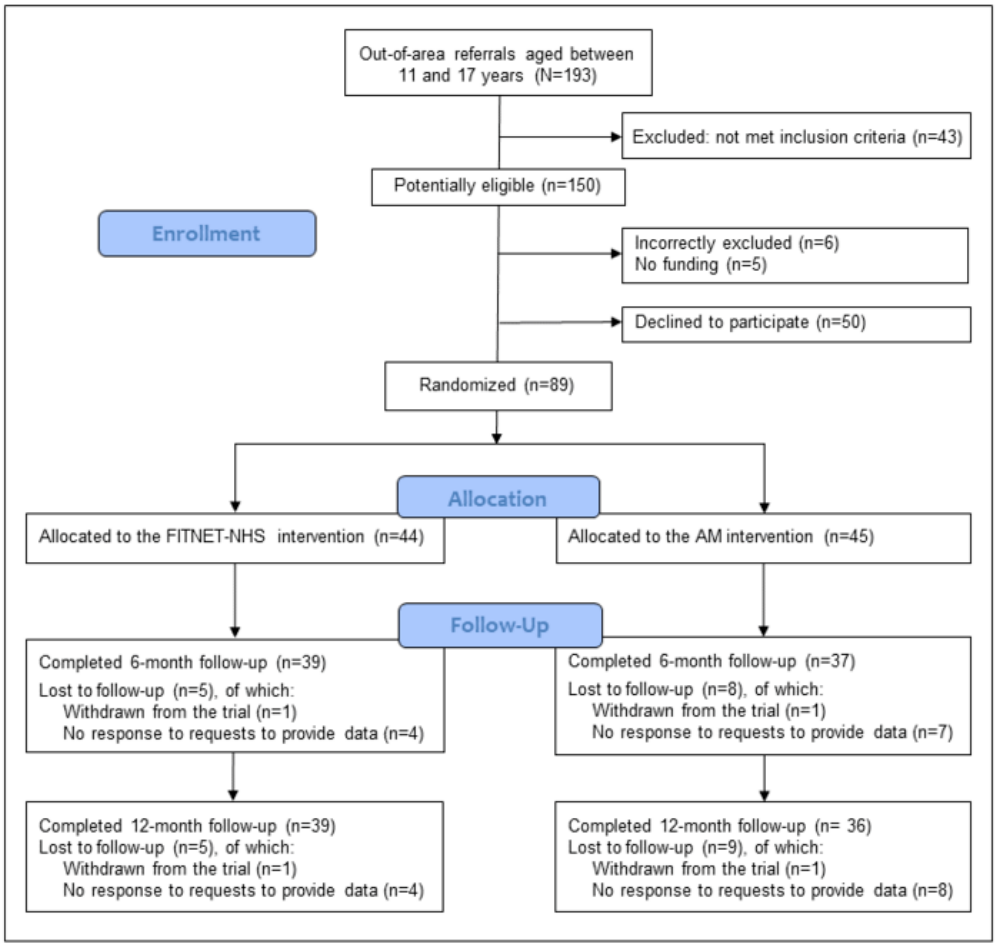
Qualitative data analysis was ongoing and iterative, commencing soon after data collection to inform further data collection [13]. Audio recordings were transcribed verbatim, imported into NVivo (QSR International) software, systematically assigned codes, and analyzed thematically using techniques of constant comparison [14]. For the in-depth interviews with families and therapists, the data were examined for patterns and themes, comparing accounts between different participants, and refining the coding framework as interviews progressed. To check coding reliability, one-fourth of the transcripts were double coded by other members of the team, and the findings were compared. Sources of difficulties identified through the qualitative data were discussed with the trial management group to improve aspects of the design, conduct, organization, or training of recruiters.

Results

Recruitment and Retention

A total of 193 out-of-area patients aged between 11 and 17 years were referred within the internal pilot phase (between November 1, 2016, and October 31, 2017). Of these, 150 out of 193 patients (77.7%) were potentially eligible, and 89 out of 150 patients (59.3%) were recruited into the trial: 44 were randomized to FITNET-NHS and 45 to AM treatment (Figure 1). Those recruited came from a wide range of locations across England. A total of 76 out of 89 patients (85.3%) provided their 6-month outcome data, and 75 out of 89 patients (84.2%) provided their 12-month outcome data. A total of 2 out of 89 patients (2.2%) actively chose to withdraw from the trial.

Figure 1. Consolidated Standards of Reporting Trials flow diagram. AM: activity management; FITNET-NHS: Fatigue in Teenagers on the InterNET in the National Health Service.



Adverse Events

No serious adverse events were reported by the 89 participants referred during the pilot phase of the trial. Nonserious health events were reported by 10 of these participants while taking part in the study, and these were reviewed by the principal investigator, sponsor, and an independent data safety monitoring committee. Only 1 adverse event was assessed as *possibly related* to the trial treatment, where a family felt that some CFS/ME symptoms worsened when following treatment

recommendations. The family took a break from treatment and then returned to it.

Exclusions and Declines

Of the 193 referrals, 43 (22.2%) were excluded at eligibility assessment, of whom 20 had a local specialist service; 12 did not have a confirmed CFS/ME diagnosis (including patients referred to be diagnosed by the South West United Kingdom center); 5 had not had the diagnostic blood tests necessary to rule out other causes of fatigue (4 due to needle phobia and 1 was unresponsive to requests by the research nurses to arrange blood tests); 2 were not disabled by their fatigue; 1 was unable to complete web-based modules due to learning difficulties; and 3 had individual exclusion reasons (clinically complicated requiring face-to-face assessment, in treatment with pain services primarily, and already completed trial treatment and rereferred to the service). After these exclusions and declines, 150 potentially eligible referrals remained.

Within the initial period after trial launch, 6 out of 150 (4.0%) potentially eligible patients were incorrectly excluded (by the clinical team) before reaching eligibility assessment by a research nurse. These patients were offered face-to-face clinical treatment as for the normal treatment pathway (outside of the trial). On discovering this, the clinical team was offered extra training, and standard operating procedures for the administrative handling of out-of-area referrals were improved, which ensured no further incorrect exclusions. A total of 5 out of 150 potentially eligible patients (3.3%) were referred by GPs in Wales, where treatment funding arrangements (between the Welsh Health Boards and the CFS/ME center) prevented these patients from entering the trial.

Of the 150 potentially eligible patients, 50 declined to participate. The main reason was that they wanted to be seen face-to-face, with 24 out of 150 (16.0%) potentially eligible patients preferring to travel to the hospital in South West United Kingdom for standard clinical treatment instead of taking part in the trial. Others declined because of symptom improvement (8/150, 5.3% patients), perceived study burden (3/150, 2.0% patients), unwillingness to use Skype (3/150, 2.0% patients), unwilling to wait for the local pediatrician to confirm CFS/ME diagnosis (3/150, 2.0% patients), or other individual reasons/unknown (9/150, 6.0% patients).

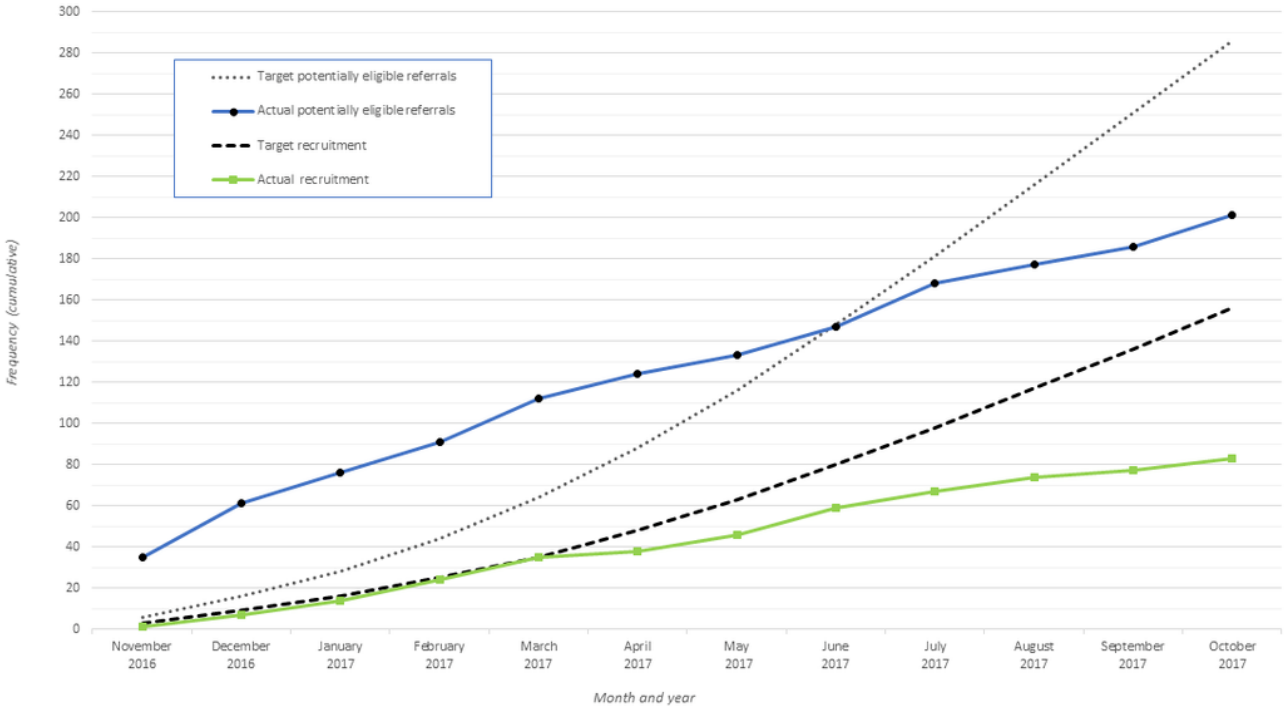
Comparison With Recruitment Projections

In advance of study launch, recruitment projections estimated 286 out-of-area referrals by the end of the first 12 months, expecting 19.9% (57/286) of these to be ineligible and to recruit 67.8% (194/286) of potentially eligible referrals, which would be 156 recruits.

Projections included an initial lag phase leading up to 35 out-of-area referrals per month. National media coverage at study launch resulted in a surge of referrals (well above projected figures), which waned 6 months later, reducing to under half of the projected out-of-area referrals per month (Figure 2). This had a knock-on effect on recruitment, which suffered a lag of 6 months into the trial (Figure 2).

As described, funding pathway issues prevented patients from Wales from accessing the trial. Similar funding pathway issues existed with other devolved nations (Scotland and Northern Ireland). Due to the distance from the center, these nations were not likely to refer patients for routine (face-to-face) clinical treatment at the service, and because of funding pathway issues, plans to promote the remote treatment opportunity across these regions to gain referrals were not able to proceed, which reduced the pool for potential referrals.

Figure 2. Target versus actual referrals and recruits within the 12-month pilot phase. Note: 6 of the patients who were referred within the pilot phase were recruited later - 5 in November 2017 and 1 in January 2018 (presented in consort, not depicted here as the trial was ongoing).



Qualitative Results

Within the pilot phase, interviews were conducted with 20 families (between February 1, 2017, and October 31, 2017; Table 1): 12 families were in the FITNET-NHS arm and 8 were in the AM arm. This included 18 children (12 females and 6 males, ranging in age from 12 to 17 years) and 22 parents (19 mothers and 3 fathers; 2 interviews included both parents). As all adult interviewees were parents, we refer to them as parents (rather than parents or caregivers) in this section. Children and their parent(s) were given the choice of being interviewed alone or together; 14 were interviewed together; 4 separately; and, for 2 families, only the parent was interviewed. A total of 10 families chose to be interviewed via Skype and 10 over the telephone. In all, 4 families canceled interviews as they were unavailable on the day. Six families contacted declined to be interviewed: 3 were too busy because of school, 1 was too ill at the time, and 2 did not wish to participate. A total of 10 interviews were undertaken with trial staff in person on hospital premises: 2 recruiters, 4 AM therapists, and 4 FITNET-NHS therapists.

Table 1. Details of the children and parents interviewed.

ID code	Child age group (years)	Treatment arm (AM[a]: 3 Skype calls; FITNET-NHS[b]: 19 web-based modules)	Interview mode	Parent interviewed	Child/parent interviewed (together/separately)
072	11-12	AM (2nd Skype call)	Skype	Mother	Separately
029	13-15	FITNET-NHS (module 15)	Telephone	Mother	Together
064	13-15	AM (2nd Skype call)	Skype	Mother	Together
082	16-17	FITNET-NHS (module 5)	Telephone	Mother	Together
093	13-15	FITNET-NHS (module 12)	Telephone	Mother	Together
155	13-15	FITNET-NHS (module 15)	Telephone	Mother	Together
198	13-15	FITNET-NHS completed	Skype	Mother and father	Together
209	11-12	AM completed	Skype	Mother	Together
254	13-15	FITNET-NHS (module 17)	Telephone	Mother	Separately
292	16-17	FITNET-NHS (module 15)	Skype	Mother	Together
313	16-17	AM (2nd Skype call)	Telephone	Mother	Parent only
345	13-15	FITNET-NHS (module 15)	Telephone	Mother and father	Together
373	11-12	AM (2nd Skype call)	Skype	Mother	Together
399	13-15	AM (2nd Skype call)	Telephone	Mother	Parent only
401	11-12	FITNET-NHS (module 10)	Skype	Mother	Together
417	11-12	FITNET-NHS (module 15)	Skype	Mother	Together
493	16-17	FITNET-NHS (module 15)	Telephone	Mother	Together
642	16-17	FITNET-NHS (module 14)	Telephone	Father	Together
727	13-15	AM (2nd Skype call)	Skype	Mother	Separately
079	13-15	AM (2nd Skype call)	Skype	Mother	Separately

[a]AM: activity management.

[b]FITNET-NHS: Fatigue in Teenagers on the InterNET in the National Health Service.

Reasons for Participating in the Trial

Families were positive about taking part in the trial, referring to it as a "lifeline," and felt they were "lucky" in the context of "absolutely no treatment" elsewhere. They reported problems before the trial: getting a diagnosis, long waiting lists, funding cuts, and only receiving general advice or self-directed treatments with little improvement. Not having to travel to an appointment was seen as beneficial. Families also wanted to "help other future sufferers".

> *Well the GP has been absolutely awful, basically saying "Oh nobody knows what chronic fatigue is" and sending him away all the time.* [Mother 064]

> *I was willing to try anything.* [Child 029]

> *We liked the idea it was online so that we didn't have to travel.* [Mother 313]

Provision and Acceptability of Patient Information and the Recruitment Process

The written patient information leaflets were found to be clear, thorough, and acceptable. However, some children were too ill to read them, with parents often explaining the study, and some families felt there was too much information. Shorter leaflets or links to a website with more images were suggested as an additional source of information. The *Frequently Asked*

Questions (FAQ)section of the FITNET-NHS website was developed to help this.

> *…we were drip-fed in the appropriate way; so, we had the written stuff, we could ask questions remotely, and we could ask questions when they called us up. It was done very-very well, the layers of information were appropriate.* [Mother 029]

> *I didn't manage to read through all of it but I did read enough and I understood it and I thought that it would be good to try.* [Child 254]

Families were particularly happy with telephone recruitment and described the research nurses as "positive," "understanding," "empathic," and "helpful," allowing them to ask questions. Although some families expressed preferences for treatments, they were often "willing to try anything." Most participants accepted randomization as part of the research process and understood the need for a "fair" comparison. However, some would have liked to choose their treatment, and a few families did not seem to understand randomization, which was fed back into research nurse training for recruitment calls. Participants preferred web-based consenting and data collection as it was "easy" and there was no need to post paper forms.

> *Yeah, I think it explained it all and explained it was two different treatments would be available and you'd just be randomly selected for one, pretty straightforward.* [Mother 082]

> *I think because [child] has had absolutely no treatment at all or help really from anywhere, she saw this as an opportunity so she was going to take it whichever she was given.* [Mother 292]

The Acceptability (Satisfaction and Adherence) of Interventions

Participants valued individually tailored advice from a *specialist* CFS/ME health professional offered in both treatment arms as they had not had the support before. Families and clinicians commented on both the advantages and disadvantages of web-based treatment.

> *The good thing is that you do have somebody to be in touch with us more often, because I felt with [child]'s illness that we were sort of left alone, and we see the paediatrician every six months...and there's no treatment that you can have.* [Mother 254]

Acceptability of AM Skype Calls

Skype was found to be easy to use once set up. Some participants felt Skype was as good as being in a face-to-face appointment. They liked that they did not have to travel and felt being in the home environment was beneficial. Therapists also felt that some patients were more comfortable in their home environment. However, some technical difficulties were often encountered with Skype calls. Skype was felt to work less well for shy children, and some families would have preferred to talk to a health professional in person rather than on Skype. Therapists indicated that a face-to-face appointment in a clinic offered a neutral "safe space" for participants to raise issues of importance and that they are more able to pick up on emotions face-to-face. Issues with confidentiality were also discussed, as it was not always clear to therapists "who is in the room" during a Skype call. Therapists also described how some younger children were harder to engage on Skype.

> *Its [Skype] kinda like face to face...it worked really well and then we didn't have to travel.* [Child 209]

> *Yeah, I mean I think it would better a doctor in person.* [Mother 064]

> *Because they are at their home, they are feeling more relaxed. They haven't had to turn out and travel somewhere, alien environment, being uptight with the traffic. The poor child being exhausted with the travelling, so in many ways you can get a better view of exactly what's going on for them at home.* [Therapist 70005]

> *Sometimes in clinic you're able to pick up on things that might be a little bit more personal. You see that they're very upset about something... It's a little bit more of a safe space to talk about things when they come into clinic, whereas when they're sitting at home, and it's just them two and then me, and sometimes that can be a frozen computer screen or just a black screen because Skype hasn't worked.* [Therapist 70008]

Acceptability of the FITNET-NHS Platform/E-Consultations With a Therapist

Participants liked that they could complete treatment (reading and answering questions on the platform) in their own time rather than having to attend appointments. E-consultations gave them time to think about their answers, and some participants found it easier to talk about personal topics over email. However, others found it difficult to portray things in writing and would have preferred some face-to-face contact. Where therapists received detailed e-consultation replies from patients, they felt they could get "a good picture of them," whereas some commented that the lack of nonverbal communication made it harder to get to know some patients. Some younger children, and particularly their parents, felt that there was too much reading on the platform, and parents often had to help and clarify the meaning of some text. Age differences were also noted by therapists, with parents more involved as a *coach* with the treatment of younger children. Two therapists also described more engagement from girls and sharing of personal information than boys.

> *Yeah, I thought it was a bit strange at the beginning but then it was fine. I think sometimes I found it easier to talk to somebody when it wasn't exactly face to face.* [Child 254]

> *Over an email, sometimes it's quite hard to portray how I feel personally or how I am and how I feel...It's okay, it's really convenient, but like I say, face to face you get the whole how I actually am rather than just words how I am.* [Child 082]

> *…it's much more difficult to get to know what a young person might need or how they might be responding when all you're getting is written information. You're*

> *not having any kind of interaction with the person to know how they're responding and what their non-verbals might be telling you if you have them in the room. I can see when a young person walks into a clinic room with me, I can see if they're looking a bit better than they did last week.* [Therapist 70006]

> *…the younger ones that you maybe have a bit more emphasis on the parent being a coach alongside you and put a bit more effort in to helping the parents with it and then when they're older and you know it's more on the young person themselves.* [Therapist 70002]

Changes to the Trial Treatments Based on Qualitative Feedback

AM

The qualitative interviews indicated that most participants did not feel 3 Skype calls were enough. Clinicians did not feel that the 2 treatment arms were equal and failed to match the current standard for usual care, which had increased to 6 calls per patient since the trial was designed. In response to this feedback, the AM arm was changed to allow up to 6 sessions (submitted as a substantial amendment, gaining ethical approval in October 2017; Multimedia Appendix 1). Other small changes to treatment were made based on feedback, such as splitting the first Skype (assessment) call into 2 shorter sessions and sending a summary email of the Skype call to families to summarize agreed actions.

> *I personally think there should be more [AM sessions] because obviously you are just getting going.* [Mother 373]

> *I do worry a little bit about how equal the two arms are. It does feel like people do FITNET for good or ill really, have a lot more to do...the activity management arm is three Skype sessions...But it feels like they don't feel comparable in terms of therapist input, which can be a factor in itself in terms of outcomes I would imagine.* [Therapist 70004]

FITNET-NHS

Several smaller changes were made to the platform based on feedback received during the pilot phase of the trial (Multimedia Appendix 2). For example, a time-out function existed within the platform for data protection, although this meant that the platform often timed out while a participant was writing a lengthy message (because of remaining on 1 page). Changes were made to warn participants of this. Changes to wording on the platform to clarify meaning were made and clearer instructions accompanied the diaries.

> *[Session] timed out so I had to write the whole email again.* [Child 642]

> *You almost need instructions to understand it [diaries].* [Child 401]

Discussion

Principal Findings

This is the first pilot of specialist web-based treatments for CFS/ME in young people in the United Kingdom, representing the first 12 months of an ongoing large, full national RCT. We demonstrated that it is feasible to recruit children with CFS/ME (and their parents or caregivers) remotely (via telephone screening and web-based consent) and retain them in a trial providing web-based specialist treatment for pediatric CFS/ME.

Strengths and Limitations

A major strength of this study is its use of entirely web-based treatment to enable families across the United Kingdom to gain treatment at home delivered from 1 specialist service. Many families appreciated the opportunity to gain access to treatment, especially when they otherwise would have no access to specialist care. In contrast to the Dutch study, the FITNET-NHS trial used an active specialist treatment comparison group (AM via Skype), rather than treatment as usual or waiting list control, as specifically recommended for this trial by the funders. This is likely to have implications for the results of the full-scale trial (in terms of relative treatment effectiveness) and may also have contributed to the families' willingness to enter the trial, as all recruited participants were offered specialist treatment. The recruitment rates were good, and the families were positive about the remote recruitment processes.

Another strength is the integration of qualitative methods into this RCT to improve recruitment and optimize the interventions in the pilot phase [12,15,16]. These methods helped us understand issues relevant to children and young people and make changes for the full study, including increasing AM sessions in line with usual care and facilitating training of research nurses to improve the recruitment processes. A range of young people of different ages were interviewed from both arms; however, fewer males and fathers were interviewed, which decreased the ability to explore any differences in acceptability [10]. Although we did not interview those declining the trial (which may have been useful to illuminate the reasons), we kept records of the main reasons for declining.

A limitation is the lower recruitment compared with projections. We planned to recruit 156 participants in the in-pilot phase, representing 68% of the potentially eligible patients (expected to be 229) identified from out-of-area referrals aged between 11 and 17 years (expected n=286). We recruited 89 out of 150 participants, representing 59% of the potentially eligible young people identified from 193 referrals. The potentially eligible denominator included 11 referred patients who were excluded due to funding issues or in error (omitting these from the denominator gives us 89 participants recruited out of 139, representing 64% uptake, which is closer to the original projections). The main issue with achieving recruitment targets was the lower than expected number of out-of-area referrals received. Although the CFS/ME service has always accepted out-of-area referrals, most were local referrals. Recruitment projections were based on reaching national pediatric CFS/ME populations at high volumes, made possible by our innovative methodology using remote processes for recruitment and

treatment delivery. Increasing out-of-area referrals was dependent upon maintaining clinician awareness of the trial across the United Kingdom, which proved to be more challenging than expected. The loss of the devolved nations as potential referral sites was part of this picture. In response, detailed plans to increase awareness of the trial across England were put into place to boost referral rates (including drawing on the National Institute for Health Research [NIHR] Clinical Research Network support to publicize the research across the United Kingdom, developing a clinician-facing video to disseminate, and presenting at national GP and pediatrician meetings). These plans, together with evidence of the acceptability of treatments and good retention were reviewed by the independent oversight groups (Trial Steering Committee [on October 3, 2017] and Data Safety Monitoring Committee [on September 5, 2017]) and the funding body (NIHR Health Technology Assessment [HTA]). These groups recommended that the study proceed to full trial after the end of the pilot phase (from November 1, 2017).

Results in the Context of Previous Literature

Recruitment rates of families referred to the service were good, with over half of potentially eligible referrals enrolled in the trial. Of the families declining to take part, approximately half opted to travel for face-to-face treatment rather than accepting web-based treatment. This is in line with other research in adults, which had similar rates of uptake [17]. By comparison, not having to travel for treatment was perceived as a benefit by those who took part. Retention at 6 months was good (76/89, 85%) and was maintained up to 12 months, with only 2 participants requesting withdrawal from the trial.

The 2 web-based treatment options—FITNET-NHS (web-based CBT) and Skype-delivered AM—are acceptable to young people with CFS/ME and their parents/caregivers. Access to tailored advice from a *specialist* CFS/ME health professional was particularly valuable to participants, and therapist support has been found to be important for increasing engagement and adherence to digital interventions in mental health [18]. Both advantages (eg, less travel and home environment) and disadvantages (eg, technical problems with Skype and preference for face-to-face contact) were discussed by participants. Other research identified a similar mix of positive and negative patient experiences of CFS/ME treatment delivery via videoconferencing [19].

This study builds on the promising results of the Dutch trial of the FITNET program [5] by demonstrating the potential for the UK-adaptation of the web-based CBT intervention to be used within the NHS. The full RCT is currently ongoing, with recruitment due to be completed by October 31, 2020 (with follow-up data collection to continue for 12 months beyond this). The results of the full trial will indicate whether the FITNET-NHS treatment is effective and cost-effective when compared with web-based AM, which will inform treatment recommendations going forward.

Conclusions

It is possible to recruit and retain families in a CFS/ME treatment trial using telephone and web-based methods. Many families are willing to accept web-based treatment. This supports the potential for remote delivery of treatment and trials for different clinical populations who cannot access local services where such an approach could benefit patients.

Acknowledgments

The trial is funded by the NIHR HTA program (award number: 14/192/109). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The sponsor is the University of Bristol. The trial is monitored and audited in accordance with the sponsor's procedures. Indemnity for the trial is provided by the University of Bristol. This study was designed and delivered in collaboration with the Bristol Randomised Trials Collaboration, a UK Clinical Research Collaboration-registered clinical trials unit, which, as part of the Bristol Trials Centre, is in receipt of NIHR CTU support funding. Study data were collected and managed using REDCap (Research Electronic Data Capture) software, version 9.5.23 (Vanderbilt University), hosted at the University of Bristol.

Authors' Contributions

EC, the chief investigator, was involved in trial conception, was responsible for the trial and reviewed the manuscript. EA, the trial manager, oversaw the trial as part of the Trial Management Group (TMG) and drafted and reviewed the manuscript. RP served as a qualitative researcher, had oversight over the trial as part of the TMG, and was involved in drafting and reviewing the manuscript. The lead on health economics, WH, oversaw the trial as part of the TMG and reviewed the manuscript. SP and KP, leads on the IT platform, had oversight over the trial as part of the TMG and reviewed the manuscript. NM, the lead on qualitative methods, oversaw the trial as part of the TMG and reviewed the manuscript. A qualitative researcher, LB, had oversight over the trial as part of the TMG and reviewed the manuscript. DG served as the trial statistician, was involved in confidential reporting to the data monitoring committee, and reviewed the manuscript. CM, the statistical lead, had oversight over the trial as part of the TMG and reviewed the manuscript. DK, lead on recruitment from primary care, had oversight over the trial as part of the TMG and reviewed the manuscript. JM had oversight over the trial as part of the TMG, was the lead on data linkage, and reviewed the manuscript review. PS, the clinical psychology support, was involved in overseeing the trial as part of the TMG and reviewed the manuscript. The lead clinical psychologist, HK, was involved in overseeing the trial as part of the TMG, training and supervising therapists, and reviewed the manuscript. EV, SN, and GB served as FITNET advisers, had oversight over the trial as part of the TMG, and reviewed the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

HK and GB received royalties for the publication of a treatment manual for CBT in CFS/ME in adults. The remaining authors have no conflicts of interest to disclose.

Multimedia Appendix 1
Details of approved study amendments.
[DOCX File , 20 KB - jmir_v22i8e17768_app1.docx]

Multimedia Appendix 2
Changes made to the FITNET-NHS Platform.
[DOCX File , 20 KB - jmir_v22i8e17768_app2.docx]

Multimedia Appendix 3
EHEALTH CONSORT checklist (v1.6.1).
[PDF File (Adobe PDF File), 3309 KB - jmir_v22i8e17768_app3.pdf]

References

1. Crawley EM, Emond AM, Sterne JA. Unidentified chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) is a major cause of school absence: surveillance outcomes from school-based clinics. BMJ Open 2011;1(2):e000252 [FREE Full text] [doi: 10.1136/bmjopen-2011-000252] [Medline: 22155938]
2. Crawley E, Hughes R, Northstone K, Tilling K, Emond A, Sterne JA. Chronic disabling fatigue at age 13 and association with family adversity. Pediatrics 2012 Jul;130(1):e71-e79. [doi: 10.1542/peds.2011-2587] [Medline: 22711716]
3. Rangel L, Garralda ME, Levin M, Roberts H. The course of severe chronic fatigue syndrome in childhood. J R Soc Med 2000 Mar;93(3):129-134 [FREE Full text] [doi: 10.1177/014107680009300306] [Medline: 10741312]
4. Crawley E, Sterne JA. Association between school absence and physical function in paediatric chronic fatigue syndrome/myalgic encephalopathy. Arch Dis Child 2009 Oct;94(10):752-756. [doi: 10.1136/adc.2008.143537] [Medline: 19001477]
5. Nijhof SL, Bleijenberg G, Uiterwaal CS, Kimpen JL, van de Putte EM. Effectiveness of internet-based cognitive behavioural treatment for adolescents with chronic fatigue syndrome (FITNET): a randomised controlled trial. Lancet 2012 Apr 14;379(9824):1412-1418. [doi: 10.1016/S0140-6736(12)60025-7] [Medline: 22385683]
6. Baos S, Brigden A, Anderson E, Hollingworth W, Price S, Mills N, et al. Investigating the effectiveness and cost-effectiveness of FITNET-NHS (fatigue in teenagers on the interNET in the NHS) compared to Activity Management to treat paediatric chronic fatigue syndrome (CFS)/myalgic encephalomyelitis (ME): protocol for a randomised controlled trial. Trials 2018 Feb 22;19(1):136 [FREE Full text] [doi: 10.1186/s13063-018-2500-3] [Medline: 29471861]
7. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
8. Parslow R, Patel A, Beasant L, Haywood K, Johnson D, Crawley E. What matters to children with CFS/ME? A conceptual model as the first stage in developing a PROM. Arch Dis Child 2015 Dec;100(12):1141-1147 [FREE Full text] [doi: 10.1136/archdischild-2015-308831] [Medline: 26453575]
9. Chronic Fatigue Syndrome/Myalgic Encephalomyelitis (or Encephalopathy): Diagnosis and Management Clinical Guideline [CG53]. National Institute for Health and Care Excellence (NICE). 2007 Aug 22. URL: https://www.nice.org.uk/guidance/cg53 [accessed 2019-10-15]
10. Ritchie J, Lewis J, Nicholls C, Ormston R, editors. Qualitative Research Practice: a Guide for Social Science Students and Researchers. Second Edition. London, UK: Sage Publications; 2014.
11. Donovan JL, Paramasivan S, de Salis I, Toerien M. Clear obstacles and hidden challenges: understanding recruiter perspectives in six pragmatic randomised controlled trials. Trials 2014 Jan 6;15:5 [FREE Full text] [doi: 10.1186/1745-6215-15-5] [Medline: 24393291]
12. Donovan JL, Rooshenas L, Jepson M, Elliott D, Wade J, Avery K, et al. Optimising recruitment and informed consent in randomised controlled trials: the development and implementation of the quintet recruitment intervention (QRI). Trials 2016 Jun 8;17(1):283 [FREE Full text] [doi: 10.1186/s13063-016-1391-4] [Medline: 27278130]
13. Glaser B, Strauss A. The Discovery of Grounded Theory: Strategies for Qualitative Research. New York, USA: Aldine; 1967.
14. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
15. Elliott D, Husbands S, Hamdy FC, Holmberg L, Donovan JL. Understanding and improving recruitment to randomised controlled trials: qualitative research approaches. Eur Urol 2017 Nov;72(5):789-798 [FREE Full text] [doi: 10.1016/j.eururo.2017.04.036] [Medline: 28578829]

16. O'Cathain A, Thomas KJ, Drabble SJ, Rudolph A, Hewison J. What can qualitative research do for randomised controlled trials? A systematic mapping review. BMJ Open 2013 Jun 20;3(6):e002889 [FREE Full text] [doi: 10.1136/bmjopen-2013-002889] [Medline: 23794542]
17. Janse A, Worm-Smeitink M, Bleijenberg G, Donders R, Knoop H. Efficacy of web-based cognitive-behavioural therapy for chronic fatigue syndrome: randomised controlled trial. Br J Psychiatry 2018 Feb;212(2):112-118. [doi: 10.1192/bjp.2017.22] [Medline: 29436329]
18. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
19. Haig-Ferguson A, Loades M, Whittle C, Read R, Higson-Sweeney N, Beasant L, et al. 'It's not one size fits all'; the use of videoconferencing for delivering therapy in a specialist paediatric chronic fatigue service. Internet Interv 2019 Mar;15:43-51 [FREE Full text] [doi: 10.1016/j.invent.2018.12.003] [Medline: 30619719]

Abbreviations

AM: activity management
CBT: cognitive behavioral therapy
CFS/ME: chronic fatigue syndrome/myalgic encephalomyelitis
FITNET: Fatigue in Teenagers on the InterNET
FITNET-NHS: Fatigue in Teenagers on the InterNET in the National Health Service
GP: general practitioner
HTA: Health Technology Assessment
NHS: National Health Service
NIHR: National Institute for Health Research
RCT: randomized controlled trial
SF-36-PFS: Short Form Health Survey Physical Function Subscale
TMG: Trial Management Group

Edited by G Eysenbach; submitted 10.01.20; peer-reviewed by S Patel, L Larun; comments to author 05.03.20; revised version received 24.04.20; accepted 14.05.20; published 12.08.20.

Please cite as:
Anderson E, Parslow R, Hollingworth W, Mills N, Beasant L, Gaunt D, Metcalfe C, Kessler D, Macleod J, Pywell S, Pitts K, Price S, Stallard P, Knoop H, Van de Putte E, Nijhof S, Bleijenberg G, Crawley E
Recruiting Adolescents With Chronic Fatigue Syndrome/Myalgic Encephalomyelitis to Internet-Delivered Therapy: Internal Pilot Within a Randomized Controlled Trial
J Med Internet Res 2020;22(8):e17768
URL: https://www.jmir.org/2020/8/e17768
doi:10.2196/17768
PMID:32784188

Original Paper

Impact of a Multicomponent Digital Therapeutic Mobile App on Medication Adherence in Patients with Chronic Conditions: Retrospective Analysis

Elyssa Wiecek[1], BS, PharmD; Andrea Torres-Robles[1], BPharm; Rachelle Louise Cutler[1], BSc, MPharm, PhD; Shalom Isaac Benrimoj[2], BPharm, PhD; Victoria Garcia-Cardenas[1], BPharm, MPharm, PhD

[1]Graduate School of Health, University of Technology Sydney, Sydney, Australia

[2]Pharmaceutical Care Research Group, University of Granada, Granada, Spain

Corresponding Author:
Victoria Garcia-Cardenas, BPharm, MPharm, PhD
Graduate School of Health
University of Technology Sydney
Sydney,
Australia
Phone: +61 (02) 9514 9297
Email: victoria.garciacardenas@uts.edu.au

Abstract

Background: Strategies to improve medication adherence are widespread in the literature; however, their impact is limited in real practice. Few patients persistently engage long-term to improve health outcomes, even when they are aware of the consequences of poor adherence. Despite the potential of mobile phone apps as a tool to manage medication adherence, there is still limited evidence of the impact of these innovative interventions. Real-world evidence can assist in minimizing this evidence gap.

Objective: The objective of this study was to analyze the impact over time of a previously implemented digital therapeutic mobile app on medication adherence rates in adults with any chronic condition.

Methods: A retrospective observational study was performed to assess the adherence rates of patients with any chronic condition using Perx Health, a digital therapeutic that uses multiple components within a mobile health app to improve medication adherence. These components include gamification, dosage reminders, incentives, educational components, and social community components. Adherence was measured through mobile direct observation of therapy (MDOT) over 3-month and 6-month time periods. Implementation adherence, defined as the percentage of doses in which the correct dose of a medication was taken, was assessed across the study periods, in addition to timing adherence or percentage of doses taken at the appropriate time (±1 hour). The Friedman test was used to compare differences in adherence rates over time.

Results: We analyzed 243 and 130 patients who used the app for 3 months and 6 months, respectively. The average age of the 243 patients was 43.8 years (SD 15.5), and 156 (64.2%) were female. The most common medications prescribed were varenicline, rosuvastatin, and cholecalciferol. The median implementation adherence was 96.6% (IQR 82.1%-100%) over 3 months and 96.8% (IQR 87.1%-100%) over 6 months. Nonsignificant differences in adherence rates over time were observed in the 6-month analysis (Fr(2)=4.314, *P*=.505) and 3-month analysis (Fr(2)=0.635, *P*=.728). Similarly, the timing adherence analysis revealed stable trends with no significant changes over time.

Conclusions: Retrospective analysis of users of a medication adherence management mobile app revealed a positive trend in maintaining optimal medication adherence over time. Mobile technology utilizing gamification, dosage reminders, incentives, education, and social community interventions appears to be a promising strategy to manage medication adherence in real practice.

(J Med Internet Res 2020;22(8):e17834) doi:10.2196/17834

KEYWORDS

medication adherence; medication compliance; mobile phone; mobile apps; mHealth; gamification

Introduction

Strategies to manage medication adherence, which is defined as the process by which patients take their medications as prescribed [1], are widespread in the literature and are reported to be modestly effective [2]. Most likely due to the multidimensional nature of medication-taking behavior and numerous determinants of nonadherence [3], multicomponent interventions with both technical and educational aspects have shown the most success [4]. However, these strategies have failed to find success in the real world; patient adherence levels tend to decrease in the long term and stay consistently at around 50% [5]. These strategies are limited not only by the capacity of the health care system delivering them but also by low levels of patient engagement. Even when patients are aware of the risks and consequences of diseases, few engage persistently in therapies to improve health outcomes [6-8].

Cognitive biases resulting in irrational and unhealthy behavior may be a key contributor to patient engagement in preventative health strategies. In contrast with traditional economic models of rational choice, modern insights have suggested that human behavior is highly influenced by the context or environment of our decision-making process rather than by price signals or factual information [9]. The field of behavioral economics combines psychology and neoclassical economics to shed light on the errors in mental processing that prevent patients from making rational and beneficial decisions to improve their health [10]. Some health behaviors may require high levels of self-control, meaning that a patient may need to endure "certain and immediate inconveniences in return for uncertain and distant benefits [11]." Obvious behaviors that create this paradigm are healthy food choices and exercise [12]. However, medication adherence, or the act of taking a medication at a certain time each day, creates inconvenience by disrupting the patient's daily lifestyle or causing adverse effects; meanwhile, this behavior is only rewarded with uncertain and distant future health outcomes.

Strategies to influence cognitive biases include incentives and rewards. Incentives and rewards not only impact motivation but also create an immediate benefit to counteract inconvenience [13]. In previous literature, financial incentives showed success in improving medication adherence but were limited by long-term viability and capacity of resources, with economic incentives often eroding the potential economic gain [13-19]. The use of lottery-based incentives has also shown success in sustainment of adherence and long-term engagement [20,21]. Frequent lotteries with small rewards can engage patients based on regret aversion, namely the understanding that the emotional cost of regret (ie, missing a reward by not taking a medication dose) is significant [21].

Methods of gamification or use of nonfinancial extrinsic motivators, such as accruing "points," can be feasible and practical ways to create similar senses of gratification and motivation [22]. Gamification is the application of game elements for purposes other than their expected use for entertainment [23]. An individual's choice to engage in an activity is affected by extrinsic and intrinsic motivation. Medication adherence requires intrinsic motivation driven by internal rewards; this sense of motivation can often be difficult to achieve for behavior that has uncertain and distant health benefits. Through the use of gamification, extrinsic motivators such as earning points and monetary rewards can create and trigger internal motivation [23]. Gamification is not only able to use both extrinsic and intrinsic motivation to create consistent engagement through rewards, such as points or daily streaks, but can also create a sense of achievement [23]. Both gamification and rewards appear to be promising strategies to potentiate the effects of frequently used adherence management approaches, such as educational components and reminders.

Currently, over 300,000 mobile health (mHealth) apps are available; they have become common and instrumental tools for health behavior change in modern times [24,25]. Success has already been demonstrated with using mobile phone apps to support health behavior changes, ranging from constructing a healthy diet to managing chronic pain or improving physical activity [26-28]. Despite the potential of mobile phone apps as a tool to manage medication adherence, there is limited evidence of the effectiveness of these innovative interventions [29-31]. Real-world evidence, which refers to health care information gathered outside clinical research settings, can help minimize this evidence gap. Generated through the analysis of multiple sources, including electronic health records and mHealth apps, real-world evidence can be used to test how health interventions work in usual practice [32]. Observational studies of real-world data can assist in evaluating the potential impact of implemented health interventions in real world settings, such as interventions delivered through mobile phone apps [33].

The objective of this study was to use real-world data to analyze the impact over time of a previously implemented digital therapeutic mobile app on medication adherence rates in adults with any chronic condition. The impact on timing adherence rates was also analyzed.

Methods

Study Design

This was a retrospective observational study using real-world data. The implementation adherence of people in Australia using a commercially available smartphone application, Perx, was evaluated. The ESPACOMP Medication Adherence Reporting Guideline (EMERGE) and STrengthening the Reporting of OBservational Studies in Epidemiology (STROBE) Statement were used [1,34].

Intervention: Perx Digital Therapeutic

Perx is a digital therapeutic that uses different components within a mobile app to improve adherence to medications. These include technical components (through dosage reminders based on the individual patient's dosing regimen and individualized visual adherence feedback), educational components (through the use of educational materials on the disease and medications used), incentives and rewards (lottery-style delivery of gift cards), a social community (through a chat forum and collaborative competition dynamics), and gamification (through the use of point-earning and minigames to enhance the

medication-taking experience). Perx enables users to input their medication schedule information while sending dosage reminders based on the individual patient's regimen. Doses taken are self-reported and recorded by mobile direct observation of therapy (MDOT) photo verification [35]. "Gold" points are rewarded to users for each dose taken on time (±1 hour). Additionally, different minigames are offered at the time of a medication dose to enhance the medication-taking experience. The patient can earn extra gold points through learning a daily fact about their medication or disease state and by completing all daily tasks. Supplementary tasks within the app include health measurements, appointment reminders, physical therapy sessions, and other health actions, which provide users with a comprehensive system to track their health in addition to visual adherence feedback on their personal progress. A social forum and leaderboard component are also included, which create a Perx community. Reward shopping vouchers for popular stores can be redeemed either with a certain amount of gold earned or randomly by taking a correct dose. Screenshots showing the different features of the app can be found in Multimedia Appendix 1.

Data Source and Patients

Deidentified user data from the Perx database were analyzed for this study to assess adherence dosing data between October 2018 and May 2019 within Australia. All information was deidentified, including medications, doses, schedules, user age, dosages taken and missed, and timestamps of dosages taken.

Users were recruited to use the app via a range of channels, including patient advocacy organizations (ie, Cystic Fibrosis Australia and Diabetes NSW & ACT), local community pharmacies, outpatient clinics at local hospitals, and app stores. App users with any chronic condition were included in the analysis. Two user cohorts were analyzed: one for users who used the app consistently for over 6 months and one for users who used the app consistently for 3 months. Users were excluded from the analysis if they used the intervention for less than 30% of the time period defined by the number of days active on the app. The 30% threshold was used because it excluded patients who appeared to decide to stop using the app during the time period of the analysis, as the objective was to analyze user medication adherence rather than adherence to the app itself.

A subanalysis of timing adherence was also performed for both time periods. Users were excluded from the subanalysis if timestamps were not available for the entire time period.

Outcome: Medication Adherence

Adherence implementation rates (where adherence implementation was defined as the extent to which a patient's actual dosing corresponded to the prescribed dosing regimen [1]) were calculated by dividing doses taken by total doses scheduled per 30-day period. This included doses taken outside the ±1-hour time period and was verified by comparing the recorded timestamps to the dosing schedules inputted within the app.

For the subanalysis, timing adherence was assessed with doses taken at the correct time (±1 hour) over total doses scheduled per 30-day period. This additional analysis was performed to understand the effects of the incentives, as users could only redeem incentives if the medication was taken within the ±1 hour time threshold. Both adherence measures are presented as percentages. Rates were compared to an optimal adherence level of 80%, which is the most commonly used cutoff point in the literature [36,37].

Data Analysis

Data were analyzed by integrating the PROC SQL (SAS University Edition 9.4) and Python (Jupyter Lab 1.0) language programs and Microsoft Excel 2019 (Microsoft Corporation) to organize and retrieve the results. The analysis was conducted in 30-day time periods. Study variables were summarized using mean (SD) and median (IQR). Adherence variables were verified for normal distribution using the Shapiro-Wilk test. Due to the distribution of the data, the Friedman test was used to compare differences in adherence rates over time. A P value <.05 was considered to indicate statistical significance.

Ethics Statement

The University of Technology Sydney Human Research Ethics Committee (HREC) approved this study (ETH19-3622). All users recruited into the program were required to actively accept and consent to the Terms of Use and Privacy Policy, which stated that de-identified data in aggregated form may be used by third parties for research and other purposes. No personal or confidential data were included in the database; therefore, informed patient consent was not required.

Results

Study Sample

A total of 130 users were included in the 6-month analysis, and 243 users were included in the 3-month analysis. For the timing adherence subanalysis, 111 users and 221 users were included in the 6-month and 3-month analyses, respectively.

6-Month Analysis Group

The distribution of users according to gender was 36/130 male (27.7%) and 88/130 (67.7%); 6/130 users (4.6%) did not disclose their gender. The average age was 45.8 years (SD 17.2). The most common medications prescribed were rosuvastatin, cholecalciferol, and atorvastatin; the mean number of medications prescribed per patient was 4.3 (SD 3.1).

3-Month Analysis Group

The distribution of users according to gender was 80/243 male (32.9%) and 156/243 female (64.2%); 7/243 users (2.9%) did not disclose their gender). The average age was 43.8 years (SD 15.5). The most common medications prescribed were varenicline, rosuvastatin, and cholecalciferol; the mean number of medications prescribed per patient was 4.0 (SD 2.9).

Implementation Adherence

Adherence rates across the 6-month time period are shown in Table 1. The overall median implementation adherence was 96.8% (IQR 87.1%-100%) across 6 months. A small decreasing trend was observed from month 4 to month 6. However, the

Friedman test revealed non-significant differences in adherence rates over time ($F_r(2)=4.314$, $P=.505$) (Figure 1).

Adherence rates across the 3-month time period are shown in Table 1. The overall median implementation adherence was 96.6% (IQR 82.1%-100%) across 3 months. A slight decreasing trend was seen from month 1 to month 3 (Figure 2). Similarly to the 6-month analysis, nonsignificant differences in adherence rates over time were found ($F_r(2)=0.635$, $P=.728$).

Table 1. Adherence rates across the 6-month and 3-month time periods.

Study period	Mean (SD)	Median (IQR)
6-month analysis (%)		
Month 1	88.6 (21.5)	96.8 (88.0-100)
Month 2	88.0 (20.3)	96.8 (82.5-100)
Month 3	89.5 (18.5)	97.1 (87.1-100)
Month 4	88.6 (20.9)	98.3 (86.5-100)
Month 5	87.0 (24.0)	97.1 (85.7-100)
Month 6	83.9 (26.9)	96.8 (83.9-100)
Overall	87.6 (16.9)	96.8 (87.1-100)
3-month analysis (%)		
Month 1	87.3 (21.1)	96.1 (86.1-99.6)
Month 2	84.1 (24.7)	96.8 (79.0-100)
Month 3	82.5 (27.5)	96.7 (80.6-100)
Overall	84.6 (20.9)	96.6 (82.1-100)

Figure 1. Mean implementation adherence rates of 130 users of the Perx app over 6 months. The shaded area below 80% indicates less than optimal adherence based on the literature.

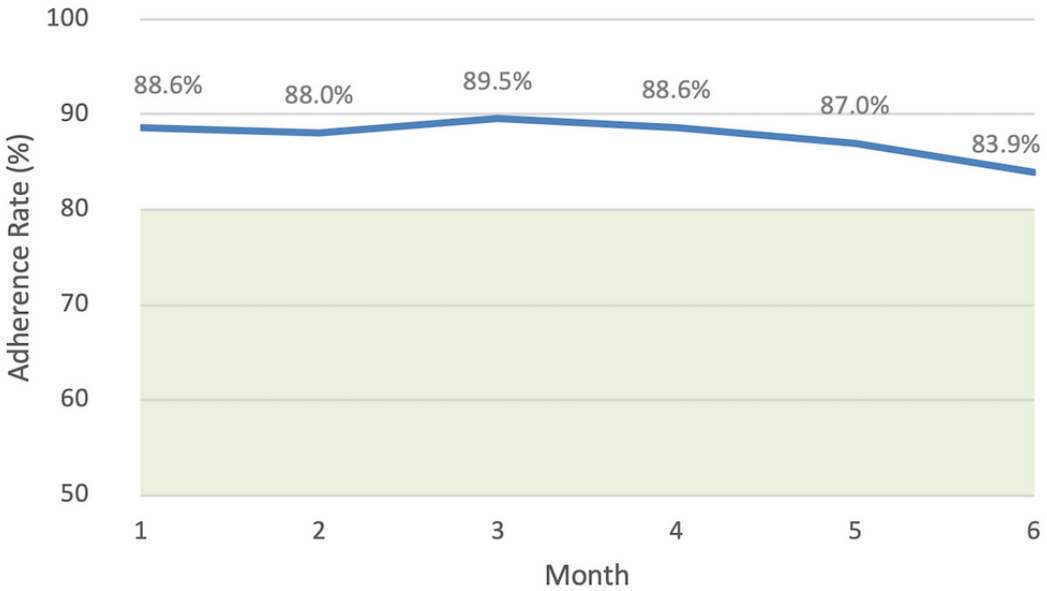
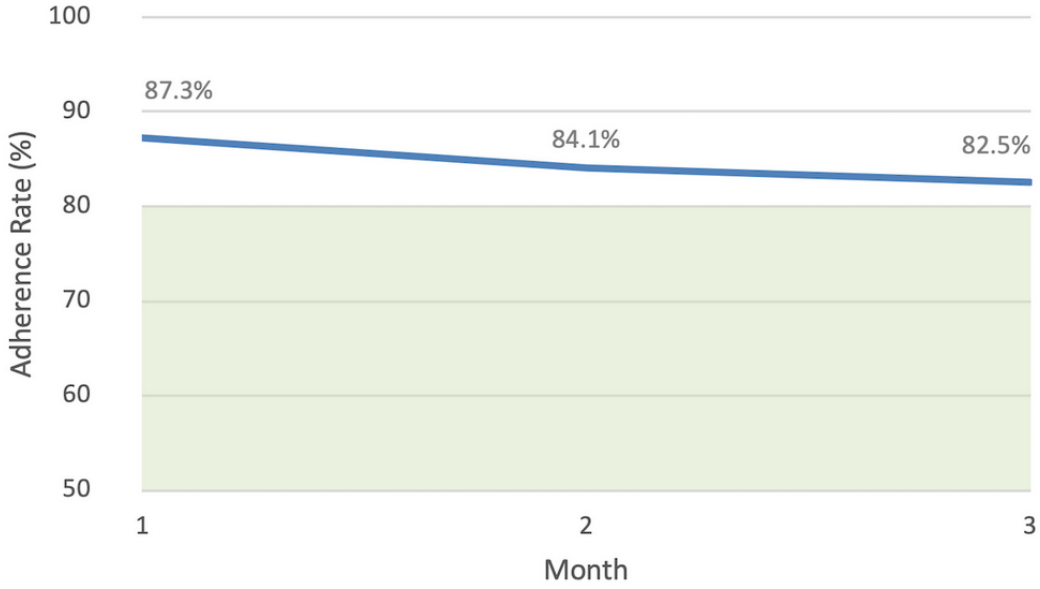


Figure 2. Mean implementation adherence rates of 243 users of the Perx app over 3 months. The shaded area below 80% indicates less than optimal adherence based on the literature.



Timing Adherence Subanalysis

Timing adherence rates across study time periods can be found in Table 2. For the 111 users included in the 6-month timing adherence analysis, their adherence remained unchanged, with medians of 77.3% (IQR 52.0%-93.1%) in month 1 and 77.4% (IQR 36.2%-94.4%) in month 6. The median value across the time periods was 79.0% (IQR 50.8%-92.9%). Overall, there were no significant changes over time ($F_r(2)=5.465$, $P=.362$) (Figure 3).

In the 3-month timing adherence analysis, 221 users' adherence remained stable (Table 2), with nonsignificant changes across time periods ($F_r(2)=2.125$, $P=.346$) (Figure 4).

Table 2. Timing adherence across the 6-month and 3-month time periods.

Study period	Mean (SD)	Median (IQR)
6-month analysis (%)		
Month 1	68.4 (27.9)	77.3 (52.0-93.1)
Month 2	70.5 (28.4)	82.3 (54.8-91.9)
Month 3	69.2 (27.7)	79.8 (50.3-93.5)
Month 4	69.8 (28.8)	81.7 (51.4-93.5)
Month 5	68.7 (28.9)	80.6 (52.8-92.7)
Month 6	63.4 (33.7)	77.4 (36.2-94.4)
Overall	68.5 (29.1)	79.0 (50.8-92.9)
3-month analysis (%)		
Month 1	63.7 (28.2)	71.0 (46.4-85.9)
Month 2	64.0 (30.8)	74.2 (41.9-90.3)
Month 3	61.4 (32.3)	71.0 (34.7-88.7)
Overall	61.1 (28.5)	72.0 (41.8-88.3)

Figure 3. Mean timing adherence rates of 111 users of the Perx app over 6 months. The users were considered to be adherent to the dose if it was taken within ±1 hour of the scheduled time.

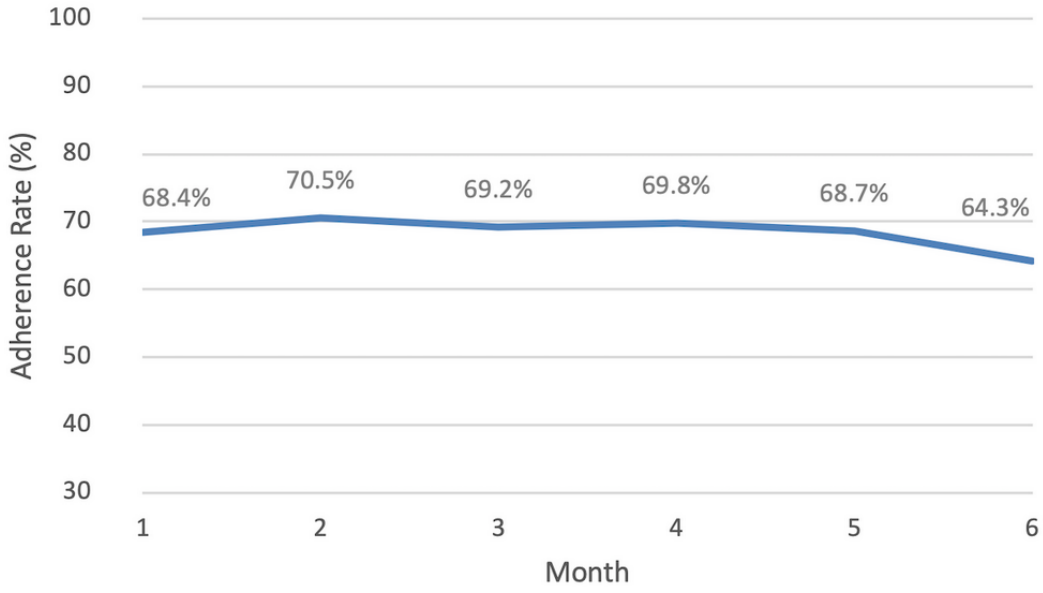
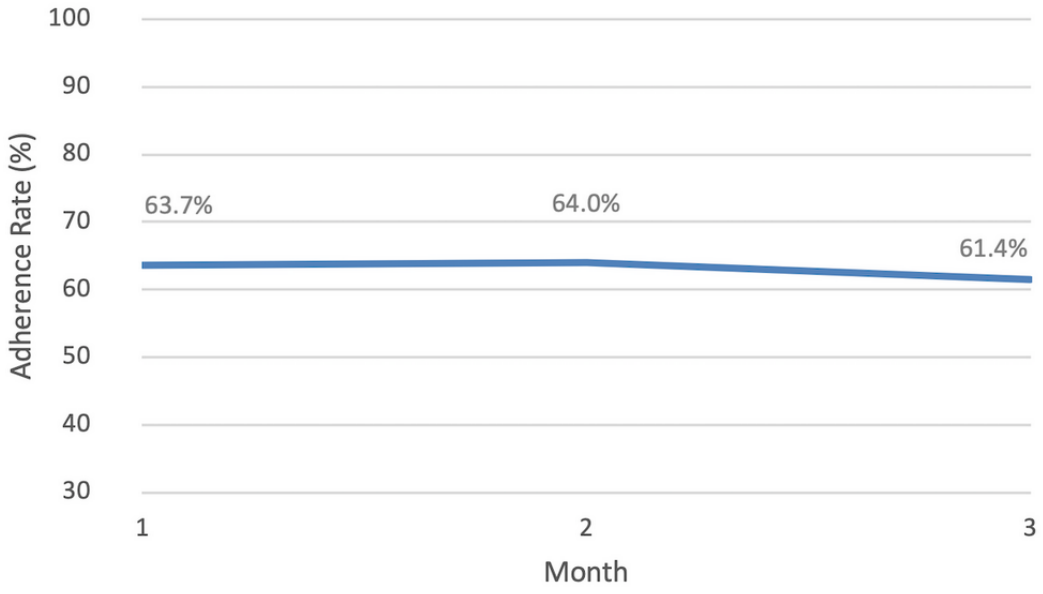


Figure 4. Mean timing adherence rates of 221 users of the Perx app over 3 months. The users were considered to be adherent to the dose if it was taken within ±1 hour of the scheduled time.



Discussion

Principal Findings

Retrospective analysis of the medication adherence of users receiving a multicomponent adherence management intervention that includes reminders, educational components, incentives, gamification, and social community components demonstrated that this intervention is a successful approach to maintaining optimal medication adherence. To our knowledge, this is the first study investigating a comprehensive multicomponent mobile intervention to maintain medication adherence across different chronic conditions.

Trends observed from the users of the mobile app showed high rates of adherence across the study periods. The adherence rates of Perx users averaged over 85% across six months. This was significantly higher than previously observed dispensing data adherence rates in Australian patients, which were found to be between 50.2% and 66.9% [38]. While a slight decrease in adherence was observed over 6 months, the long-term rates remained above 80%, which is often considered to be an optimal threshold for medication adherence [36]. The decrease in adherence rates was found to be statistically insignificant [36][39]. The gradual decrease was less pronounced than that in previous literature examining the long-term effects and multidimensional, dynamic nature of medication adherence; in a previous study, average adherence was estimated to decrease by 1.1% per month [40]. This suggests that the addition of gamification and incentive components to more traditional management interventions (eg, educational components and reminders) is a viable option to inspire long-term motivation and adherence to medications.

A recent network meta-analysis examined the impact of adherence interventions across time and identified multicomponent interventions as the most effective long-term solution [4]. Although there is limited evidence, interventions that include incentives and technical aspects (ie, dosage reminders) have been shown to be the most effective in sustaining long-term results [4]. The Perx digital therapeutic presents an advantageous alternative to existing medication adherence interventions due to its incorporation of multiple and innovative components into one platform to continuously motivate and empower users. A main component of the Perx app, medication reminders, has long been identified as a successful intervention component to improve adherence [41,42]. However, although medication reminders help to enhance adherence, they only affect one dimension of the multiple nonadherence determinants and are frequently used in combination with additional interventions, such as education [4]. Educational interventions are also a common long-term strategy to improve adherence to medications [2,43]. Delivered by numerous methods, these interventions can be moderately effective; however, they are not a sole solution to improve adherence for all patients [44]. When combined with technical and attitudinal components such as motivational interviewing, education-based strategies are found to be even more successful [4].

Motivation is another common determinant of medication adherence [3]. Patients can be fully aware of the positive health benefits medications provide as well as the consequences of poor health behavior; however, some patients consistently make poor health choices [11]. Present-biased preferences explain the "human tendency to grab immediate rewards and to avoid immediate costs in a way that our 'long-run selves' do not appreciate [10]." An individual may analyze immediate costs or immediate rewards to make a decision; such decisions often result from impatience or immediate gratification and place greater value on achieving gratification in the present moment than obtaining the same reward in the future. Positive and negative health outcomes remain too distant of a reward and consequence, respectively [45]. The Perx digital therapeutic aims to create instant gratification through gamification elements. Through receiving instant praise and reward after each medication dose taken on time, users may be motivated to continue to be adherent. Motivation can additionally be created through intrinsic forces, as stated by the self-determination theory. The self-determination theory suggests that the nature of perceptible motivational types determines the predictability and force of how people behave, rather than the amount of motivation [46,47]. Therefore, it is necessary for gamified systems to promote a sense of autonomy, competence, and relatedness to create the intrinsic motivation needed to continue the value of the extrinsic motivating factors [46,48]. The Perx digital therapeutic intervention may be successful because it meets the users' need for competence, autonomy, and relatedness. Competence and autonomy are created by setting challenging yet manageable goals, where adherence is the challenge and financial incentives are the goals. Users can also follow their progress through points, leaderboards, and personal visualized feedback graphs on their individual adherence. This feedback provides additional positive reinforcement and has been proven to be a successful component of interventions to improve medication adherence; it is estimated that adherence increases 8.8% for interventions where feedback is included compared to those that do not include feedback [40,49]. The social community component meets the need for relatedness by fostering a feeling of belonging to a community that shares the common goal of better health [46].

While gamification is a main force in creating motivation in the app, the impact of rewards and incentives cannot be dismissed. The use of incentives in public policy has long been used as an extrinsic force to influence behavior and intrinsic motivation [9,50,51]. However, the use of incentives to encourage health behaviors is relatively new, and more research is needed in this area. Financial incentives have proven to improve medication adherence in certain populations; however, their long-term viability can be questioned due to the resources needed [13-19]. Although incentives can be critiqued on their superficial nature or short-term viability, they may be a powerful motivating factor in creating habit-based behavior, a proven successful key in improving medication adherence, and an intrinsic source of motivation [51,52]. In the case of Perx, the extrinsic nature of the incentives may create habit-based adherence behavior in addition to intrinsic motivation to improve health outcomes. Additionally, the Perx app uses lottery-based incentives rather than predictable rewards. These incentives can enhance health behavior based on regret aversion or the human tendency to place a significant cost on regret [20]. If users believe that missing a medication dose can prevent them from winning a reward, they are still likely to improve their adherence, even without a guaranteed instant reward [21].

Limitations

Although our analysis proved that the Perx digital therapeutic is an effective intervention in managing medication adherence, it does have some limitations. First, the number of app users with available data was limited, did not extend past 6 months for the majority of users, and did not include information on the users' clinical conditions. Due to this, we were unable to perform subanalyses based on patient age, gender, medication, or condition. Second, we could not establish baseline adherence rates before the intervention was implemented or evaluate a control group due to the retrospective nature of the study. Third, while we believe that our sample reflects an accurate sample of patients who would be likely to use a mobile app to manage medications, the users who downloaded the app may also have been likely to adhere to their medications without the app. Conversely, it could also be argued that patients who need adherence management support would be more likely to download the app. Finally, while we could measure the number of active days per patient, it was not possible to determine full user engagement of the intervention in this analysis to understand the extent to which the intervention was used by each user.

Strengths

One strength of our study is our measure of adherence, self-reporting with MDOT [35]. Similar to electronic methods such as the Medication Event Monitoring System (MEMS), MDOT enables objective measurement while simultaneously

providing timestamps to additionally measure timing adherence, which is an important component of the multidimensional medication-taking process [35]. Additionally, our analysis of the gamification of mobile apps to maintain positive health behaviors is part of a new and emerging research landscape within the pharmacy and health care sector that has not been previously examined [53]. With the increasing number of health apps entering the market, supportive evidence is necessary to demonstrate the effectiveness of these tools and to indicate whether they should be recommended by health care professionals as a component of medication therapy [54,55]. Finally, our use of real-world data generated from users of this commercially available mobile app was a strength in that the data can be applied to a broader population of patients and reflect actual use in practice [32].

Future Work

A 12-month clinical trial is currently being conducted with the objective of assessing the efficacy of the Perx intervention in adherence and clinical outcomes. Future research should aim to assess the effectiveness of this intervention in improving adherence to medications and other gamification- or incentive-based strategies in addition to observing the impact on clinical health outcomes. Furthermore, a longer analysis period of 12 to 24 months would be beneficial in observing the long-term effects to determine if these types of interventions can sustain gold-standard adherence rates above 80% for longer than 6 months. It would additionally be useful to analyze the impact of the intervention across different points in the medication-taking process, such as initiation of medication, implementation and persistence adherence, and time to discontinuation of medication [1]. Finally, the opinions of stakeholders, specifically users, regarding the app and intervention components are vital to understand the main motivating factor in promoting adherence. A full engagement analysis identifying the components of the app on which the most time is spent as well as a user survey analysis are required to obtain a complete understanding of the success of the intervention.

Conclusion

Retrospective analysis of a digital therapeutic mobile app that merges gamification, education, reminders, a social community, and incentive-based components indicates that this intervention is successful in maintaining optimal medication adherence over time. Extrinsic external monetary motivators combined with fundamental game mechanics and other common behavioral change components may be a key force to promote intrinsic motivation and habit-based behavior, which can spark long-term changes in health behavior. Future research should evaluate the long-term impact of mobile apps using these components over a longer time period using experimental designs.

Acknowledgments

We would like to thank Lucky Health Pty Ltd for providing the data. This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the most important features of the Perx app.
[PNG File , 1021 KB - jmir_v22i8e17834_app1.png]

References

1. De Geest S, Zullig LL, Dunbar-Jacob J, Helmy R, Hughes DA, Wilson IB, et al. ESPACOMP Medication Adherence Reporting Guideline (EMERGE). Ann Intern Med 2018 Jul 03;169(1):30-35. [doi: 10.7326/M18-0543] [Medline: 29946690]
2. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014 Dec 20(11):CD000011 [FREE Full text] [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
3. Kardas P, Lewek P, Matyjaszczyk M. Determinants of patient adherence: a review of systematic reviews. Front Pharmacol 2013;4:91 [FREE Full text] [doi: 10.3389/fphar.2013.00091] [Medline: 23898295]
4. Wiecek E, Tonin FS, Torres-Robles A, Benrimoj SI, Fernandez-Llimos F, Garcia-Cardenas V. Temporal effectiveness of interventions to improve medication adherence: A network meta-analysis. PLoS One 2019;14(3):e0213432 [FREE Full text] [doi: 10.1371/journal.pone.0213432] [Medline: 30861014]
5. Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc 2011 May;86(4):304-314 [FREE Full text] [doi: 10.4065/mcp.2010.0575] [Medline: 21389250]
6. Yang K, Lee Y, Chasens ER. Outcomes of health care providers' recommendations for healthy lifestyle among U.S. adults with prediabetes. Metab Syndr Relat Disord 2011 Jul;9(3):231-237. [doi: 10.1089/met.2010.0112] [Medline: 21361822]
7. Okosun IS, Lyn R. Prediabetes awareness, healthcare provider's advice, and lifestyle changes in American adults. Int J Diabetes Mellit 2015 May;3(1):11-18 [FREE Full text] [Medline: 32595044]

8. Ruge T, Nyström L, Lindahl B, Hallmans G, Norberg M, Weinehall L, et al. Recruiting high-risk individuals to a diabetes prevention program: how hard can it be? Diabetes Care 2007 Jul;30(7):e61. [doi: 10.2337/dc06-2466] [Medline: 17596484]
9. Vlaev I, King D, Darzi A, Dolan P. Changing health behaviors using financial incentives: a review from behavioral economics. BMC Public Health 2019 Aug 07;19(1):1059 [FREE Full text] [doi: 10.1186/s12889-019-7407-8] [Medline: 31391010]
10. O'Donoghue T, Rabin M. Doing It Now or Later. American Economic Review 1999 Mar 01;89(1):103-124. [doi: 10.1257/aer.89.1.103] [Medline: 22489612]
11. Kullgren JT, Hafez D, Fedewa A, Heisler M. A Scoping Review of Behavioral Economic Interventions for Prevention and Treatment of Type 2 Diabetes Mellitus. Curr Diab Rep 2017 Oct;17(9):73 [FREE Full text] [doi: 10.1007/s11892-017-0894-z] [Medline: 28755061]
12. Arno A, Thomas S. The efficacy of nudge theory strategies in influencing adult dietary behaviour: a systematic review and meta-analysis. BMC Public Health 2016 Jul 30;16:676 [FREE Full text] [doi: 10.1186/s12889-016-3272-x] [Medline: 27475752]
13. Volpp KG, Loewenstein G, Troxel AB, Doshi J, Price M, Laskin M, et al. A test of financial incentives to improve warfarin adherence. BMC Health Serv Res 2008 Dec 23;8:272 [FREE Full text] [doi: 10.1186/1472-6963-8-272] [Medline: 19102784]
14. Mooney M, Babb D, Jensen J, Hatsukami D. Interventions to increase use of nicotine gum: a randomized, controlled, single-blind trial. Nicotine Tob Res 2005 Aug;7(4):565-579. [doi: 10.1080/14622200500185637] [Medline: 16085528]
15. Rigsby MO, Rosen MI, Beauvais JE, Cramer JA, Rainey PM, O'Malley SS, et al. Cue-dose training with monetary reinforcement: pilot study of an antiretroviral adherence intervention. J Gen Intern Med 2000 Dec;15(12):841-847 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.00127.x] [Medline: 11119180]
16. Nunes EV, Rothenberg JL, Sullivan MA, Carpenter KM, Kleber HD. Behavioral therapy to augment oral naltrexone for opioid dependence: a ceiling on effectiveness? Am J Drug Alcohol Abuse 2006;32(4):503-517. [doi: 10.1080/00952990600918973] [Medline: 17127538]
17. Choudhry NK, Avorn J, Glynn RJ, Antman EM, Schneeweiss S, Toscano M, et al. Full coverage for preventive medications after myocardial infarction. N Engl J Med 2011 Dec 01;365(22):2088-2097. [doi: 10.1056/NEJMsa1107913] [Medline: 22080794]
18. Malotte CK, Hollingshead JR, Larro M. Incentives vs outreach workers for latent tuberculosis treatment in drug users. Am J Prev Med 2001 Mar;20(2):103-107. [doi: 10.1016/s0749-3797(00)00283-x] [Medline: 11165450]
19. Morisky DE, Malotte CK, Choi P, Davidson P, Rigler S, Sugland B, et al. A patient education program to improve adherence rates with antituberculosis drug regimens. Health Educ Q 1990;17(3):253-267. [doi: 10.1177/109019819001700303] [Medline: 2228629]
20. Sen AP, Sewell TB, Riley EB, Stearman B, Bellamy SL, Hu MF, et al. Financial incentives for home-based health monitoring: a randomized controlled trial. J Gen Intern Med 2014 May;29(5):770-777 [FREE Full text] [doi: 10.1007/s11606-014-2778-0] [Medline: 24522623]
21. Humphrey CH, Small DS, Jensen ST, Volpp KG, Asch DA, Zhu J, et al. Modeling lottery incentives for daily adherence. Stat Med 2019 Jul 10;38(15):2847-2867 [FREE Full text] [doi: 10.1002/sim.8149] [Medline: 30941805]
22. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
23. Stieglitz S, Lattemann C, Robra-Bissantz S, Zarnekow R, Brockmann T. Introduction to Gamification: Foundation and Underlying Theories. In: Gamification - Using Game Elements in Serious Context. Cham, Switzerland: Springer International Publishing; 2017:3-18.
24. Aitken M, Clancy B, Nass D. The Growing Value of Digital Health.: IQVIA Institute; 2017. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health [accessed 2019-11-10]
25. Brand J, Todhunter S, Jervis J. Digital Australia Report 2018. Eveleigh, NSW: Interactive Games & Entertainment Association; 2018. URL: https://www.igea.net/wp-content/uploads/2017/07/Digital-Australia-2018-DA18-Final-1.pdf
26. Zhao J, Freeman B, Li M. Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
27. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Mar 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
28. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, et al. Effect of a Game-Based Intervention Designed to Enhance Social Incentives to Increase Physical Activity Among Families: The BE FIT Randomized Clinical Trial. JAMA Intern Med 2017 Nov 01;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]
29. Ahmed I, Ahmad NS, Ali S, Ali S, George A, Saleem Danish H, et al. Medication Adherence Apps: Review and Content Analysis. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e62 [FREE Full text] [doi: 10.2196/mhealth.6432] [Medline: 29549075]
30. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]

31. Hightow-Weidman L, Muessig K, Knudtson K, Srivatsa M, Lawrence E, LeGrand S, et al. A Gamified Smartphone App to Support Engagement in Care and Medication Adherence for HIV-Positive Young Men Who Have Sex With Men (AllyQuest): Development and Pilot Study. JMIR Public Health Surveill 2018 May 30;4(2):e34 [FREE Full text] [doi: 10.2196/publichealth.8923] [Medline: 29712626]
32. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence - What Is It and What Can It Tell Us? N Engl J Med 2016 Dec 08;375(23):2293-2297. [doi: 10.1056/NEJMsb1609216] [Medline: 27959688]
33. Barnish MS, Turner S. The value of pragmatic and observational studies in health care and public health. Pragmat Obs Res 2017;8:49-55 [FREE Full text] [doi: 10.2147/POR.S137701] [Medline: 28546780]
34. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg 2014 Dec;12(12):1495-1499 [FREE Full text] [doi: 10.1016/j.ijsu.2014.07.013] [Medline: 25046131]
35. Shields MD, ALQahtani F, Rivey MP, McElnay JC. Mobile direct observation of therapy (MDOT) - A rapid systematic review and pilot study in children with asthma. PLoS One 2018 Feb 5;13(2):e0190031 [FREE Full text] [doi: 10.1371/journal.pone.0190031] [Medline: 29401500]
36. Baumgartner PC, Haynes RB, Hersberger KE, Arnet I. A Systematic Review of Medication Adherence Thresholds Dependent of Clinical Outcomes. Front Pharmacol 2018;9:1290 [FREE Full text] [doi: 10.3389/fphar.2018.01290] [Medline: 30524276]
37. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 04;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
38. Torres-Robles A, Wiecek E, Cutler R, Drake B, Benrimoj SI, Fernandez-Llimos F, et al. Using Dispensing Data to Evaluate Adherence Implementation Rates in Community Pharmacy. Front Pharmacol 2019;10:130 [FREE Full text] [doi: 10.3389/fphar.2019.00130] [Medline: 30863308]
39. Burnier M. Is There a Threshold for Medication Adherence? Lessons Learnt From Electronic Monitoring of Drug Adherence. Front Pharmacol 2018;9:1540 [FREE Full text] [doi: 10.3389/fphar.2018.01540] [Medline: 30687099]
40. Demonceau J, Ruppar T, Kristanto P, Hughes DA, Fargher E, Kardas P, ABC Project Team. Identification and assessment of adherence-enhancing interventions in studies assessing medication adherence through electronically compiled drug dosing histories: a systematic literature review and meta-analysis. Drugs 2013 May;73(6):545-562 [FREE Full text] [doi: 10.1007/s40265-013-0041-3] [Medline: 23588595]
41. Choudhry NK, Krumme AA, Ercole PM, Girdish C, Tong AY, Khan NF, et al. Effect of Reminder Devices on Medication Adherence: The REMIND Randomized Clinical Trial. JAMA Intern Med 2017 May 01;177(5):624-631 [FREE Full text] [doi: 10.1001/jamainternmed.2016.9627] [Medline: 28241271]
42. Vervloet M, Linn AJ, van Weert JCM, de Bakker DH, Bouvy ML, van Dijk L. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature. J Am Med Inform Assoc 2012;19(5):696-704 [FREE Full text] [doi: 10.1136/amiajnl-2011-000748] [Medline: 22534082]
43. Sapkota S, Brien JE, Greenfield JR, Aslani P. A Systematic Review of Interventions Addressing Adherence to Anti-Diabetic Medications in Patients with Type 2 Diabetes--Components of Interventions. PLoS One 2015;10(6):e0128581 [FREE Full text] [doi: 10.1371/journal.pone.0128581] [Medline: 26053004]
44. Kini V, Ho PM. Interventions to Improve Medication Adherence: A Review. JAMA 2018 Dec 18;320(23):2461-2473. [doi: 10.1001/jama.2018.19271] [Medline: 30561486]
45. Wang Y, Sloan FA. Present bias and health. J Risk Uncertain 2018 Oct;57(2):177-198 [FREE Full text] [doi: 10.1007/s11166-018-9289-z] [Medline: 31244508]
46. van Roy R, Zaman B. Why Gamification Fails in Education and How to Make It Successful: Introducing Nine Gamification Heuristics Based on Self-Determination Theory. In: Serious Games and Edutainment Applications Volume II. Cham, Switzerland: Springer International Publishing; 2017:485-509.
47. Deci EL, Ryan RM. Facilitating optimal motivation and psychological well-being across life's domains. Canadian Psychology/Psychologie canadienne 2008 Feb;49(1):14-23. [doi: 10.1037/0708-5591.49.1.14]
48. Mekler ED, Brühlmann F, Tuch AN, Opwis K. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. Comput Hum Behav 2017 Jun;71:525-534 [FREE Full text]
49. van Heuckelum M, van den Ende CHM, Houterman AEJ, Heemskerk CPM, van Dulmen S, van den Bemt BJF. The effect of electronic monitoring feedback on medication adherence and clinical outcomes: A systematic review. PLoS One 2017;12(10):e0185453 [FREE Full text] [doi: 10.1371/journal.pone.0185453] [Medline: 28991903]
50. Burgess S, Ratto M. The Role of Incentives in the Public Sector: Issues and Evidence. Oxf Rev Econ Policy 2003 Jun 01;19(2):285-300 [FREE Full text] [doi: 10.1093/oxrep/19.2.285]
51. Kreps DM. Intrinsic Motivation and Extrinsic Incentives. Am. Econ. Rev 1997 May;87(2):359-364 [FREE Full text] [doi: 10.1257/aer.98.2.201]
52. Conn VS, Ruppar TM. Medication adherence outcomes of 771 intervention trials: Systematic review and meta-analysis. Prev Med 2017 Jul;99:269-276 [FREE Full text] [doi: 10.1016/j.ypmed.2017.03.008] [Medline: 28315760]
53. Morrissey EC, Corbett TK, Walsh JC, Molloy GJ. Behavior Change Techniques in Apps for Medication Adherence: A Content Analysis. Am J Prev Med 2016 May;50(5):e143-e146. [doi: 10.1016/j.amepre.2015.09.034] [Medline: 26597504]

54. Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: potential benefits to patients and providers: response to Aungst. J Am Pharm Assoc (2003) 2013;53(4):345. [doi: 10.1331/JAPhA.2013.13121] [Medline: 23892801]
55. Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: potential benefits to patients and providers. J Am Pharm Assoc (2003) 2013;53(2):172-181 [FREE Full text] [doi: 10.1331/JAPhA.2013.12202] [Medline: 23571625]

Abbreviations

EMERGE: ESPACOMP Medication Adherence Reporting Guideline
MDOT: mobile direct observation of therapy
MEMS: Medication Event Monitoring System
mHealth: mobile health
STROBE: STrengthening the Reporting of OBservational Studies in Epidemiology

Edited by G Eysenbach; submitted 16.01.20; peer-reviewed by R Bartlett, K Knudtson; comments to author 02.03.20; revised version received 15.05.20; accepted 03.06.20; published 12.08.20.

Please cite as:
Wiecek E, Torres-Robles A, Cutler RL, Benrimoj SI, Garcia-Cardenas V
Impact of a Multicomponent Digital Therapeutic Mobile App on Medication Adherence in Patients with Chronic Conditions: Retrospective Analysis
J Med Internet Res 2020;22(8):e17834
URL: http://www.jmir.org/2020/8/e17834/
doi:10.2196/17834
PMID:32784183

Original Paper

A Novel Multimodal Digital Service (Moderated Online Social Therapy+) for Help-Seeking Young People Experiencing Mental Ill-Health: Pilot Evaluation Within a National Youth E-Mental Health Service

Mario Alvarez-Jimenez[1,2], BSc(Hons), DClinPsy, MAResearchMeth, PhD; Simon Rice[1,2], BBSc, DipEd, BSc (Hons), GradDipClinEpi, MPsych (Clin), PhD; Simon D'Alfonso[1,3], PhD; Steven Leicester[4], BEd, MPsych, PhD; Sarah Bendall[1,2], BA, PGDipClinPsych, MA, PhD; Ingrid Pryor[1,2], BA (Hons), MClinPsych (Clin); Penni Russon[1,5], BA (Hons); Carla McEnery[1,2,6], BPsychSci (Hons), PhD; Olga Santesteban-Echarri[7], MSc (ClinPsych), PhD; Gustavo Da Costa[3], BEng (Elect & Elect) (Hons), MEng (Elect & Elect) (Distinct), MIT; Tamsyn Gilbertson[1], BA (Hons), MSc, MPsych (Clin); Lee Valentine[1,2], BA, MSocWk; Laia Solves[8], BSc; Aswin Ratheesh[1,2], MD Psychiatry, PhD; Patrick D McGorry[1,2], MD Psychiatry, PhD; John Gleeson[9,10], BA (Hons), MPsych (Clin), PhD

[1]Orygen, Parkville, Australia

[2]Centre for Youth Mental Health, The University of Melbourne, Melbourne, Australia

[3]The School of Computing and Information Systems, The University of Melbourne, Melbourne, Australia

[4]headspace National Youth Mental Health Foundation, Melbourne, Australia

[5]School of Culture and Communications, The University of Melbourne, Melbourne, Australia

[6]Centre for Mental Health, Swinburne University of Technology, Hawthorn, Australia

[7]Hotchkiss Brain Institute, Psychiatry Department, University of Calgary, Calgary, AB, Canada

[8]Consorci Sanitari del Maresme, Department of Psychiatry, Hospital of Mataro, Mataro, Spain

[9]School of Behavioural and Health Sciences, Australian Catholic University, Melbourne, Australia

[10]Healthy Brain and Mind Research Centre, Australian Catholic University, Melbourne, Australia

Corresponding Author:
Mario Alvarez-Jimenez, BSc(Hons), DClinPsy, MAResearchMeth, PhD
Orygen
35 Poplar Rd
Parkville, 3052
Australia
Phone: 61 9966 9383
Email: mario.alvarez@orygen.org.au

Abstract

Background: Mental ill-health is the leading cause of disability worldwide. Moreover, 75% of mental health conditions emerge between the ages of 12 and 25 years. Unfortunately, due to lack of resources and limited engagement with services, a majority of young people affected by mental ill-health do not access evidence-based support. To address this gap, our team has developed a multimodal, scalable digital mental health service (Enhanced Moderated Online Social Therapy [MOST+]) merging real-time, clinician-delivered web chat counseling; interactive user-directed online therapy; expert and peer moderation; and peer-to-peer social networking.

Objective: The primary aim of this study is to ascertain the feasibility, acceptability, and safety of MOST+. The secondary aims are to assess pre-post changes in clinical, psychosocial, and well-being outcomes and to explore the correlations between system use, perceived helpfulness, and secondary outcome variables.

Methods: Overall, 157 young people seeking help from a national youth e-mental health service were recruited over 5 weeks. MOST+ was active for 9 weeks. All participants had access to interactive online therapy and integrated web chat counseling. Additional access to peer-to-peer social networking was granted to 73 participants (46.5%) for whom it was deemed safe. The intervention was evaluated via an uncontrolled single-group study.

Results: Overall, 93 participants completed the follow-up assessment. Most participants had moderate (52/157, 33%) to severe (96/157, 61%) mental health conditions. All a priori feasibility, acceptability, and safety criteria were met. Participants provided mean scores of ≥3.5 (out of 5) on ease of use (mean 3.7, SD 1.1), relevancy (mean 3.9, SD 1.0), helpfulness (mean 3.5, SD 0.9), and overall experience (mean 3.9, SD 0.8). Moreover, 98% (91/93) of participants reported a positive experience using MOST+, 82% (70/93) reported that using MOST+ helped them feel better, 86% (76/93) felt more socially connected using it, and 92% (86/93) said they would recommend it to others. No serious adverse events or inappropriate use were detected, and 97% (90/93) of participants reported feeling safe. There were statistically significant improvements in 8 of the 11 secondary outcomes assessed: psychological distress (d=−0.39; P<.001), perceived stress (d=−0.44; P<.001), psychological well-being (d=0.51; P<.001), depression (d=−0.29; P<.001), loneliness (d=−0.23; P=.04), social support (d=0.30; P<.001), autonomy (d=0.36; P=.001), and self-competence (d=0.30; P<.001). There were significant correlations between system use, perceived helpfulness, and a number of secondary outcome variables.

Conclusions: MOST+ is a feasible, acceptable, and safe online clinical service for young people with mental ill-health. The high level of perceived helpfulness, the significant improvements in secondary outcomes, and the correlations between indicators of system use and secondary outcome variables provide initial support for the therapeutic potential of MOST+. MOST+ is a promising and scalable platform to deliver standalone e-mental health services as well as enhance the growing international network of face-to-face youth mental health services.

(J Med Internet Res 2020;22(8):e17155) doi:10.2196/17155

KEYWORDS

mHealth; youth; social media; social networking; mobile phone; internet-based intervention

Introduction

Background

Mental ill-health is the number one cause of disability worldwide [1] and accounts for 8 million deaths annually [2]. Mental illness strikes the young, with 75% of all mental disorders having their onset before the age of 25 years [3], and for many, it follows a relapsing course [4,5]. The timing and course of mental illness disrupts the attainment of critical developmental milestones for young people, such as completing their education, entering the job market, and developing intimate relationships [6]. This can result in devastating lifelong consequences, including increased risk of chronic unemployment, lower income and living standards, homelessness, social isolation, suicide, and early mortality [7-9].

Despite the prevalence and impact of mental illness, between 35% and 57% of people with mental health disorders do not access treatment in high-income countries [1]. The corresponding range for low- and middle-income countries is 76% to 85% [1]. The mismatch between the prevalence of mental illness and the rate at which treatment is accessed is the greatest for young people. The most recent Australian national survey of young people's mental health revealed that only 13% of men and 31.2% of women aged 16-24 years who had experienced a mental disorder in the preceding 12 months received professional help [10]. The reasons for low rates of treatment access among young people include low help-seeking due to fear of stigma and embarrassment, confidentiality concerns, negative prior experiences of treatment, poor mental health literacy, lack of knowledge of available resources, a preference for self-reliance, and social isolation [11-13]. Additional structural and logistical barriers include geographical distance, poverty, out-of-pocket expenses associated with treatment, and lack of availability of services [14].

The internet, mobile technologies, and social media have the potential to address the global crisis in the rate at which young people access evidence-based mental health care. Internet-enabled mobile devices are a pervasive element of young people's lives, with 45% of adolescents being on the web almost constantly [15]. Social media has become a key vehicle for young people to communicate with one another. Almost all young people have at least one active social media account, with over 70% using social media multiple times a day—a rate that has doubled between 2012 and 2018 [16]. Recent surveys demonstrate that young people experiencing mental ill-health are also avid users of social media: 97% use it regularly, 2.6 hours per day on average [17]. Particularly relevant to the clinical potential of social media, engaging with peers online about mental health issues increases the likelihood of seeking professional support [18], and many young people use social media to obtain mental health information [19]. Similarly, initial studies showed that 74% of young people who experience mental ill-health would like to obtain help from mental health clinicians via social media [17] and value web-based services run by professionals [20]. Thus, social media offers a unique opportunity to provide and boost web-based youth mental health interventions.

Currently, there are 4 main types of digital interventions for mental health: self-guided web-based interventions, standalone mental health mobile apps, online peer support groups or interventions, and web-based counseling with registered professionals. Previous trials have shown that the first generation of self-guided web-based interventions, particularly those delivering cognitive behavioral therapy (CBT) and including human support [21], are effective in alleviating anxiety and depressive symptoms [22,23]. However, the impact of these web-based interventions has been hindered by two constraints. First, these interventions were designed to *mimic* face-to-face interventions, resulting in high attrition rates (particularly for interventions with no human support) and little treatment

innovation (eg, novel intervention models harnessing mobile technology to provide intensive, real-time support) [24]. Second, these interventions were developed as *separate products* [24], with little consideration as to how they fit in with existing clinical services, resulting in an overall lack of integration of web-based support with clinical services [25].

The number of mobile apps targeting mental health has grown exponentially over the past few years. According to a 2017 report, almost 500 unique apps were targeting mental health disorders in 2017 [26]. More recently, a systematic search of mobile apps focused on wellness and stress management found over 1000 publicly available apps [27]. However, although mental health apps are flooding the consumer market, very few studies have examined their effectiveness [28], and many apps do not follow evidence-based guidelines and principles [29]. For example, of the apps focused on psychosocial wellness and stress management, only 1% involved therapist support, and less than 2% were designed to supplement in-person treatment and 2% had any research publications (with most of these being a single feasibility or efficacy study) [27]. Examining the existing evidence, two meta-analyses found significant but small effects for reductions in anxiety [30] and depression [31] from smartphone interventions as compared with control conditions. However, a more recent meta-analysis including only mobile apps designed to treat mental health symptoms found small significant effects on depression, with no significant effects on anxiety, suicidal ideation, self-injury, or alcohol use [32]. It must be noted that the effects of mental health apps (g=−0.14 to 0.18) were significantly smaller than those of web-based mental health interventions [23,33-35] (g=0.84-1.09). These smaller effects may be explained by the lack of human support in most mental health apps; poor adherence, which is considered to be a major pitfall of mental health apps [36], or low use of key app features; and possibly a lack of sustained and structured attention to content and features that may be needed for an intervention to yield significant benefits [32]. To date, most mobile mental health apps have been designed in academic labs or by commercial companies, usually outside of clinical services, resulting in an overall lack of integration with existing clinical services [37].

A growing number of online peer support groups and social networking sites (SNS) exist for people with mental health problems. Overall, the extant evidence suggests that online peer support groups can foster a sense of social connectedness, empowerment, and improved quality of life as well as reduce depression and emotional distress [38-40]. That said, the type and function of online peer support groups and SNS appear to have a significant effect on their outcomes. For example, unmoderated forums and SNS can lead to increased contagion, distress, and collusion among users [39]. Conversely, SNS interventions that have been moderated, ideally by professionals, have been found to be safe, engaging, supportive, and useful [39,41-44]. This is in keeping with the findings that young people with mental ill-health have a strong preference for purpose-built, moderated, social media–based interventions enabling peer-to-peer contact as well as clinician support [45]. Despite this emerging evidence, the majority of available peer support groups do not provide professional moderation and clinical support or incorporate evidence-based, user-directed web-based therapy.

The third main type of digital support is web-based counseling (ie, real-time web chat with clinicians). There is initial evidence that web chat is an effective way to deliver mental health support [46], with recent trials showing comparable levels of effectiveness [47], therapeutic alliance [48], and perceived helpfulness with face-to-face therapy [48]. Recent studies also indicate that young people are increasingly turning to web chat to receive mental health support [49]. However, these web-based services are also constrained by capacity and scalability due to their reliance on one-to-one web-based support.

Our group has developed a novel and evolving model of web-based behavioral interventions entitled Moderated Online Social Therapy (MOST). The MOST model merges (1) interactive web therapy, (2) peer-to-peer web-based social networking, (3) peer, and (4) clinical moderation. Successive iterations and evolutions of MOST have been successfully adapted for, and trialed with, young people with psychosis [42,50], at clinical risk of psychosis [43], suicidal risk [51], depression [44,52], and social anxiety [53], as well as relatives of young people with psychosis [54] and depression [55]. To respond to the crisis in access to care by young people who experience mental ill-health, our group has developed a new model of web-based clinical support entitled Enhanced Moderated Online Social Therapy (MOST+). MOST+ fully merges the effective components of web-based interventions, online peer support groups, and web-based counseling into a single platform. As such, MOST+ combines (1) a wide range of evidence-based, interactive, user-directed web-based interventions; (2) secure and supportive peer-to-peer web-based social networking; (3) peer moderator support; (4) clinical moderation; and (5) on demand web chat with registered clinicians. Thus, MOST+ was designed to deliver an accessible and scalable web-based mental health service catering to the varying needs and preferences of young people by flexibly integrating multiple modes of effective web-based support.

Objectives

The overarching aim of this study (trial registration: ACTRN12617000370303) was to determine the feasibility, acceptability, and safety of MOST+ for young people seeking online mental health support. The secondary aims of the project were (1) to assess changes in psychological distress, well-being, depression, stress, social support, loneliness, basic psychological needs (self-competence, relatedness, and autonomy), strengths usage, and mindfulness skills from the point of engagement to post intervention and (2) to explore the associations between system usage, perceived helpfulness, and secondary outcome variables. We hypothesized that MOST+ would be regularly used, favorably received, and safe against a priori established criteria (described in detail in the *Results* section) [56].

Methods

Study Design and Setting

The methods of this study have been described in detail elsewhere [56]. Briefly, this study employed an uncontrolled

single-group design [57]. The MOST+ intervention was embedded within eheadspace services, a national web counseling service funded by the federal government for young people aged 12 to 25 years in Australia [49]. eheadspace is staffed by qualified and supervised mental health clinicians providing synchronous web chat, email support, and telephone-based mental health support. For this study, 27 eheadspace clinicians were trained to deliver online support using the MOST+ platform. Two of these clinicians, based at the eheadspace operation center, provided web chat support via MOST+ from 4 PM to 12 AM every day. They were registered with mental health clinicians with prior specialist training and experience in the delivery of e-mental health support to young people in distress. The MOST+ platform was available for a period of 9 weeks, including a 5-week recruitment period (ACTRN12617000370303).

Participants and Recruitment

The sample included 157 young people recruited via an opt-in process at the point of entry to eheadspace through a link on the home page. The inclusion criteria were as follows: (1) help-seeking young people with concerns about their own mental health, (2) people aged 16 to 25 years, and (3) people with the ability to provide informed consent and comply with study procedures. The participants who indicated on the web that they understood and consented to the study procedures were recruited into the study.

To ensure the safety of the online social network, some participants were excluded from access to the social networking component of MOST+ (defined as partial access; see under *Web Counseling*). These exclusions were (1) acute risk of self-harm requiring urgent intervention (ie, suicidal ideation with a current plan and intent) indicated by a young person on web chat or by endorsing predetermined screening questions; (2) participation in the MOST+ social network was deemed likely to interfere with appropriate clinical management of mental health symptoms (eg, psychosis) or increase the risk of harm to self or others by an eheadspace clinician; and (3) inability to confirm age and conduct induction via a research assistant telephone contact (Multimedia Appendix 1). Participants with partial access could access all other components of MOST+ (ie, web chat and interactive web-based therapy). This design feature enabled us to compare outcomes across these 2 levels of access.

The mean age of the participants was 19.1 (SD 2.3) years, with 77% female participants. A total of 87% (137/157) of the participants were born in Australia and 11% (17/157) spoke languages other than English. Moreover, 70% (110/157) of the participants were from metropolitan areas, 28% (44/157) from rural areas, and 2% (3/157) from remote areas. In addition, 3% (5/157) of the participants identified themselves as Aboriginals and/or Torres Strait Islanders. Furthermore, 59% (93/157) of the participants had not previously used youth mental health services and 37% (58/157) had never received any mental health support. A total of 57% (89/157) of the participants were engaged in paid work and 77% (121/157) were studying part-time or full-time. The main reasons for seeking help included sadness (38%: 60/157) and anxiety (22%:35/157), followed by feelings of distress (9.6%:15/157). Baseline clinical measures indicated that the majority of participants had mental ill-health. Specifically, the mean baseline Kessler 10 (K10) score was 32.03 (SD 7.72), with 61% (57/93) scoring 30 (indicative of a severe mental health disorder) and 33% (31/93) scoring 25-29 (indicative of a moderate mental health disorder) [58] (Table 1). Similarly, the mean baseline Patient Health Questionnaire-9 (PHQ-9) score (measuring depression) was 15.76 (SD 6.32), with 59% (55/93) scoring 15 (indicative of moderately severe depression) and 24% (22/93) scoring 10-14 (indicative of moderate depression) [59] (Table 1 [60-71]).

Table 1. Mean (SD) and within-group effect sizes (Cohen d) for outcome measures (N=93).

Characteristics	Baseline, mean (SD)	Follow-up, mean (SD)	*P* values	Cohen *d* (95% CI)
K10[a]	32.03 (7.680)	29.43 (8.119)	<.001	−0.39 (−0.68 to −0.10)
WEMWS[b]	6.58 (2.174)	7.60 (2.232)	<.001	0.51 (0.21 to 0.80)
PSS[c]	10.65 (2.483)	9.52 (2.940)	<.001	−0.44 (−0.72 to −0.14)
PHQ-9[d]	15.76 (6.322)	13.98 (6.514)	.008	−0.29 (−0.57 to −0.01)
UCLA[e]	9.23 (1.984)	8.83 (2.224)	.04	−0.23 (−0.52 to −0.06)
Competence[f]	20.69 (6.449)	22.27 (6.494)	.005	0.30 (0.01 to 0.60)
Relatedness[g]	35.61 (8.900)	36.85 (7.412)	.08	0.17 (−0.12 to 0.46)
Autonomy[h]	25.68 (6.663)	27.61 (7.148)	.001	0.36 (0.07 to 0.65)
FS[i]	10.06 (5.303)	11.28 (4.935)	.004	0.30 (0.01 to 0.59)
SUS[j]	54.20 (16.621)	56.40 (17.361)	.21	0.13 (−0.15 to 0.42)
FMI[k]	28.77 (6.513)	30.08 (7.184)	.08	0.20 (−0.10 to 0.48)

[a]K10: Kessler 10.

[b]WEMWS: 3 items from the Warwick-Edinburgh Mental Well-being Scale.

[c]PSS: Perceived Stress Scale.

[d]PHQ-9: Patient Health Questionnaire-9.

[e]UCLA: UCLA Loneliness Scale (Version 3).

[f]Competence: subscale of the Basic Psychological Need Satisfaction Scale.

[g]Relatedness: subscale of the Basic Psychological Need Satisfaction Scale.

[h]Autonomy: subscale of the Basic Psychological Need Satisfaction Scale.

[i]FS: Friendship Scale.

[j]SUS: Strengths Use Scale.

[k]FMI: Freiburg Mindfulness Inventory-Short Form.

Intervention: MOST+

A large multidisciplinary team of researchers, clinical psychologists, programmers, creative writers, graphic artists, and experts in human-computer interaction worked in collaboration with end users to iteratively develop the MOST+ platform [42,45]. MOST+ merged (1) interactive user-directed web-based therapy (*Steps*), (2) peer-to-peer online social networking, (3) peer moderator support, (4) expert moderation, and (5) on demand web chat with clinicians.

MOST+ was conceived as an accessible web-based youth mental health service delivering immediate, short-term, flexible, and evidence-based support to help-seeking young people with mental ill-health. MOST+ was designed to be scalable through the integration of multiple modes of web-based support, thus enabling varying levels of direct support by peer moderators and clinicians.

MOST+ Intervention Components

Interactive, User-Directed Psychosocial Interventions

MOST+ adopted a strengths-based approach [42] through which users were guided and prompted to identify (via an interactive card-sort game) and exercise their core personal strengths to foster psychological well-being, including social connectedness and self-efficacy. Each strength included behavioral prompts or *do its* designed to support young people in applying their core strengths (eg, courage) for specific purposes (eg, dealing with social anxiety).

Psychosocial interventions in MOST+ took the form of brief web-based comics called *steps*. Comics have been used in physical health interventions as a nonthreatening, easy-to-understand medium for patient education [72]. We chose to use comics in MOST+ because of their ability to merge persuasive metaphors and character-driven narratives [73], potentially making therapeutic concepts more accessible, engaging, and compelling for young people. Comics were developed by clinical psychologists, professional novelists, and comic artists in partnership with young people (Figure 1). This process included focus groups with young people to identify salient therapeutic themes as well as continual feedback and co-design sessions involving artists, novelists, psychologists, and young people across all phases of comic development (from initial scripting to comic drawing and coloring) [74,75]. MOST+ included 52 discrete, interactive therapy comics designed to address the main concerns for helping young people to access web counseling at eheadspace, including managing immediate distress, low mood, anxiety, social anxiety, relationship issues, and difficulties at school or work [49]. The comics were informed by evidence-based psychological therapies including CBT [76-78], mindfulness [79,80], self-compassion [81], and positive psychology interventions [82]. The comics were

designed to be race, age, and gender neutral, with the purpose of maximizing their cultural, sexuality, and gender acceptance. The steps further incorporated *do its* to help participants exercise therapeutic skills (eg, empathy) in a real-life context (eg, school). Young people were able to store any *do it* they wanted to complete in the future in a *playlist*. Finally, the participants had the ability to rate, like, comment on, and share any step or *do it* with others via the social networking newsfeed.

Figure 1. Extract of a mindfulness online comic.

Web-Based Social Networking

Participants with full access were able to communicate with one another and the peer moderators in the *Café* and *Talk it Out* sections to foster social support and problem-solving skills. The Café included a newsfeed where participants and moderators were able to create posts to share thoughts, information, pictures, and videos and respond to other users' posts by commenting, *liking*, or *reacting* to content. The *reactions* were designed to facilitate social support (eg, *I get you* and *thinking of you*). Each participant could personalize their own profile with images and a bio and could visit the *wall* of fellow users, where their posts and general activities were displayed.

Talk it Out was designed as a collaborative problem-solving feature informed by the evidence-based problem-solving framework [83,84]. Participants could suggest topics (eg, how to make new friends at a new school in year 11?) and discuss solutions with trained peer moderators and other young people. Peer moderators encouraged young people to define the problem, brainstorm possible solutions, identify the pros and cons, and summarize possible solutions. Previous problems and group solutions were stored in the system, providing an easily accessible solution wiki for all young people.

Web Counseling

The web chat was fully integrated within MOST+. Young people using the system could request access to a clinician-delivered web chat between 4 PM and 12 AM. This included real-time web counseling focused on reducing immediate distress, supporting positive self-care, and facilitating referral to additional support where appropriate. Following a web chat session and based on the context of the consultation, MOST+ clinicians suggested specific, relevant content from MOST+ (eg, web-based comic, *Talk it Out*, or *do it*) to the young person. These suggestions appeared in the users' notifications sections as well as in the chat window. Web chats were classified based on the level of risk, with chat requests indicating that suicidal ideation and psychotic symptoms were prioritized.

Partial Versus Full Access to MOST+

Participants who consented to the study completed a 15-min web-based survey [56]. Following this survey, all young people were granted partial access to MOST+ consisting of real-time, clinician-delivered web chat as well as user-directed psychosocial interventions. Full access to MOST+, which additionally included peer-to-peer web-based social networking, was granted based on a three-tiered screening process designed to ascertain the safety and appropriateness of this component for each young person (Multimedia Appendix 1).

Duration of Access to MOST+

Irrespective of the level of access, all participants were enrolled in the MOST+ intervention for 1 week, with the option to extend their enrollment on a weekly basis over the duration of the intervention period (ie, a minimum of 1 week to a maximum of 9 weeks). The participants were shown a *count-down watch* depicting the time left before their account deactivation. Following account deactivation, all information and activity pertaining to the participant's account (eg, profile, posts) were hidden from the MOST+ system. However, the participants were able to reactivate an expired account at any time during the study intervention period. Multimedia Appendix 2 provides examples of the participants' enrollment timelines, with each colored line representing a different potential timeline over the duration of the pilot. This feature was implemented to protect the privacy of the large group of one-time users of eheadspace services [49] while ensuring a lively and safe online social network for regular users. Specifically, based on the patterns of usage of eheadspace, we anticipated that a significant proportion of users would only use MOST+ for a short period. By automatically hiding their activity following their short-term use of MOST+, we aimed to protect their privacy in the long term, while maintaining the currency and dynamism of the social network and facilitating the safety management of the social network at scale.

Moderation

MOST+ incorporated clinician as well as peer moderation. Clinical moderation primarily focused on ensuring the safety

of the social network. Specifically, MOST+ clinicians monitored new contributions to the network for indicators of clinical risk. The social network was moderated by an on-duty MOST+ clinician daily. Safety checks (ie, monitoring any indications of risk on the social network) were undertaken a minimum of 2 times per week day and once daily on weekends and public holidays (see *Safety Protocol* details below). The clinicians were regularly supervised throughout the trial by their usual eheadspace shift supervisors.

The *Cafe* was led and moderated by trained young people with the experience of having lived with mental illness (*Peer workers*). Peer workers were peer moderators who facilitated social learning using MOST+ in desired ways (eg, self-disclosing, using therapy content to deal with difficulties) [85]. Drawing on the growing evidence for peer support having a positive impact on the levels of hope, empowerment, and quality of life [86], peer workers also provided general guidance and peer-to-peer support. In addition, peer workers guided the problem-solving discussions in *Talk it Out* and posted links to relevant therapeutic resources and *tips*. Finally, peer workers seeded discussion threads and *icebreakers* to enable useful, enjoyable conversations and facilitate meaningful relationships. Peer moderators were supervised weekly by members of the research team (clinical psychologists and peer support coordinators).

Safety Protocol

The safety protocol comprised 3 levels of security: (1) system and privacy protection, (2) web safety, and (3) clinical safety. MOST+ had built-in security and data protection to prevent unauthorized access to the platform, which has been described elsewhere [56]. These measures conform to industry best practices as defined by the Open Web Application Security Project. Privacy and web safety were managed in accordance with the Australian Communications and Media Authority. Specifically, the participants were informed of, and were required to accept, the terms of use of MOST+, which included clauses on protecting their privacy and that of others as well as guidelines on proscribed behavior (ie, disrespectful behavior or offensive comments).

The MOST+ clinical safety protocol included manual and automated procedures. First, information related to clinical risk (posts or messages) was screened by clinical moderators twice each weekday and daily on weekends and public holidays. Second, MOST+ incorporated an automatic alert system that monitored self-harm–related terms posted on the social feed. Any detected increased risk or inappropriate use activated the safety protocol (Multimedia Appendix 3). Finally, a *report function* enabled the participants to alert the moderators to inappropriate use of the system (eg, discriminatory comments posted on the social network).

Participants' Data Management

The participants were able to control the extent to which they could be identified by other users within the social network, including whether they used their first name or a nickname and whether their profile picture included a photo. As noted above, following account deactivation, the participants' accounts and activities were hidden from MOST+. The participants could also choose to *switch off* their profiles to hide all past posts and comments and anonymize their contributions to *Talk it Out*. The participants were informed that any records of user activity hidden from the social network were retained by the researchers for the purpose of analysis. Specifically, all user-generated data were encrypted and retained in the MOST+ database throughout the trial. Upon completion of the trial, MOST+ was decommissioned and the database was exported and stored in a deidentified format on an Orygen server for research purposes.

Outcome Measures

The primary outcome variables were intervention feasibility, acceptability, and safety. All outcomes were assessed at baseline and at follow-up. Baseline assessments were conducted on the web as part of the onboarding process. Follow-up assessment occurred approximately 4 days after the initial account deactivation (ie, 4 days after a participant opted not to renew their account for an additional week). For those participants who maintained active enrollment across the intervention period, follow-up occurred as soon as possible following the conclusion of the pilot. The participants received a short message service notification indicating that their web follow-up survey was due and that they were able to complete survey items either via the web or telephone.

A self-report user feedback questionnaire was developed based on the user experience approach [87] assessing the following themes: (1) acceptability, (2) helpfulness, and (3) safety. Acceptability was determined against the following a priori indicators: (1) participants provided ratings of the MOST+ platform averaging above three out of five across feedback questions regarding ease of use, relevancy, helpfulness, and overall experience; (2) at least 60% (56/93) of the participants reported that the MOST+ intervention provided relevant and helpful support; and (3) at least 80% (74/93) of the participants would recommend MOST+ to other young people experiencing similar difficulties (Table 2). In addition, the MOST+ intervention was considered safe if (1) at least 90% (84/93) of the participants reported the web-based intervention to be safe, (2) none of the participants experienced a serious adverse event as a result of their engagement with the system, and (3) there were no unlawful entries into the MOST+ system detected by study programmers during the 8-week pilot.

Secondary outcome measures included self-report measures of psychological distress, well-being, depression, stress, social support, loneliness, basic psychological needs (self-competence, relatedness, and autonomy), strengths usage, and mindfulness skills (Table 3).

Table 2. Acceptability, safety, and perceived helpfulness ratings using Enhanced Moderated Online Social Therapy (N=93).

Questions	Mean (SD)	Median	Values, n (%)[a]
Overall acceptability			
How would you describe your overall experience on MOST+[b]?[c]	3.9 (0.8)	4	91 (98)
Please rate the helpfulness of using MOST+[d]	3.5 (0.9)	4	80 (86)
Please rate how quickly you were able to find what you needed on MOST+ (ease of use)[e]	3.7 (1.1)	4	80 (86)
Please rate how relevant you found the content on MOST+[f]	3.9 (1.0)	4	82 (88)
Safety and support			
Has using MOST+ helped you to better access support from others?[g]	3.59 (1.125)	4	79 (85)
Please rate whether using MOST+ helped you feel better[h]	3.38 (1.03)	3	76 (82)
Please rate whether using MOST+ helped you feel more socially connected[h]	3.18 (1.15)	3	70 (86)
Please rate whether you felt safe using MOST+[i]	4.43 (0.82)	5	90 (97)

[a]Number of cases responding in the positive range (3 or higher) based on complete responses.

[b]MOST+: Enhanced Moderated Online Social Therapy.

[c]Items rated from 1=not at all positive to 5=very positive.

[d]Items rated from 1=not at all helpful to 5=very helpful.

[e]Items rated from 1=not at all quickly to 5=very quickly.

[f]Items rated from 1=not at all relevant to 5=very relevant.

[g]Items rated from 1=not at all to 5=very much.

[h]Items rated from 1=not at all safe to 5=very safe.

[i]Items rated from 1=not at all confidential to 5=very confidential, asked of participants will full access only.

Table 3. Overview of secondary outcomes and measures used.

Outcomes of interest	Measures	Descriptions
Psychological distress	K10[a]	10-item, widely recommended measure of psychological distress; validated in adolescents [1]
Psychological well-being	WEMWS[b]	3 items of the WEMWS are included in the eheadspace Minimum Data Set and assessed in this study: "I've been interested in new things," "I've been feeling useful," and "I've been feeling good about myself" [2]
Perceived stress	PSS[c]	10-item measure of the degree to which situations in one's life are appraised as stressful. Widely used, with acceptable psychometric properties [3]
Depression	PHQ-9[d]	9-item measure of severity of depression. Validated in psychiatric and primary care populations [4,5]
Loneliness	UCLA[e]	20-item measure assessing how often the respondent feels disconnected from others. Highly acceptable reliability and validity [6]
Basic psychological needs of competence, relatedness, and autonomy	BPNS[f]	21-item measure with 3 subscales (competence, autonomy, and relatedness), drawing from self-determination theory [7]. Widely used [8,9]
Social support	FS[g]	6-item measure of perceived social isolation, with acceptable psychometric properties in the older adult population [10]
Strengths use	SUS[h]	14-item measure assessing the extent to which respondents use their strengths, drawing from positive psychology literature [11]
Mindfulness skills	FMI[i]	14-item measure of mindfulness. Appropriate for use in contexts where little experience or knowledge of mindfulness can be expected. Acceptable reliability and validity, including in clinical samples [12]

[a]K10: Kessler 10.

[b]WEMWS: 3 items from the Warwick-Edinburgh Mental Well-being Scale.

[c]PSS: Perceived Stress Scale.

[d]PHQ-9: Patient Health Questionnaire-9.

[e]UCLA: UCLA Loneliness Scale (Version 3).

[f]BPNS: Basic Psychological Need Satisfaction Scale.

[g]FS: Friendship Scale.

[h]SUS: Strengths Use Scale.

[i]FMI: Freiburg Mindfulness Inventory-Short Form.

Analyses

The patterns of intervention use were tracked in real time. Aggregated data from the user feedback questionnaire were compared with the a priori acceptability and safety criteria to determine the success of the pilot. Paired samples *t* tests were conducted and within-group effect sizes (Cohen *d*) were reported for changes in the baseline and posttest study measures. To estimate Cohen *d*, Morris and DeShon's [88] equation was applied to correct for dependence among the means in within-group designs. Parametric and nonparametric correlations were performed as appropriate to explore the associations between the usage of MOST+, acceptability ratings, and changes in secondary outcome measures.

Results

Feasibility, Acceptability, and Safety

A total of 93 of the 157 participants recruited for the study were contactable and assessed at follow-up. There were no statistically significant differences in any baseline demographic or clinical variables between those who completed the follow-up assessment and those who were lost to follow-up. All a priori indicators of acceptability were met (Table 2). The participants provided positive ratings of their experience using MOST+, with mean scores of 3.5 or more (out of 5) on each of the core domains of ease of use (mean 3.7, SD 1.1), relevancy (mean 3.9, SD 1.0), helpfulness (mean 3.5, SD 0.9), and overall experience (mean 3.9, SD 0.8). In addition, 98% (91/93) of participants reported a positive experience using MOST+, 86% (80/93) considered it easy to use, 88% (88/93) reported that MOST+ was relevant to their needs, 86% (80/93) considered it helpful, 82% (76/93) reported that using MOST+ helped them feel better, 86% (70/93) felt more socially connected using it, and 92% (86/93) said that they would recommend it to other young people experiencing similar difficulties. Moreover, 46% (72/157) of the participants had full access to MOST+ and 53% (83/157) had partial access (ie, excluded from the web-based social networking). The reasons for partial access included high clinical risk either detected by the system (automatically triggering a web chat; 18%: 28/157 (of all participants) or based on the clinician's assessment (4%: 6/157) and not being able to contact eligible participants to verify age (31%: 49/157). Participants with full access to MOST+ reported a significantly more positive overall experience (mean 4.1, SD 0.7) compared with those with partial access (mean 3.7, SD 0.8; two-tailed $t(91)=-2.89$; $P<.001$). The follow-up retention rate was also significantly higher in participants with full access (57/73, 78%)

than in those with partial access (43%; X^2_{155}=20.1; P<.001). There were no other significant differences in any demographic or outcome variable at baseline or acceptability ratings at follow-up between participants with full versus partial access.

A priori set safety criteria were also met. Specifically, no adverse events, inappropriate use, reports by participants, or unlawful entries pertaining to MOST+ were detected during the study. A total of 97% (90/93) of the participants reported feeling safe using MOST+. Moreover, all clinical measures showed a trend toward improved clinical status at follow-up (Table 1). There were no significant differences in safety indicators between those with full versus partial access.

Regarding the overall use of MOST+, there were a total of 1058 log-ins during the 9-week study, with 45.2% (71/157) logging in once, 14% (22/157) logging in twice, and 40.8% (64/157) logging in 3 or more times (Table 4). A total of 585 *steps* and 244 *do its* were completed during the study. All indicators of system usage were significantly higher in those with full access to MOST+ compared with those with partial access (Table 5). In terms of the duration of access to MOST+, 66.2% (104/157) of participants did not extend their initial default period of 7 days of enrollment, 14% (22/157) requested one extension, and 19.7% (31/157) requested 2 or more extensions. Finally, 41% (64/157) of participants did not request any chats with eheadspace clinicians, 34% (53/157) requested 1 chat, and 24% (38/157) requested 3 or more chats (including automatic web chats triggered by screening items indicating possible acute risk on initial registration into MOST+; Multimedia Appendix 1).

Table 4. Log-ins and individual usage of the main components of Enhanced Moderated Online Social Therapy (N=157) during the pilot study.

Full sample	Characteristics		
Site component	Mean (SD)	Range	Percentage, n (%)
Log-ins	6.74 (15.21)	1-103	86 (54.8[a])
Posts and comments	1.14 (4.69)	0-45	21 (14[b])
Steps	3.73 (9.88)	0-87	78 (49.7[c])
“Do its”	1.55 (6.53)	0-74	49.9 (31.8[d])

[a]Percentage of participants with more than 2 log-ins.

[b]Percentage of participants with more than 1 posts/comments.

[c]Percentage of participants completing more than 1 step.

[d]Percentage of participants completing more than 1 *do it*.

Table 5. Comparison of log-ins and individual use of the main components between participant groups with full access (n=73) and participant group with partial access (n=84).

Variables	Participants with full access (n=73)			Participants with partial access (n=84)				
	Mean (SD)	Range	Participants, n (%)	Mean (SD)	Range	Participants, n (%)	*t* test (df)	*P* value
Log-ins	12.34 (18.75)	1-103	58 (79.5[a])	1.87 (2.23)	1-18	30 (33.3[a])	−5.10 (155)	<.001
Post and comments	2.46 (6.66)	0-45	22 (30.1[b])	N/A[e]	N/A	N/A	N/A	N/A
Steps	6.52 (13.55)	0-87	49 (67.1[c])	1.30 (3.35)	1-23	29 (34.5[c])	−3.41 (155)	.001
“Do its”	2.78 (9.36)	0-74	31 (42.5[d])	0.49 (1.30)	0-7	19 (22.6[d])	−2.22 (155)	.03

[a]Percentage of participants with more than 2 log-ins.

[b]Percentage of participants with more than 1 posts/comments.

[c]Percentage of participants completing more than 1 step.

[d]Percentage of participants completing more than 1 *do it*.

[e]N/A: not applicable.

Secondary Outcome Variables

There were statistically significant improvements between baseline and follow-up assessments, with a small to medium size, in psychological distress (d=−0.39; P<.001), perceived stress (d=−0.44; P<.001), psychological well-being (d=0.51; P<.001), depression (d=−0.29; P=.008), loneliness (d=−0.23; P=.04), social support (d=0.30; P<.001), autonomy (d=0.36; P=.001), and self-competence (d=0.30; P<.001; Table 3). Moreover, the proportion of participants with K10 and PHQ-9 scores indicative of severe mental health disorder (K1030) or moderately severe depression (PHQ-915) was significantly lower at follow-up (72% for K10; 53% for PHQ-9) compared with baseline (82% for K10; 58% for PHQ-9; X^2_{91}=18.8; P<.001 for K10; X^2_{91}=19.7; P<.001 for PHQ-9).

For those with full access, a secondary analysis revealed that there were significant improvements in psychological distress (d=−0.38; P<.001), perceived stress (d=−0.37; P=.01), psychological well-being (d=0.38; P<.001), loneliness (d=−0.33;

P=.02), social support (*d*=0.25; *P*=.05), and autonomy (*d*=0.50; *P*<.001). Similarly, for those with partial access to MOST+, there were significant improvements in psychological distress (*d*=−0.40; *P*=.03), perceived stress (*d*=−0.55; *P*<.001), psychological well-being (*d*=0.72; *P*<.001), depression (*d*=−0.40; *P*=.03), social support (*d*=0.39; *P*=.03), and self-competence (*d*=0.42; *P*<.001).

Exploratory Correlations Between System Use, Acceptability Ratings, and Outcome Variables

Given the significantly lower system usage and overall retention rate in participants with partial access compared with those with full access, we reported exploratory correlations between system usage, acceptability ratings, and secondary outcome variables for participants who had full access to MOST+. In terms of system usage and acceptability ratings, there were significant correlations between (1) participants reporting that MOST+ helped them feel better and the number of web-based messages between clinicians and young people (Spearman rho, r_s=0.53; *P*<.01) and the number of weeks logging in (r_s=0.42; *P*<.01) and (2) participants reporting feeling more socially connected and the number of comments posted on the newsfeed (r_s=0.42; *P*<.01) as well as the number of contributions to *Talk it Out* (r_s=0.32; *P*=.01). With respect to system usage and secondary outcome variables, there were statistically significant correlations between increased relatedness and the number of log-ins (r_s=0.28; *P*=.03), number of steps completed (r_s=0.37*;* *P*<.01), as well as the combined number of *do its* and steps completed (r_s=0.40; *P*<.01). Increased social support also correlated positively with the number of completed *steps* (r_s=0.30*;* *P*=.02) and combined number of *do its* and steps. Moreover, increased strengths usage correlated positively with the number of completed *do its* (r_s=0.27; *P*=.03). Finally, there was a nonsignificant correlation in the expected direction between lower loneliness and the number of log-ins (r_s=−0.25; *P*=.06).

Discussion

To the best of our knowledge, this is the first study to develop and test a multimodal nationwide web-based mental health service for young people experiencing mental ill-health. As such, MOST+ was designed to be an all-in-one digital mental health app merging engaging, evidence-based therapy modules with expert clinician guidance, peer support, social networking, and real-time clinical support. Baseline clinical measures indicated that the majority of participants using MOST+ had moderate (52/157, 33%) to severe mental health conditions (96/157, 61%) and moderate to severe depressive symptoms (130/157, 83%). The results of this study showed that MOST+ was feasible, acceptable, and safe, with all acceptability and safety indicators exceeding the a priori established criteria. The high level of overall satisfaction and perceived helpfulness provided strong support for the relevance of the intervention content and features for help-seeking young people experiencing significant mental ill-health.

Secondary outcome variables showed significant improvements, with small to medium effect sizes, in 8 of the 11 outcomes assessed. These included psychological distress, perceived stress, psychological well-being, depression, loneliness, social support, autonomy, and self-competence. Similarly, the proportions of participants with severe mental health disorders and moderate to severe depression (as indicated by the K-10 and PHQ-9) were significantly lower at follow-up. Although the uncontrolled design of this study did not allow any causal inferences, it was worth noting that there were a number of significant correlations in the expected direction between several indicators of system usage (ie, number of log-ins, number of steps completed, number of *do its* completed, and number of posts made on the social network) and both perceived helpfulness (ie, participants reporting that MOST+ made them feel better and more socially connected) as well as secondary outcome variables (ie, relatedness and loneliness, strengths usage, and social support). These initial findings provide *proof of concept* for MOST+ and lend preliminary support to the potential therapeutic effects of the intervention.

MOST+ was designed as a scalable and efficient online youth mental health service integrating multiple modes of digital therapy, available 24/7, thus catering to individual needs and preferences of young people. The combination of treatment modalities integrated by MOST+ was reported in previous research [38,47,48,89-91]. Specifically, recent studies have found that web e chat can be as effective as face-to-face interventions and even out-perform face-to-face treatment, possibly via increased focus on essential treatment goals [47,48]. Writing in web chats can also have significant advantages, such as enabling users to re-read and reflect on the therapist's responses and their own emotions and promote a sense of control of the pace, content, and depth of disclosure [48]. Moreover, online chat integrated with a web-based intervention was shown to increase clinical effects compared with web-based support alone [89]. Finally, online peer support showed promise in improving mental health outcomes [38,90], potentially improving intervention adherence and satisfaction [45,92], although some studies found no additional therapeutic effect of peer support compared with traditional online interventions [91]. The patterns of use of MOST+, with different young people interacting with different features (ie, steps, web chat with clinicians, and web-based social network) over different timeframes (ie, one-time usage vs multiple extensions) coupled with the finding that 88% (81/93) of the participants reported that the intervention was relevant to their needs, provide support to this approach. Furthermore, the fact that 41% (64/157) of young people did not request a web chat while using MOST+ indicated that this model could increase the efficiency and scalability of traditional web chat services by reducing their reliance on real-time clinician-delivered web chat. Although promising, the cost-effectiveness of MOST+ against traditional web chat services needs to be established via controlled studies.

With the purpose of ensuring the safety of a (potentially) population-level social network, participants could be granted either full or partial access (excluded access to web-based social networking) to MOST+. This provided an opportunity to examine the differences in satisfaction levels, perceived safety, usage, and secondary outcomes in relation to the level of access to the system. Interestingly, participants with access to the social

network reported significantly higher levels of overall satisfaction, showed higher levels of usage of MOST+, and were more likely to be interviewed at follow-up (60/73, 78%) compared with their counterparts in the partial access group (36/84, 43%). Moreover, although loneliness and autonomy improved significantly in the group with full access, these domains remained unchanged in the partial access group. Taken together, these findings suggested that limiting access to the social network may have thwarted the sense of autonomy and the motivation to use the system and remained in the study in those with partial access. Conversely, having full access to the system and social network may lead to increased autonomy and reduced loneliness, irrespective of whether young people posted or not. It must be noted that there were no significant differences in any baseline variables (including clinical severity and basic psychological needs) between those with partial access and those with full access. These findings were in keeping with the self-determination theory, which posited that environments that addressed the basic psychological needs of autonomy (ie, sense that one's own behavior is freely chosen and of one's own volition), relatedness (ie, feelings of safety, belonging, and connectedness in their social interactions), and self-competence enhanced intrinsic motivation [93,94] as well as engagement in web-based interventions [95]. Alternatively, it could have been that participants with partial access were more difficult to contact because they were less motivated to use the system and participate in the study in the first place. Taken as a whole, these findings highlighted that, when designing web-based social media–based interventions, safety considerations should be weighed against potential negative effects on engagement or lack of positive effects on social outcomes. For example, future iterations of MOST+ could include an opt-in feature whereby young people decide whether they want to participate in the social network, thus preserving their sense of autonomy. In addition, less burdensome procedures could be introduced to verify age, which accounted for 57% (48/84) of all participants with partial access (eg, uploading an ID card picture via mobile phone as part of the onboarding process or linking their MOST+ account to the existing national web health records to automatically verify age). For those at increased clinical risk, additional safety measures (eg, enhanced automatic and manual monitoring of posts; personalized, detailed information on the purpose and terms of use of the social network) could be implemented to ensure clinical safety while including them in the social network. We successfully implemented this approach in a recent study of a social media–based intervention for young people with elevated suicidal ideation [96].

Web-based social media interventions provide a unique opportunity to address the pervasive rates of social isolation and lack of social support among young people with mental ill-health. For example, young people with psychosis report an average of three lonely days per week [97]. Come adulthood, 75%-94% will experience significant loneliness [98,99]. Unfortunately, recent studies have shown that frequent use of social media among young people is linked to lower self-esteem and increased anxiety, depression, and psychological distress [100,101], with young people with lower well-being being more vulnerable to experience adverse effects when using social media [102]. Against this backdrop, it is essential to develop evidence-based social media–enabled interventions that promote social support while ensuring safety and diminishing harmful consequences. A number of studies from our research lab [42-44,55] and others [39] have shown that carefully designed moderated web-based social media interventions can be safely implemented and are not associated with harmful effects. This study adds to this growing body of evidence by demonstrating that web-based social media interventions can be safely deployed to help-seeking young people via a national web counseling service. That said, the optimal size, functioning, and operations of social media–based interventions remain uncertain. For example, what level of engagement or participation is needed for participants to benefit from these interventions? What is the optimal balance between messages of distress and requests for help (which may lead to contagion [103]) versus messages of hope and positivity (which may alienate some participants [104])? Can web-based social networks provide a safe and transitional *training environment* that leads to real-world improvements? These and other questions will need to be examined in future research using mixed methods, including qualitative studies as well as novel methodologies and analytic techniques such as machine learning and natural language processing.

This study has several limitations. First, the uncontrolled design precluded any causal inferences about the efficacy of MOST+. Second, given that the study was implemented nationally and all assessments were conducted remotely, there was a 40% attrition rate at follow-up, which may have positively biased the results (ie, young people who felt more positively about the intervention may be more likely to be assessed at follow-up). That being said, the reported attrition rate is among the lowest reported by studies evaluating web-based interventions of the equivalent duration via remote assessments (35%-74%) [105-107], and there were no differences in any baseline demographic or clinical variables between those who completed the follow-up assessment and those who did not. Moreover, it is worth noting that MOST+ was not designed to promote sustained engagement, and it did not implement any strategies to foster engagement over time. On the basis of the purpose and function of the eheadspace web counseling service, MOST+ was intended to provide immediate short-term support to help-seeking young people and, where appropriate, redirect young people with long-term needs to youth mental health services. Third, multiple correlations were estimated, which is likely to increase the number of type I errors. Given the exploratory nature of these analyses, we did not adjust for multiple comparisons. Thus, these findings should be considered to be exploratory and interpreted with caution. Fourth, the short-term duration of the study precluded the examination of long-term outcomes. However, this design was considered appropriate, given the purpose of the intervention (ie, addressing immediate psychological distress) and the typically short-term engagement of young people with eheadspace. Finally, the multimodal nature of MOST+ precludes the examination of the effectiveness of the specific elements of MOST+, leading to improved outcomes. Although these research questions were outside the scope of this study, future research adopting novel approaches (eg, rapid A/B testing) will need to determine the optimal combination of features, therapeutic content, and level

of usage to improve technological markers (eg, penetration rate, satisfaction, and perceived helpfulness) as well as clinical (eg, symptoms) and social (eg, loneliness, social support) outcomes. Moreover, further research should determine whether MOST+ is effective in fostering help-seeking in, and addressing the needs of, young people with lower socio-economic status or young people at the risk of social exclusion.

Conclusions and Future Research

The results of this pilot investigation demonstrated that MOST+ is a highly promising and relevant web-based clinical service for young people with clinically significant mental ill-health as it yielded high satisfaction, safety, and perceived helpfulness as well as encouraging improvements in a wide range of clinical and social outcomes. These initial findings provide *proof of concept* for MOST+ and lend support to the multimodal, integrated approach of the intervention.

The effectiveness and cost-effectiveness of MOST+ will need to be established via controlled evaluations addressing the limitations of this study. For example, MOST+ could be implemented as a national service and evaluated through hybrid trial designs that blend components of clinical effectiveness and implementation research [108]. Alternatively, the national deployment of MOST+ would enable fast, well-powered, efficient randomized controlled trials evaluating the effectiveness of the different components of the intervention (eg, dismantling trials) as well as successive iterations of the service. The results and innovations of these trials could be rapidly assimilated into the mainstream service, thereby breaking the current divide and long-term delays between research and clinical implementation [109].

The results from this study indicate that MOST+ is a scalable web-based mental health service that enhances the capacity of traditional web counseling services. Future iterations of MOST+ will incorporate artificial intelligence (AI) and machine learning technologies to further enhance the efficiency of the service (eg, via triaging human support as required) as well as the personalization of the intervention [110]. This could also include chatbots harnessing natural language processing and AI to support participants in finding relevant content within the system, delivering basic therapeutic counseling, and initiating human involvement as needed. Finally, optional built-in video conferencing capabilities may enable faster therapeutic sessions while being able to assess nonverbal cues when necessary.

Finally, in addition to providing mental health support to young people who are not able to access face-to-face care, MOST+ could be integrated with the growing international network of youth mental health services to address wait-list issues, provide continuity of care in between therapy sessions, and offer relapse prevention support after initial treatment response. Meanwhile, MOST+, in its current form, stands to deliver an accessible and scalable web-based mental health service, providing multiple and integrated modalities of web-based support, to cater to the needs of an increasingly growing number of young people with mental ill-health.

Acknowledgments

The authors wish to thank the young people who generously participated in this study. In addition, the authors wish to thank the eheadspace senior management team and clinicians for supporting this study. This work was supported by a grant from the Young and Well Cooperative Research Centre funded by the National Health and Medical Research Council (NHMRC). MA was supported via a Career Development Fellowship (APP1082934) and an Investigator Grant (APP1177235) by NHMRC. The sponsors did not participate in the design or conduct of this study; in the collection, management, analysis, or interpretation of data; in the writing of the manuscript; or in the preparation, review, approval, or decision to submit this manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study procedure. MOST+: Enhanced Moderated Online Social Therapy.
[PNG File , 576 KB - jmir_v22i8e17155_app1.png]

Multimedia Appendix 2
Example participant timelines through the Enhanced Moderated Online Social Therapy (MOST+) intervention.
[PNG File , 342 KB - jmir_v22i8e17155_app2.png]

Multimedia Appendix 3
Enhanced Moderated Online Social Therapy (MOST+) intervention safety algorithm.
[PNG File , 288 KB - jmir_v22i8e17155_app3.png]

References

1. Saxena S, Funk MK, Chisholm D. Comprehensive mental health action plan 2013-2020. East Mediterr Health J 2015 Sep 28;21(7):461-463 [FREE Full text] [Medline: 26442884]

2. Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. JAMA Psychiatry 2015 Apr;72(4):334-341 [FREE Full text] [doi: 10.1001/jamapsychiatry.2014.2502] [Medline: 25671328]
3. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
4. Alvarez-Jimenez M, Priede A, Hetrick S, Bendall S, Killackey E, Parker A, et al. Risk factors for relapse following treatment for first episode psychosis: a systematic review and meta-analysis of longitudinal studies. Schizophr Res 2012 Aug;139(1-3):116-128. [doi: 10.1016/j.schres.2012.05.007] [Medline: 22658527]
5. Melvin GA, Dudley AL, Gordon MS, Ford S, Taffe J, Tonge BJ. What happens to depressed adolescents? A follow-up study into early adulthood. J Affect Disord 2013 Oct;151(1):298-305. [doi: 10.1016/j.jad.2013.06.012] [Medline: 23829999]
6. Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. Lancet 2007 Apr 14;369(9569):1302-1313. [doi: 10.1016/S0140-6736(07)60368-7] [Medline: 17434406]
7. Naicker K, Galambos NL, Zeng Y, Senthilselvan A, Colman I. Social, demographic, and health outcomes in the 10 years following adolescent depression. J Adolesc Health 2013 May;52(5):533-538. [doi: 10.1016/j.jadohealth.2012.12.016] [Medline: 23499382]
8. Draine J, Salzer MS, Culhane DP, Hadley TR. Role of social disadvantage in crime, joblessness, and homelessness among persons with serious mental illness. Psychiatr Serv 2002 May;53(5):565-573. [doi: 10.1176/appi.ps.53.5.565] [Medline: 11986504]
9. Gibb SJ, Fergusson DM, Horwood LJ. Burden of psychiatric disorder in young adulthood and life outcomes at age 30. Br J Psychiatry 2010 Aug;197(2):122-127. [doi: 10.1192/bjp.bp.109.076570] [Medline: 20679264]
10. Slade J, Teesson W, Burgess P. The Mental Health of Australians 2: Report on the 2007 National Survey of Mental Health and Wellbeing. Australian Government Department of Health. 2009. URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/mental-pubs-m-mhaust2 [accessed 2020-07-21]
11. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
12. Thornicroft G. Stigma and discrimination limit access to mental health care. Epidemiol Psichiatr Soc 2008;17(1):14-19. [doi: 10.1017/s1121189x00002621] [Medline: 18444452]
13. Salaheddin K, Mason B. Identifying barriers to mental health help-seeking among young adults in the UK: a cross-sectional survey. Br J Gen Pract 2016 Oct;66(651):e686-e692 [FREE Full text] [doi: 10.3399/bjgp16X687313] [Medline: 27688518]
14. Jorm AF, Wright A, Morgan AJ. Where to seek help for a mental disorder? National survey of the beliefs of Australian youth and their parents. Med J Aust 2007 Nov 19;187(10):556-560. [Medline: 18021042]
15. Anderson M, Jiang J. Teens, Social Media & Technology 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2020-07-21]
16. Rideout V, Robb M. Social Media, Social Life: Teens Reveal Their Experiences, 2018. Common Sense Media: Age-Based Media Reviews for Families. 2018. URL: https://www.commonsensemedia.org/research/social-media-social-life-2018 [accessed 2020-07-21]
17. Birnbaum ML, Rizvi AF, Confino J, Correll CU, Kane JM. Role of social media and the internet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. Early Interv Psychiatry 2017 Aug;11(4):290-295 [FREE Full text] [doi: 10.1111/eip.12237] [Medline: 25808317]
18. Lawlor A, Kirakowski J. Online support groups for mental health: a space for challenging self-stigma or a means of social avoidance? Comput Hum Behav 2014 Mar;32:152-161. [doi: 10.1016/j.chb.2013.11.015]
19. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
20. Best P, Gil-Rodriguez E, Manktelow R, Taylor BJ. Seeking help from everyone and no-one: conceptualizing the online help-seeking process among adolescent males. Qual Health Res 2016 Jul;26(8):1067-1077. [doi: 10.1177/1049732316648128] [Medline: 27179020]
21. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
22. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
23. Richards D, Richardson T, Timulak L, McElvaney J. The efficacy of internet-delivered treatment for generalized anxiety disorder: a systematic review and meta-analysis. Internet Interv 2015 Sep;2(3):272-282. [doi: 10.1016/j.invent.2015.07.003]
24. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research . . . and three things we can do about them. Psychiatr Serv 2017 May 1;68(5):427-429 [FREE Full text] [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
25. Mohr DC, Riper H, Schueller SM. A solution-focused research approach to achieve an implementable revolution in digital mental health. JAMA Psychiatry 2018 Feb 1;75(2):113-114. [doi: 10.1001/jamapsychiatry.2017.3838] [Medline: 29238805]

26. Aitken M, Clancy B, Nass D. The Growing Value of Digital Health. IQVIA. 2017. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health [accessed 2020-07-21]
27. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iphone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]
28. Leigh S, Flatt S. App-based psychological interventions: friend or foe? Evid Based Ment Health 2015 Nov;18(4):97-99. [doi: 10.1136/eb-2015-102203] [Medline: 26459466]
29. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
30. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
31. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
32. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
33. Buntrock C, Ebert DD, Lehr D, Smit F, Riper H, Berking M, et al. Effect of a web-based guided self-help intervention for prevention of major depression in adults with subthreshold depression: a randomized clinical trial. J Am Med Assoc 2016 May 3;315(17):1854-1863. [doi: 10.1001/jama.2016.4326] [Medline: 27139058]
34. Kampmann IL, Emmelkamp PM, Morina N. Meta-analysis of technology-assisted interventions for social anxiety disorder. J Anxiety Disord 2016 Aug;42:71-84. [doi: 10.1016/j.janxdis.2016.06.007] [Medline: 27376634]
35. Kuester A, Niemeyer H, Knaevelsrud C. Internet-based interventions for posttraumatic stress: a meta-analysis of randomized controlled trials. Clin Psychol Rev 2016 Feb;43:1-16. [doi: 10.1016/j.cpr.2015.11.004] [Medline: 26655959]
36. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
37. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153 [FREE Full text] [doi: 10.2196/jmir.7725] [Medline: 28490417]
38. Rice SM, Goodall J, Hetrick SE, Parker AG, Gilbertson T, Amminger GP, et al. Online and social networking interventions for the treatment of depression in young people: a systematic review. J Med Internet Res 2014 Sep 16;16(9):e206 [FREE Full text] [doi: 10.2196/jmir.3304] [Medline: 25226790]
39. Easton K, Diggle J, Ruethi-Davis M, Holmes M, Byron-Parker D, Nuttall J, et al. Qualitative exploration of the potential for adverse events when using an online peer support network for mental health: cross-sectional survey. JMIR Ment Health 2017 Oct 30;4(4):e49 [FREE Full text] [doi: 10.2196/mental.8168] [Medline: 29084710]
40. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
41. Ridout B, Campbell A. The use of social networking sites in mental health interventions for young people: systematic review. J Med Internet Res 2018 Dec 18;20(12):e12244 [FREE Full text] [doi: 10.2196/12244] [Medline: 30563811]
42. Alvarez-Jimenez M, Bendall S, Lederman R, Wadley G, Chinnery G, Vargas S, et al. On the HORYZON: moderated online social therapy for long-term recovery in first episode psychosis. Schizophr Res 2013 Jan;143(1):143-149. [doi: 10.1016/j.schres.2012.10.009] [Medline: 23146146]
43. Alvarez-Jimenez M, Gleeson J, Bendall S, Penn D, Yung A, Ryan R, et al. Enhancing social functioning in young people at ultra high risk (UHR) for psychosis: a pilot study of a novel strengths and mindfulness-based online social therapy. Schizophr Res 2018 Dec;202:369-377. [doi: 10.1016/j.schres.2018.07.022] [Medline: 30031616]
44. Rice S, Gleeson J, Davey C, Hetrick S, Parker A, Lederman R, et al. Moderated online social therapy for depression relapse prevention in young people: pilot study of a 'next generation' online intervention. Early Interv Psychiatry 2018 Aug;12(4):613-625. [doi: 10.1111/eip.12354] [Medline: 27311581]
45. Lederman R, Wadley G, Gleeson J, Bendall S, Álvarez-Jiménez M. Moderated online social therapy. ACM Trans Comput-Hum Interact 2014 Feb;21(1):1-26. [doi: 10.1145/2513179]
46. Dowling M, Rickwood D. Online counseling and therapy for mental health problems: a systematic review of individual synchronous interventions using chat. J Tech Hum Serv 2013 Jan;31(1):1-21. [doi: 10.1080/15228835.2012.728508]
47. Gieselmann A, Pietrowsky R. The effects of brief chat-based and face-to-face psychotherapy for insomnia: a randomized waiting list controlled trial. Sleep Med 2019 Sep;61:63-72. [doi: 10.1016/j.sleep.2019.03.024] [Medline: 31262551]
48. Richards D, Viganó N. Online counseling: a narrative and critical review of the literature. J Clin Psychol 2013 Sep;69(9):994-1011. [doi: 10.1002/jclp.21974] [Medline: 23630010]

49. Rickwood D, Webb M, Kennedy V, Telford N. Who are the young people choosing web-based mental health support? Findings from the implementation of Australia's national web-based youth mental health service, eHeadSpace. JMIR Ment Health 2016 Aug 25;3(3):e40 [FREE Full text] [doi: 10.2196/mental.5988] [Medline: 27562729]
50. Alvarez-Jimenez M, Bendall S, Koval P, Rice S, Cagliarini D, Valentine L, et al. HORYZONS trial: protocol for a randomised controlled trial of a moderated online social therapy to maintain treatment effects from first-episode psychosis services. BMJ Open 2019 Feb 19;9(2):e024104 [FREE Full text] [doi: 10.1136/bmjopen-2018-024104] [Medline: 30782893]
51. Bailey E, Alvarez-Jimenez M, Robinson J, D'Alfonso S, Nedeljkovic M, Davey CG, et al. An enhanced social networking intervention for young people with active suicidal ideation: safety, feasibility and acceptability outcomes. Int J Environ Res Public Health 2020 Apr 3;17(7):2435 [FREE Full text] [doi: 10.3390/ijerph17072435] [Medline: 32260111]
52. Rice S, O'Bree B, Wilson M, McEnery C, Lim MH, Hamilton M, et al. Leveraging the social network for treatment of social anxiety: pilot study of a youth-specific digital intervention with a focus on engagement of young men. Internet Interv 2020 Apr;20:100323 [FREE Full text] [doi: 10.1016/j.invent.2020.100323] [Medline: 32435600]
53. McEnery C, Lim MH, Knowles A, Rice S, Gleeson J, Howell S, et al. Social anxiety in young people with first-episode psychosis: pilot study of the EMBRACE moderated online social intervention. Early Interv Psychiatry 2019 Dec 30:- epub ahead of print. [doi: 10.1111/eip.12912] [Medline: 31889431]
54. Gleeson J, Lederman R, Herrman H, Koval P, Eleftheriadis D, Bendall S, et al. Moderated online social therapy for carers of young people recovering from first-episode psychosis: study protocol for a randomised controlled trial. Trials 2017 Jan 17;18(1):27 [FREE Full text] [doi: 10.1186/s13063-016-1775-5] [Medline: 28095883]
55. Gleeson J, Lederman R, Koval P, Wadley G, Bendall S, Cotton S, et al. Moderated online social therapy: a model for reducing stress in carers of young people diagnosed with mental health disorders. Front Psychol 2017;8:485 [FREE Full text] [doi: 10.3389/fpsyg.2017.00485] [Medline: 28421012]
56. Rice S, Gleeson J, Leicester S, Bendall S, D'Alfonso S, Gilbertson T, et al. Implementation of the enhanced moderated online social therapy (MOST+) model within a national youth e-mental health service (eHeadSpace): protocol for a single group pilot study for help-seeking young people. JMIR Res Protoc 2018 Feb 22;7(2):e48 [FREE Full text] [doi: 10.2196/resprot.8813] [Medline: 29472177]
57. Mueser K, Drake RE. How does a practice become evidenced-based? In: Drake RE, Merrens MR, Lynde DW, editors. Evidence-Based Mental Health Practice: A Textbook. Manhattan, New York: Norton Professional Books; 2005.
58. Andrews G, Slade T. Interpreting scores on the Kessler psychological distress scale (K10). Aust N Z J Public Health 2001 Dec;25(6):494-497. [doi: 10.1111/j.1467-842x.2001.tb00310.x] [Medline: 11824981]
59. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals 2002 Sep 1;32(9):509-515 [FREE Full text] [doi: 10.3928/0048-5713-20020901-06]
60. Chan SM, Fung TC. Reliability and validity of K10 and K6 in screening depressive symptoms in Hong Kong adolescents. Vulnerable Child Youth Stud 2013 Nov 25;9(1):75-85. [doi: 10.1080/17450128.2013.861620]
61. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
62. Lee E. Review of the psychometric evidence of the perceived stress scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127 [FREE Full text] [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]
63. Beard C, Hsu K, Rifkin L, Busch A, Björgvinsson T. Validation of the PHQ-9 in a psychiatric sample. J Affect Disord 2016 Mar 15;193:267-273. [doi: 10.1016/j.jad.2015.12.075] [Medline: 26774513]
64. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010;8(4):348-353 [FREE Full text] [doi: 10.1370/afm.1139] [Medline: 20644190]
65. Russell DW. UCLA loneliness scale (version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
66. Deci EL, Ryan RM. The 'what' and 'why' of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
67. Gagné M. The role of autonomy support and autonomy orientation in prosocial behavior engagement. Motiv Emot 2003;27(3):199-223. [doi: 10.1023/A:1025007614869]
68. Johnston MM, Finney SJ. Measuring basic needs satisfaction: evaluating previous research and conducting new psychometric evaluations of the basic needs satisfaction in general scale. Contemp Educ Psychol 2010 Oct;35(4):280-296. [doi: 10.1016/j.cedpsych.2010.04.003]
69. Hawthorne G. Perceived social isolation in a community sample: its prevalence and correlates with aspects of peoples' lives. Soc Psychiatry Psychiatr Epidemiol 2008 Feb;43(2):140-150. [doi: 10.1007/s00127-007-0279-8] [Medline: 17994175]
70. Wood AM, Linley PA, Maltby J, Kashdan TB, Hurling R. Using personal and psychological strengths leads to increases in well-being over time: a longitudinal study and the development of the strengths use questionnaire. Pers Individ Differ 2011 Jan;50(1):15-19. [doi: 10.1016/j.paid.2010.08.004]
71. Walach H, Buchheld N, Buttenmüller V, Kleinknecht N, Schmidt S. Measuring mindfulness—the Freiburg mindfulness inventory (FML). Pers Individ Differ 2006 Jun;40(8):1543-1555. [doi: 10.1016/j.paid.2005.11.025]

72. Green MJ, Myers KR. Graphic medicine: use of comics in medical education and patient care. Br Med J 2010 Mar 3;340:c863. [doi: 10.1136/bmj.c863] [Medline: 20200064]
73. Williams IC. Graphic medicine: comics as medical narrative. Med Humanit 2012 Jun;38(1):21-27. [doi: 10.1136/medhum-2011-010093] [Medline: 22282425]
74. McEnery C, Lim MH, Knowles A, Rice S, Gleeson J, Howell S, et al. Development of a moderated online intervention to treat social anxiety in first-episode psychosis. Front Psychiatry 2019;10:581 [FREE Full text] [doi: 10.3389/fpsyt.2019.00581] [Medline: 31474889]
75. Rice S, O'Bree B, Wilson M, McEnery C, Lim MH, Hamilton M, et al. Development of a graphic medicine - enabled social media - based intervention for youth social anxiety. Clin Psychol 2020 May 5:- (forthcoming). [doi: 10.1111/cp.12222]
76. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. Can J Psychiatry 2013 Jul;58(7):376-385. [doi: 10.1177/070674371305800702] [Medline: 23870719]
77. Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G. Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. Br J Psychiatry 2010 Mar;196(3):173-178. [doi: 10.1192/bjp.bp.109.066001] [Medline: 20194536]
78. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
79. Sevilla-Llewellyn-Jones J, Santesteban-Echarri O, Pryor I, McGorry P, Alvarez-Jimenez M. Web-based mindfulness interventions for mental health treatment: systematic review and meta-analysis. JMIR Ment Health 2018 Sep 25;5(3):e10278 [FREE Full text] [doi: 10.2196/10278] [Medline: 30274963]
80. Goldberg SB, Tucker RP, Greene PA, Davidson RJ, Wampold BE, Kearney DJ, et al. Mindfulness-based interventions for psychiatric disorders: a systematic review and meta-analysis. Clin Psychol Rev 2018 Feb;59:52-60 [FREE Full text] [doi: 10.1016/j.cpr.2017.10.011] [Medline: 29126747]
81. Wilson AC, Mackintosh K, Power K, Chan SW. Effectiveness of self-compassion related therapies: a systematic review and meta-analysis. Mindfulness 2018 Oct 24;10(6):979-995. [doi: 10.1007/s12671-018-1037-6]
82. Bolier L, Haverman M, Westerhof GJ, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. BMC Public Health 2013 Feb 8;13:119 [FREE Full text] [doi: 10.1186/1471-2458-13-119] [Medline: 23390882]
83. D'Zurilla TJ, Goldfried MR. Problem solving and behavior modification. J Abnorm Psychol 1971 Aug;78(1):107-126. [doi: 10.1037/h0031360] [Medline: 4938262]
84. Nezu AM. Problem solving and behavior therapy revisited. Behav Ther 2004;35(1):1-33. [doi: 10.1016/s0005-7894(04)80002-9]
85. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
86. Bellamy C, Schmutte T, Davidson L. An update on the growing evidence base for peer support. Ment Health Soc Incl 2017 Jun 12;21(3):161-167. [doi: 10.1108/mhsi-03-2017-0014]
87. Bargas-Avila J, Hornbæk K. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI'11; May 7-12, 2011; Vancouver, BC, Canada. [doi: 10.1145/1978942.1979336]
88. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol Methods 2002 Mar;7(1):105-125. [doi: 10.1037/1082-989x.7.1.105] [Medline: 11928886]
89. Schaub MP, Wenger A, Berg O, Beck T, Stark L, Buehler E, et al. A web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: three-arm randomized controlled trial. J Med Internet Res 2015 Oct 13;17(10):e232 [FREE Full text] [doi: 10.2196/jmir.4860] [Medline: 26462848]
90. Ali K, Farrer L, Gulliver A, Griffiths KM. Online peer-to-peer support for young people with mental health problems: a systematic review. JMIR Ment Health 2015;2(2):e19 [FREE Full text] [doi: 10.2196/mental.4418] [Medline: 26543923]
91. Low KG, Charanasomboon S, Lesser J, Reinhalter K, Martin R, Jones H, et al. Effectiveness of a computer-based interactive eating disorders prevention program at long-term follow-up. Eat Disord 2006;14(1):17-30. [doi: 10.1080/10640260500403816] [Medline: 16757446]
92. Biagianti B, Quraishi SH, Schlosser DA. Potential benefits of incorporating peer-to-peer interactions into digital interventions for psychotic disorders: a systematic review. Psychiatr Serv 2018 Apr 1;69(4):377-388 [FREE Full text] [doi: 10.1176/appi.ps.201700283] [Medline: 29241435]
93. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
94. Ryan RM. Facilitating health behavior change and its maintenance: interventions based on self-determination theory. Eur Health Psychol 2008;10(1):2-5 [FREE Full text]

95. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
96. Bailey E, Rice S, Robinson J, Nedeljkovic M, Alvarez-Jimenez M. Theoretical and empirical foundations of a novel online social networking intervention for youth suicide prevention: a conceptual review. J Affect Disord 2018 Oct 1;238:499-505. [doi: 10.1016/j.jad.2018.06.028] [Medline: 29936387]
97. Sündermann O, Onwumere J, Bebbington P, Kuipers E. Social networks and support in early psychosis: potential mechanisms. Epidemiol Psychiatr Sci 2013 Jun;22(2):147-150 [FREE Full text] [doi: 10.1017/S2045796012000601] [Medline: 23098565]
98. Badcock JC, Shah S, Mackinnon A, Stain HJ, Galletly C, Jablensky A, et al. Loneliness in psychotic disorders and its association with cognitive function and symptom profile. Schizophr Res 2015 Dec;169(1-3):268-273. [doi: 10.1016/j.schres.2015.10.027] [Medline: 26527247]
99. Lim MH, Gleeson JF, Alvarez-Jimenez M, Penn DL. Loneliness in psychosis: a systematic review. Soc Psychiatry Psychiatr Epidemiol 2018 Mar;53(3):221-238. [doi: 10.1007/s00127-018-1482-5] [Medline: 29327166]
100. Vannucci A, Flannery KM, Ohannessian CM. Social media use and anxiety in emerging adults. J Affect Disord 2017 Jan 1;207:163-166. [doi: 10.1016/j.jad.2016.08.040] [Medline: 27723539]
101. Lin L, Sidani JE, Shensa A, Radovic A, Miller E, Colditz JB, et al. Association between social media use and depression among US Young adults. Depress Anxiety 2016 Apr;33(4):323-331 [FREE Full text] [doi: 10.1002/da.22466] [Medline: 26783723]
102. Rideout VJ. The Common Sense Census: Media Use by Tweens and Teens, 2019. Common Sense Media: Age-Based Media Reviews for Families. 2015. URL: https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-tweens-and-teens-2019 [accessed 2020-07-21]
103. Kendal S, Kirk S, Elvey R, Catchpole R, Pryjmachuk S. How a moderated online discussion forum facilitates support for young people with eating disorders. Health Expect 2017 Feb;20(1):98-111 [FREE Full text] [doi: 10.1111/hex.12439] [Medline: 26725547]
104. Lederman R, Gleeson J, Wadley G, D'alfonso S, Rice S, Santesteban-Echarri O, et al. Support for carers of young people with mental illness. ACM Trans Comput-Hum Interact 2019 Feb 23;26(1):1-33. [doi: 10.1145/3301421]
105. Donker T, Bennett K, Bennett A, Mackinnon A, van Straten A, Cuijpers P, et al. Internet-delivered interpersonal psychotherapy versus internet-delivered cognitive behavioral therapy for adults with depressive symptoms: randomized controlled noninferiority trial. J Med Internet Res 2013 May 13;15(5):e82 [FREE Full text] [doi: 10.2196/jmir.2307] [Medline: 23669884]
106. Manicavasagar V, Horswood D, Burckhardt R, Lum A, Hadzi-Pavlovic D, Parker G. Feasibility and effectiveness of a web-based positive psychology program for youth mental health: randomized controlled trial. J Med Internet Res 2014 Jun 4;16(6):e140 [FREE Full text] [doi: 10.2196/jmir.3176] [Medline: 24901900]
107. Powell J, Hamborg T, Stallard N, Burls A, McSorley J, Bennett K, et al. Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial. J Med Internet Res 2012 Dec 31;15(1):e2 [FREE Full text] [doi: 10.2196/jmir.2240] [Medline: 23302475]
108. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
109. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med 2011 Dec;104(12):510-520 [FREE Full text] [doi: 10.1258/jrsm.2011.110180] [Medline: 22179294]
110. D'Alfonso S, Santesteban-Echarri O, Rice S, Wadley G, Lederman R, Miles C, et al. Artificial intelligence-assisted online social therapy for youth mental health. Front Psychol 2017;8:796 [FREE Full text] [doi: 10.3389/fpsyg.2017.00796] [Medline: 28626431]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
K10: Kessler 10
MOST: Moderated Online Social Therapy
MOST+: Enhanced Moderated Online Social Therapy
NHMRC: National Health and Medical Research Council
PHQ-9: Patient Health Questionnaire-9
SNS: social networking sites

Edited by G Eysenbach; submitted 21.11.19; peer-reviewed by T Campellone, L Schwab-Reese, J Chu, K Fauvelle; comments to author 29.02.20; revised version received 12.05.20; accepted 13.06.20; published 13.08.20.

Please cite as:
Alvarez-Jimenez M, Rice S, D'Alfonso S, Leicester S, Bendall S, Pryor I, Russon P, McEnery C, Santesteban-Echarri O, Da Costa G, Gilbertson T, Valentine L, Solves L, Ratheesh A, McGorry PD, Gleeson J
A Novel Multimodal Digital Service (Moderated Online Social Therapy+) for Help-Seeking Young People Experiencing Mental Ill-Health: Pilot Evaluation Within a National Youth E-Mental Health Service
J Med Internet Res 2020;22(8):e17155
URL: https://www.jmir.org/2020/8/e17155
doi:10.2196/17155
PMID:32788151

Original Paper

Text Messaging Adherence Intervention for Adolescents and Young Adults with Chronic Kidney Disease: Pilot Randomized Controlled Trial and Stakeholder Interviews

Cyd Eaton[1], PhD; Margaret Comer[1], BA; Cozumel Pruette[2], MD, MHS; Kevin Psoter[3], PhD; Kristin Riekert[1], PhD

[1]Pulmonary & Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[2]Division of Pediatric Nephrology, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[3]Division of General Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD, United States

Corresponding Author:
Cyd Eaton, PhD
Pulmonary & Critical Care Medicine
Johns Hopkins University School of Medicine
5501 Hopkins Bayview Circle
Baltimore, MD, 21224
United States
Phone: 1 410 550 3988
Email: ceaton4@jhmi.edu

Abstract

Background: Up to one-third of adolescents and young adults (11-21 years old) with chronic kidney disease exhibit suboptimal rates of adherence to renal-protective antihypertensive medications. Mobile health interventions may promote higher adherence to these medicines in these individuals, but empirical research is needed to inform best practices for applying these modalities.

Objective: In this multiphase investigation, we developed and tested a theoretically informed text messaging intervention based on the COM-B model, a well-established health intervention framework stating that capability, opportunity, and motivation interactively modify health behaviors, to improve participants' antihypertensive medication adherence in a pilot randomized controlled trial. Qualitative data on user experiences were obtained.

Methods: In phase 1, intervention messages (Reminder+COM-B Message) were developed via stakeholder engagement of participants and pediatric nephrologists. In phase 2, the Reminder+COM-B Message intervention was tested against a Reminder-only Message active control condition in an 8-week pilot randomized controlled trial. The primary outcome was daily electronically monitored antihypertensive medication adherence and secondary outcomes included pre-post participant surveys of adherence self-efficacy, adherence barriers, outcome expectancies for taking medicine, and motivation for and importance of taking medicine. In phase 3, qualitative interviews related to user experiences were conducted with participants in the Reminder+COM-B Message intervention group.

Results: Following phase 1, 34 participants (mean age 16.59 years, 41% female, 38% African American/Black, 35% hypertension diagnosis) completed the phase 2 pilot randomized controlled trial (n=18 in the Reminder+COM-B Message intervention group, n=16 in the Reminder-only Message active control group). All participants in the Reminder+COM-B Message intervention group completed a phase 3 qualitative interview. Overall, study procedures were feasible and the Reminder+COM-B Message intervention was acceptable to the participants (eg, 15/18 participants reported reading the majority of messages sent to them, 0/18 reported that the messages reduced their desire to take medicine). Prerandomization, there were no significant group differences in the rate of change in daily adherence over time. However, postrandomization, there was a significant group by time interaction (B=.01, *P*=.04) in which daily adherence decreased significantly over time in the Reminder-only Message active control group but remained stable in the Reminder+COM-B Message intervention group. There were no significant differences between groups in pre-post changes in survey responses. Qualitative interviews revealed participants' perceptions of how the Reminder+COM-B Message intervention changed adherence behavior and highlighted several areas for improving the intervention (eg, adapt messaging timing, intensity, and content to match daily adherence, send praise when medicine is taken).

Conclusions: The Reminder+COM-B Message intervention was feasible and acceptable to adolescents/young adults and demonstrated potential to promote participants' daily medication adherence beyond simple reminders. Further research is needed

to determine the Reminder+COM-B Message intervention's mechanisms of adherence behavior change and to incorporate qualitative participant feedback into a modified version of this intervention to enhance its acceptability.

Trial Registration: ClinicalTrials.gov NCT03651596; https://clinicaltrials.gov/ct2/show/NCT03651596

(J Med Internet Res 2020;22(8):e19861) doi:10.2196/19861

KEYWORDS

medication adherence; mobile health; pediatrics; kidney diseases; kidney; mHealth; adherence; adolescent; young adult; intervention

Introduction

Several antihypertensive medications serve as renal protective agents by lowering blood pressure or decreasing urinary protein excretion with the goal of slowing disease progression in children with chronic kidney disease (CKD) [1,2]. However, up to one-third of adolescents and young adults (AYAs) with CKD do not consistently take these medicines [3]. AYAs with CKD could benefit from behavior intervention to improve antihypertensive medication adherence, which in turn could improve long-term health outcomes. Adherence interventions exist for AYAs with other chronic medical conditions [4], but typically involve face-to-face visits, thereby limiting efficient dissemination and patient access [5]. Mobile health (mHealth) interventions offer solutions to these barriers by delivering the intervention to nearly anyone with a smartphone, including 95% of adolescents in the United States [6]. However, to date, there has been very little rigorous empirical investigation of mHealth adherence intervention efficacy or the application of mHealth approaches for AYAs with CKD [7,8].

mHealth interventions have been developed for medication adherence in children and AYAs with other medical conditions (eg, sickle cell disease, asthma, type 1 diabetes, migraine). However, these interventions have primarily relied on daily dose reminders [9-11], which show short-term effects on adherence and have barriers to long-term practicality (eg, repeated reminders could be viewed as intrusive instead of helpful) [12,13], or involve self-tracking daily adherence (eg, self-reported logs or video recordings of taking medicine), which have low engagement and completion rates [14,15]. Other limitations to existing interventions include no evidence for their efficacy on objective adherence behavior [10,16], no theoretical basis to an a priori outline of the intervention's mechanisms of behavior change [17], relying on components with barriers to long-term sustainability (eg, monetary incentives for taking medicine) [15], or not actively involving stakeholders in intervention development and refinement, which is essential for AYA engagement and intervention uptake [18].

The aim of the current study was to address these limitations of prior interventions by developing and testing a theoretically informed antihypertensive medication adherence-promoting mHealth messaging intervention for AYAs with CKD using objective adherence outcomes and qualitative stakeholder feedback. Our messaging intervention was based on the COM-B model, a well-established health intervention framework stating that *capability* (ie, skills, knowledge, and ability to complete a behavior), *opportunity* (ie, environmental factors prompting a behavior), and *motivation* (ie, conscious, reflective, and learned reasons for engaging in a behavior) interactively modify health *behaviors*, including adherence [19]. Opportunity is often targeted in mHealth interventions via medicine reminder messages. Longitudinal research involving AYAs with CKD has shown that higher capability and motivation to adhere to medications is associated with higher objective antihypertensive medication adherence [20]. Hence, it was expected that sending a reminder message (adherence opportunity) with additional messaging aimed at improving AYAs' adherence capability (eg, teach skills to reduce barriers and improve adherence self-efficacy) and motivation (eg, highlight the benefits of adherence to activate desire to take medicine) would improve medication adherence, which formed the basis for the design of our Reminder+COM-B Message intervention.

In the current study, we conducted a mixed methods investigation, which involved (1) developing the Reminder+COM-B Message intervention for antihypertensive medication adherence in AYAs with CKD (Phase 1), (2) preliminarily evaluation of the Reminder+COM-B Message intervention against a Reminder-only Message active control condition in an 8-week pilot randomized controlled trial (RCT; Phase 2), and (3) obtaining poststudy qualitative feedback from AYAs randomized to the Reminder+COM-B Message intervention (Phase 3). We hypothesized that (1) study procedures would be feasible (few to no technical issues implementing study procedures, low cost), and acceptable and engaging to the majority of AYAs (ie, few to no AYA-reported issues with study procedures in either group, majority of AYAs in the Reminder+COM-B Message intervention group would report reading or noticing the messages sent, <10% of AYAs would report that the Reminder+COM-B Message intervention reduced their desire to take medicine); (2) postrandomization (Phase 2), daily adherence (dose taken or not) would show a faster rate of improvement in the Reminder+COM-B Message intervention group compared to the Reminder-only Message active control group; and (3) pre-post survey scores representing AYA perceptions of adherence motivation and capability would show greater improvement in the Reminder+COM-B Message intervention group compared to the Reminder-only Message active control group (Phase 2). Given conventions for analyzing adherence as an overall mean, we also examined overall mean changes in adherence from baseline to postrandomization according to group allocation (Phase 2). During qualitative interviews (Phase 3), we probed AYAs' perceptions on the Reminder+COM-B Message intervention's mechanisms of behavior change and suggestions for improving its acceptability.

Methods

Procedure

Recruitment

The Johns Hopkins School of Medicine Institutional Review Board approved all study procedures prior to recruitment, which occurred at a single pediatric nephrology clinic from October 2018 to November 2019. Trained study staff identified potentially eligible AYAs via clinic roster and electronic medical record review. Inclusion criteria were: aged 11-21 years old, CKD diagnosis, current antihypertensive medication prescription (pill form only), and access to a mobile phone that received text messages. Exclusion criteria were: underwent solid organ transplantation, received dialysis, had a sibling participating in the study, unable to understand spoken English, had a developmental delay or significant cognitive impairment precluding their ability to complete study procedures, or declined to use the electronic adherence monitoring device or to be audio-recorded during qualitative interviews. Potentially eligible AYAs were mailed or emailed a letter describing the study and providing the opportunity to opt out of recruitment. AYAs who did not opt out were contacted by telephone to provide study details, conduct further eligibility screening, and, if eligible and interested in enrolling, coordinate informed consent/assent procedures. Informed consent/assent procedures were conducted with participants during a telephone call with study staff (AYAs >18 years old provided consent for themselves; for AYAs <18 years old, a primary caregiver provided informed consent for their child and the AYA provided informed assent for themselves).

Phase 1: Reminder+COM-B Message Intervention Development

Message content was developed by the study team (experts on AYA medication adherence, behavior change theory, and pediatric nephrology) to target COM-B model components [19] and incorporate effective public health communication strategies; specifically, framing messages based on what the AYA could gain by taking their medicine as opposed to what they could lose by not taking their medicine, minimizing use of fear appeals, and incorporating novelty and relevance to the intended audience [21-24] (eg, providing different content each day, gearing content toward AYAs with CKD taking antihypertensive medicine, writing content at reading level and style appropriate for age group). Poor framing of health messages can have iatrogenic effects by reducing engagement in the targeted behavior [25]. Messages were written at the ≤5th-grade reading level and contained <140 characters. Messages targeting the COM-B model's capability and motivation components were based on validated self-report measures of AYA adherence barriers and beliefs [26,27]. The COM-B model's opportunity component was targeted with a simple medicine reminder message. Pediatric nephrologists (N=6) at the study site provided feedback on the messages' medical accuracy.

The initial intervention message pool was presented to 10 AYAs with CKD (mean age 16.50, SD 3.41, range 12-21 years) during a semistructured telephone interview (~60 minutes; audio-recorded). AYAs rated messages on acceptability, effectiveness, helpfulness, and comprehension, and provided open-ended suggestions for improving content. AYAs received US $20 each for completing the interviews.

Messages rated by <85% of AYAs as acceptable, effective, helpful, or understood were excluded or revised before inclusion in the final pool tested in Phase 2. Messages were revised based on interview responses and suggestions, including adjusting the message wording for enhanced comprehension and age appropriateness, and adding new content generated by AYAs. The final Reminder+COM-B Message pool included 14 messages targeting adherence capability (eg, "Tip: Put your medicine in a safe place you look each day [like the kitchen counter or by your bed] to help remember to take them"), 14 messages targeting adherence motivation (eg, "Think about your future goals and how being healthier by taking your medicine may help you achieve them!"), and a simple reminder targeting adherence opportunity ("Please remember to take your medicine"). The Reminder+COM-B Message intervention involved sending AYAs a daily message bundle, which included the opportunity message (simple reminder) and a capability or motivation message (alternated each day) at the time(s) when the AYA reportedly took their antihypertensive medicine. All participants in this group received the same Reminder+COM-B Message bundle in the same order (eg, opportunity+capability message bundle on day 1, opportunity+motivation message bundle on day 2). If the participant was prescribed a twice-a-day antihypertensive medication regimen, they received the Reminder+COM-B Message bundle with their first dose and the Reminder-only Message with their second dose.

Reminder-Only Message Active Control Condition

The Reminder-only Message active control condition involved sending AYAs the daily opportunity message only (the same simple reminder). Reminder messages were sent at the time(s) when each AYA reportedly took their antihypertensive medicine. If the participant's antihypertensive medication regimen was twice a day, they received a separate reminder message when each expected dose was due.

Phase 2: Pilot RCT

AYAs were mailed electronic pill bottles for monitoring adherence (see Objective Medication Adherence section). After the bottle was delivered, study staff conducted a telephone call with the AYA to describe how the bottle worked, transfer the monitored antihypertensive medicine into the bottle, and answer AYA or caregiver questions. AYAs were instructed to use the study bottle for their antihypertensive medicine during the 8-week pilot RCT and to transfer any refills to the study bottle during this period. AYAs were informed that the study text messages would be sent to their mobile phone at some point during the pilot RCT but were not provided with a specific date. Text messages (all one-way) containing the Reminder+COM-B Message intervention or Reminder-only Message active control content were sent from a single study number via REDCap's Twilio interface (research team costs included US $1/month for the phone number and US $0.007/outgoing text message; participants used their personal cellular devices and plans). A

brief demographic survey was sent to AYAs ≥18 years old and caregivers of AYAs <18 years old via REDCap. AYAs completed online surveys (Qualtrics) at baseline and after the pilot RCT ended.

The first 4 weeks of the pilot RCT (baseline) evaluated AYA adherence without sending any messages. AYAs were randomized to the Reminder+COM-B Message intervention or Reminder-only Message active control group and received their respective group's messages for the last 4 weeks of the pilot RCT. AYAs were randomized at a 1:1 basis to either group at the end of baseline. A random number generator was used for simple randomization and AYAs were assigned consecutively to a group. Blinding the study team to group assignment was not possible; however, the messages were delivered remotely via text message rather than face to face. At the end of the 8 weeks, AYAs in the Reminder+COM-B Message intervention group completed a qualitative interview (see Qualitative Interviews section). AYAs in both groups received US $40 each for completing the pilot RCT and pre and postsurveys.

Objective Medication Adherence

AdhereTech bottles electronically assessed daily antihypertensive medication adherence via a cellular-connected cap that recorded the date and time when the bottle cap was opened and closed. Timestamps were automatically transferred to AdhereTech's secure online portal. If more than one antihypertensive medication was prescribed, AYAs selected which medicine they wanted to monitor in the AdhereTech bottle. At the end of the monitoring period, the study team debriefed with participants to query for major problems when using the AdhereTech bottle (eg, issues that led them to stop using the device) and whether the monitored medicines were taken from a container other than the AdhereTech bottle during the study period [28]. Daily adherence was coded as whether the bottle was opened that day (1) or not (0). If the AYA took the monitored medicine twice a day, daily adherence was coded as whether the bottle was opened twice that day (1) or not (0). Overall mean adherence was calculated as the number of bottle cap openings recorded divided by the total number of expected openings based on the prescribed regimen and days in the monitoring period.

Surveys of Adherence Capability and Motivation

Adherence Capability

To represent capability as defined in the COM-B model (eg, skills, knowledge, ability to take medicine) [19], the Riekert Self-Efficacy Scale [20] and the Adolescent Medication Barriers Scale (AMBS) [26] were administered.

The Riekert Self-Efficacy Scale (12 items) asked AYAs to rate their ability to take medicine in different situations on a 10-point Likert scale ranging from "not at all sure" to "completely sure" (eg, "How sure are you that you can take your blood pressure medicine the way your doctor said when you want to do something else?"). Ratings were summed and divided by 12 for a scaled score ranging from 1 to 10 (higher scores indicate higher self-efficacy). Internal consistency (Cronbach α) in this study ranged from .94 to .95.

The AMBS (17 items) assessed AYAs' adherence barriers. AYAs rated each item using a 5-point Likert scale ranging from "strongly disagree" to "strongly agree" (eg, "I find it hard to stick to a fixed medication schedule"). The AMBS contains subscales but only the total score was analyzed in this study (ratings were summed and divided by 17 for a scaled score ranging from 1 to 5; higher scores indicate higher adherence barriers). Internal consistency in this study ranged from .81 to .89.

Adherence Motivation

To represent motivation as defined in the COM-B model (conscious, reflective, and learned reasons for taking medicine as prescribed) [19], several scales from the Beliefs About Medication Scale [27,29,30] were administered.

The Positive Outcome Expectancies (POE; 20 items) and Negative Outcome Expectancies (NOE; 13 items) scales assessed AYAs' expectations of favorable/unfavorable outcomes for taking medications as prescribed (eg, "When I take my medicine the way I should, I feel well enough to do things I enjoy" [POE]; "Taking my medicine the way I should makes me miss out on doing fun things" [NOE]). Items are rated on a 7-point Likert scale ranging from "definitely do not agree" to "definitely agree." Item ratings were summed and divided by the number of scale items to obtain scaled scores ranging from 1 to 7 (higher scores indicate more positive expectations [POE] or more negative expectations [NOE]). Internal consistency in this study ranged from .85 to .87 for POE and from .90 to .92 for NOE.

The Adherence Motivation and Importance scales (3 items each) assessed AYAs' perspectives on the importance of taking medicine (eg, "How important do you think it is for you to take your blood pressure medication the way the doctor said when you feel just fine?") and motivation to do so (eg, "How much do you want to take your blood pressure medication the way the doctor said everyday?") on a 10-point Likert scale. Item ratings were summed and divided by 3 to obtain scaled scores (higher scores indicate higher importance or motivation). Internal consistency in this study ranged from .76 to .96 for Importance and from .84 to .96 for Motivation.

Phase 3: Qualitative Interviews

Following Phase 2 completion, AYAs in the Reminder+COM-B Message intervention group were invited to complete audio-recorded and transcribed qualitative interviews (~30 minutes) by telephone. The interviewer (CE) followed an iterative interview guide evaluating AYAs' perceptions of the intervention, including perspectives on message acceptability and engagement (eg, "Did you read all of the messages?" "In what, if any, ways did the messages make you not want to take your medicine?"), mechanisms of behavior change, and suggestions for improving the intervention. Interviews lasted, on average, 31 minutes (SD 11). No repeat interviews were conducted. AYAs received US $20 for their time.

Data Analysis

The primary feasibility, acceptability, and engagement results are reported as proportions and percentages. Descriptive statistics (mean, SD, range) were calculated for primary study

variables by randomization group and measurement time point. For hypothesis testing, statistical significance was assumed when $P<.05$. A linear mixed model (PROC MIXED; SAS 9.4 Software, Cary, NC, USA) was used to examine whether changes in daily adherence over time differed by group allocation, controlling for AYA age, gender, race, and hypertension diagnosis (covariates were based on variables commonly associated with pediatric adherence [4], including AYAs with CKD [20]). Change in daily adherence during both the baseline and postrandomization phases were initially examined to determine the functional form of the time variable (linear or quadratic) and whether to include an individual-level random slope. The best-fitting model for time was selected using the Akaike information criterion (AIC) values [31]. Based on this initial examination, time was modeled as a linear function with an individual-level random slope during baseline (AIC=750.6) and postrandomization (AIC=750.9). When time was modeled as a quadratic function with an individual-level random slope, the AIC value was 765.8 during baseline and during postrandomization. Models were fitted using restricted maximum likelihood estimation. An autoregressive covariance structure was used to account for expected correlations between repeated daily adherence assessments within participants. There were no investigator-identified or participant-reported technical problems when transmitting data. In instances when a participant reported taking their antihypertensive medicine from a different container than the electronic pill bottle with specific dates (n=3), these data were imputed; thus, missed doses were assumed to be medication nonadherence and not missing data.

Repeated-measures analysis of variance (IBM SPSS Statistics Version 26) with one within-subject factor (time point) and one between-subject factor (treatment group) was used to examine whether the change in overall mean adherence during baseline and postrandomization and pre-post mean survey scores differed by group (controlling for AYA age, gender, race, and hypertension diagnosis [4,20]). The magnitude of within-treatment group change between study time points was examined with Cohen *d*.

All qualitative interview transcripts were content-analyzed to identify major themes. The codebook was developed by the principal investigator and study team. Each interview was coded separately by two coders (CE and MC) in nVIVO Pro v.11 (QSR International). The coders identified, discussed, and resolved coding discrepancies in each transcript. The study authors reviewed final codes to identify the number of AYAs reporting that they read the messages, noticed the messages were sent and delivered (including AYAs who reportedly did not read the message content), and perceived the Reminder+COM-B Message intervention to reduce their desire to take their medicine, as well as AYA perceptions of the intervention's mechanisms of behavior change and suggestions for improving it.

Results

Participants

Figure 1 shows the Consolidated Standards of Reporting Trials (CONSORT) diagram for screening, enrollment, and overall study participation. Of the potentially eligible AYAs, 34 declined to enroll. Of the 44 AYAs who were screened as eligible and initially expressed verbal interest in enrolling, 35 AYAs completed informed consent/assent procedures to enroll. There were no significant differences in age, gender, or race between those who enrolled and those who declined to enroll.

Of the 35 enrolled AYAs, one lost contact with the study team after enrolling and did not complete any further procedures. Hence, 34 AYAs (mean age 16.59, SD 3.26, range 11-21 years) were randomized in Phase 2 (Reminder+COM-B Message intervention group n=18, Reminder-only Message active control group n=16). The final sample was primarily male, White, without a hypertension diagnosis, and taking an angiotensin-converting enzyme inhibitor once a day (see Table 1 for detailed demographic characteristics). There were no statistically significant differences in demographic characteristics by randomization group allocation.

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) flow diagram of screening, recruitment, and enrollment of potentially eligible participants and study procedure completion for enrolled adolescents and young adults (AYA). CKD: chronic kidney disease.

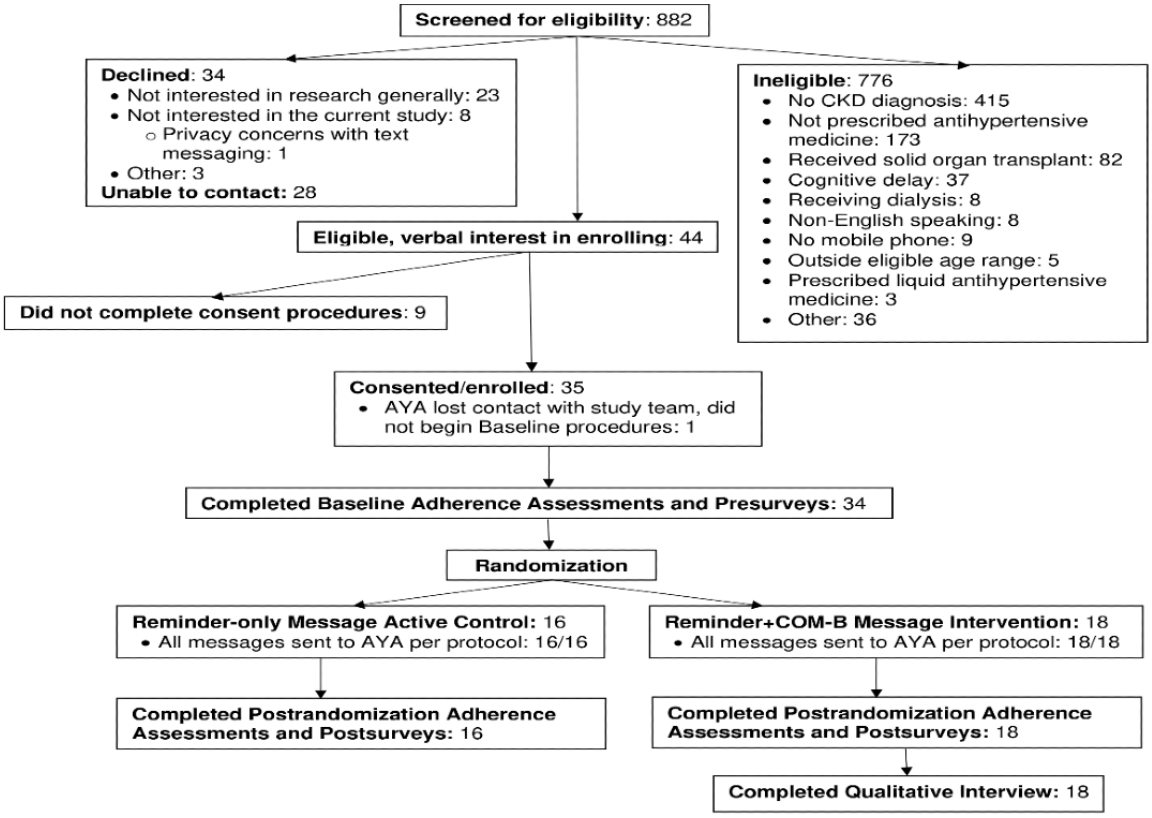


Table 1. Demographic characteristics of participants.

Variable	Overall Sample (N=34)	Reminder-only Message Active Control (n=16)	Reminder+COM-B Message Intervention (n=18)
Age (years), mean (SD)	16.59 (3.26) (range 11-21)	16.13 (3.44)	17.00 (3.22)
Gender n (%)			
Female	14 (41)	7 (44)	7 (39)
Male	20 (59)	9 (56)	11 (61)
Race/ethnicity, n (%)			
White	19 (56)	8 (50)	11 (61)
African American/Black	13 (38)	7 (44)	6 (33)
Other	2 (6)	1 (6)	1 (6)
Hypertension diagnosis, n (%)			
Yes	12 (35)	5 (31)	7 (39)
No	22 (65)	11 (69)	11 (61)
Hypertension medication, n (%)			
ACE[a] inhibitor	28 (82)	14 (87)	14 (78)
Other[b]	6 (18)	2 (13)	4 (22)
Dose frequency, n (%)			
Once a day	32 (94)	15 (94)	17 (94)
Twice a day	2 (6)	1 (6)	1 (6)

[a]ACE: angiotensin-converting enzyme.

[b]"Other" hypertension medicines included calcium channel blocker or angiotensin II receptor blocker.

Feasibility, Acceptability, and Engagement

Only 9 AYAs were ineligible to be enrolled in the study due to not having a mobile phone. No AYAs were ineligible due to having a mobile phone that did not receive text messages. No AYA declined to participate due to concerns about the expense of receiving study text messages on their mobile phone. All AYAs in the final sample (34/34, 100%) completed the Phase 2 pilot RCT procedures and all AYAs in the Reminder+COM-B Message intervention group (18/18, 100%) completed a Phase 3 qualitative interview. The decline rate for study enrollment was relatively high (34/106, 32.1%), although the primary reason for declining was lack of interest in research participation generally (23/34, 68%). For AYAs who declined for reasons related to this study, the primary reason was discomfort discussing topics related to having a medical condition during qualitative interviews (n=3); one person cited concerns with privacy related to text messaging and one person reported that they did not want their adherence to be monitored (the concern was specific to adherence monitoring, not the monitoring device itself). However, retention was high for AYAs who enrolled and began the pilot RCT in Phase 2 (34/34, 100%).

All AYAs (34/34, 100%) reportedly used the electronic pill bottles without indicating major problems that led them to stop using the device. The research team did not identify any technical issues with data transmission from the electronic pill bottles to the data collection portal. The text messaging plan cost the research team up to US $0.59 per AYA for 1 month plus US $1.00/month for the dedicated phone number used to send text messages to all participants.

No AYAs in either group (0/34) asked for the messages to stop or reported that the messages bothered them. The majority of AYAs in the Reminder+COM-B Message intervention group reportedly read the messages sent to them (15/18, 83%) and all AYAs in this group (18/18, 100%) reportedly noticed that the messages were sent and received (even if they did not read the content). No AYAs in the Reminder+COM-B Message intervention group (0/18) reported that the messages reduced their desire to take their medicine.

Changes in Daily Adherence by Group Over Time During the Phase 2 Pilot RCT

Detailed linear mixed model results are shown in Table 2. In the final models accounting for covariates, daily adherence did not significantly change over time during baseline and there was no significant group by time interaction. As no significant group by time interaction was observed for adherence slopes during baseline, the decision was made to report changes in daily adherence separately during baseline and postrandomization.

Postrandomization, daily adherence significantly decreased over time in the overall sample but there was a significant group by time interaction. Specifically, the Reminder-only Message active control group showed a higher initial response to receiving the text messages and then a steeper decline in daily adherence over the postrandomization phase compared to the Reminder+COM-B Message intervention group's daily adherence, which remained relatively stable with some improvement over time (see Figure 2).

Table 2. Changes in adolescent/young adults' daily objective medication adherence by randomization group over time.

Effect	Baseline phase (4 weeks)			Postrandomization phase (4 weeks)		
	Estimate[a] (SE)	95% CI/z[b]	*P* value	Estimate (SE)	95% CI/z[b]	*P* value
Fixed effects						
Intercept	.91 (.08)	.75- 1.06	<.001	.95 (.10)	.75-1.15	<.001
Time (Day)	–.003 (.003)	-.01-.003	.27	–.01 (.003)	–.01-.001	.02
Group[c]	.03 (.08)	–.12-.18	.71	–.13 (.10)	–.33-.06	.18
Age	.002 (.01)	–.02-.02	.86	–.01 (.01)	–.04-.02	.47
Gender[d]	–.07 (.07)	–.21-.07	.35	.005 (.09)	–.18-.19	.96
Hypertension diagnosis[e]	–.10 (.07)	–.25-.05	.19	–.08 (.10)	–.27-.11	.42
Race[f]	–.003 (.07)	–.14-.14	.97	–.11 (.09)	–.30-.07	.23
Group × Time	–.002 (.004)	–.01-.01	.65	.01 (.004)	.0004-.02	.04
Random effects						
Intercept	.03 (.01)	2.60	.005	.06 (.02)	3.05	.001
Random slope	<.001 (<.001)	2.29	.01	<.001 (<.001)	2.26	.01

[a]Estimate is beta for fixed-effects models and is variance for random-effects models.

[b]95% CI for fixed-effects variables and *z* for random-effects variables.

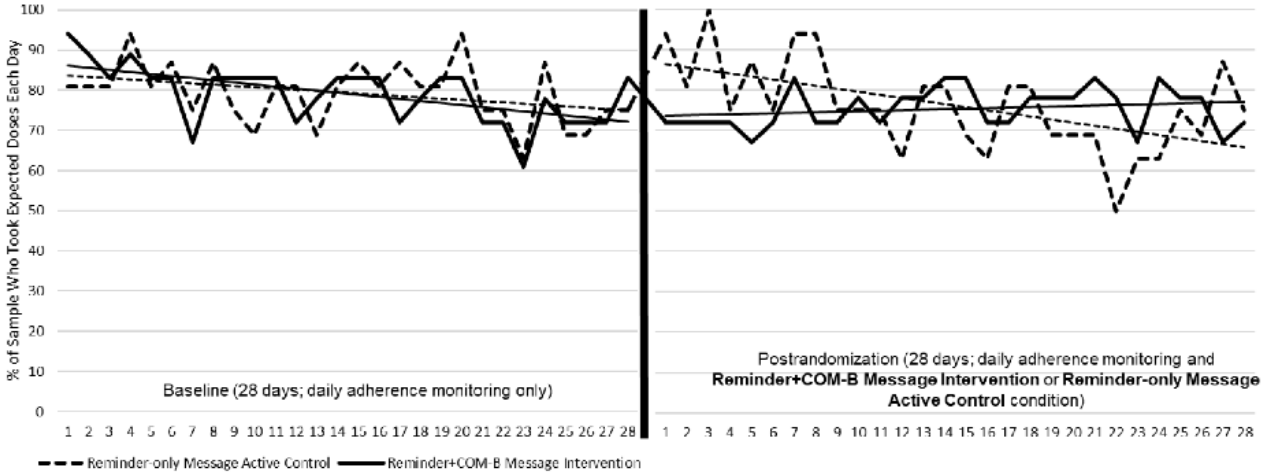
[c]Randomization group was coded as Reminder-only Message active control=0, Reminder+COM-B Message intervention=1.

[d]Gender was coded as male=0, female=1.

[e]Hypertension diagnosis was coded as no hypertension=0, hypertension=1.

[f]Race was coded as White=0, non-White=1.

Figure 2. Daily objective medication adherence by randomization group during the baseline and postrandomization phases. The y-axis refers to the percentage of the sample who were recorded as having taken their medicine dose(s) each day.



Posthoc Sensitivity Analysis for Baseline Adherence

A posthoc sensitivity analysis was conducted to exclude the first 7 days of adherence data due to potential Hawthorne effects of knowing one's adherence was being electronically monitored. Results were similar whether the first 7 days of baseline data were excluded or included (adherence demonstrated a nonsignificant decline over time, with no significant group by time interaction).

Changes in Overall Mean Adherence and AYA Pre-Post Surveys During the Pilot RCT

During baseline, overall mean adherence was 80.60% (SD 22.47) in the Reminder-only Message active control group and was 79.46% (SD 22.62) in the Reminder+COM-B Message intervention group. During postrandomization, overall mean adherence was 76.76% (SD 27.31) in the Reminder-only Message active control group and 76.00% (SD 27.00) in the Reminder+COM-B Message intervention group. There were no significant differences between treatment groups in changes in overall mean adherence by measurement time point (no group by time interactions) or in mean survey scores. The magnitudes of within-group mean changes between time points were small (all Cohen d<.25). Table 3 shows the mean scores for survey responses by group allocation and measurement time point.

Posthoc Sensitivity Analysis for Mean Baseline Adherence

A posthoc sensitivity analysis was conducted to exclude the first 7 days of adherence data from the baseline mean scores due to potential Hawthorne effects of knowing that one's adherence was being electronically monitored. Results were similar whether the first 7 days of baseline adherence data were excluded or included (no significant group by time interactions).

Table 3. Descriptive data for mean medication adherence and adolescent/young adult survey scores by group allocation and time point.

Variable	Baseline		Postrandomization		*P* value[a]	Cohen *d*[b]
	mean (SD)	Range	mean (SD)	Range		
Overall medication adherence					.97	
Active control[c]	80.60 (22.47)	32.14-100	76.76 (27.31)	21.43-100		.15
Intervention[d]	79.46 (22.62)	35.71-100	76.00 (27.00)	7.14-100		.14
Medication self-efficacy					.57	
Active control	7.90 (1.74)	4.08-9.75	8.03 (1.88)	3.83- 9.83		–.07
Intervention	7.95 (1.95)	3.33- 9.67	7.83 (2.36)	2.92- 10.00		.06
Positive outcome expectancies					.42	
Active control	5.81 (0.56)	5.05- 6.75	5.85 (0.40)	5.30- 6.80		–.08
Intervention	5.70 (1.00)	4.00-7.00	5.94 (0.90)	3.85-7.00		–.25
Negative outcome expectancies					.61	
Active control	2.07 (1.12)	1.00-4.31	2.19 (1.24)	1.00-5.31		–.10
Intervention	1.72 (0.94)	1.00-4.62	1.91 (1.09)	1.00-5.46		–.19
Barriers to adherence					.79	
Active control	2.23 (0.55)	1.35-3.41	2.25 (0.69)	1.12-3.65		–.03
Intervention	1.94 (0.56)	1.18-2.94	1.92 (0.66)	1.00-3.35		.03
Motivation for adherence					.46	
Active control	7.88 (2.41)	2.00-10.00	8.44 (2.92)	1.00-10.00		–.21
Intervention	8.30 (2.44)	1.00-10.00	7.87 (2.71)	3.33-10.00		.17
Importance of adherence					.18	
Active control	9.25 (1.72)	3.33-10.00	8.81 (1.85)	3.33-10.00		.25
Intervention	9.24 (1.58)	4.00-10.00	9.19 (1.98)	3.33-10.00		.03

[a]*P* values are for group by time interactions for changes in mean adherence or surveys.

[b]Cohen *d* reflects the magnitude of change within randomization group from baseline to postrandomization.

[c]Reminder-only Message active control group.

[d]Reminder+COM-B Message intervention group.

Qualitative Interviews with AYAs in the Reminder+COM-B Message Intervention Group

Mechanisms of Adherence Behavior Change

The majority of AYAs tended to view all Reminder+COM-B Message bundles irrespective of content as cues to take medicine: "When I got the text, I would look at my phone and then remembered that it was time to take the medicine" (14-year-old male). However, a 17-year-old male pointed out that the different content helped him attend to the messages: "Because [the messages] are different every time, I thought, 'I'm going to read this and then I'm going to take my medicine.' So it did help that they were different."

Occasionally, AYAs recalled messages that provided helpful content regarding capability (skills) or motivation to maintain adherence. For example, an 18-year-old female discussed how the messages enhanced her adherence motivation: "I think [the messages] helped me by reading them and then thinking over that it is really important for me to take [my medicine] and it does motivate me to take it." Other key quotes are shown in Textbox 1.

Textbox 1. Additional qualitative adolescent/young adults' feedback about the Reminder+COM-B Message intervention.

Theme 1: Mechanisms of behavior change

- Subtheme: Reminder+COM-B Message intervention bundles are cues to take medicine irrespective of content

  When the messages were sent, it was a reminder to take my medicine. I looked at it and said, "I've got to go take my medicine." [16-year-old male]

  [The messages] gave me a reminder, like, "Oh, I need to take my medicine." And I would usually do it after I got the text message. [19-year-old female]

[The messages are] sort of like a reminder on your phone. [15-year-old male]

- Subtheme: Message content facilitated behavior changes to support adherence

There was one [message] about planning ahead for me not being at home. That helped me a lot...because on the weekends and Fridays, I'm away from home. I made sure I set another reminder for the time when I knew I'd get home. That way, I'd get a reminder fairly quick that I should take my medicine. [14-year-old male]

When I got that message [about refilling the prescription], it actually made me do a double take and look and see how much medicine I had and if I needed to get a refill or not. [18-year-old male]

Theme 2: Suggestions to improve intervention acceptability

- Subthemes: Adjust Reminder+COM-B Message intensity, timing, and content to match adherence behavior, barriers, and adolescent/young adult availability, receptivity, and engagement

Gauge how frequent to send the messages based on how consistent the person is with taking the medicine. [18-year-old male]

I think that you should send [the messages] twice because sometimes people don't get them. [14-year-old female]

At first, I thought, "These are useful." And then because it happened every day, I started getting a little bit annoyed - but [the messages] did help me on the weekends [when] I forget. [17-year-old male]

I forget to check my phone some of the time. When we would go to a friend's house, I wouldn't get [my phone] or bring it. And when we got home, I would be tired. Maybe we could set [the messages] at a different time or set more than one reminder to [take my medicine]. [11-year-old female]

[Send messages] at a time when the person is up and able to look at their phone. [18-year-old female]

AYA Suggestions for Improving the Reminder+COM-B Message Intervention Acceptability

Tailoring message intensity, timing, and content to correspond to actual adherence behavior, current adherence barriers, or AYA availability, receptivity, and engagement could improve the intervention's acceptability. A 21-year-old male suggested,

> *If you see a trend of a person taking their medicine spot-on, lessen the amount of messages. If you see that they're not taking the medicine, make the reminders more repetitive and more prominent.*

Sending too many messages could lead to disengagement: "I wasn't really paying attention. I get like 1000 messages a day" (11-year-old male). Providing positive reinforcement for taking medications as prescribed could also improve intervention acceptability: "I think if they did take it, they should [get a message saying], 'Good job taking your medicine and don't forget next time.' Because I think people will feel good inside" (14-year-old female). Message content should ideally match the AYA's current situation and potential barriers to adherence: "Maybe if you're running low, that message [about getting a refill] pops up, but if your pill bottle is completely full, that message doesn't really help you" (21-year-old male). Other key quotes appear in Textbox 1.

Discussion

Principal Findings

The current study preliminarily investigated a new mHealth messaging intervention based on the COM-B model [19] that aimed to promote objectively measured antihypertensive medication adherence in AYAs with CKD and obtained qualitative stakeholder feedback on user experiences. The Reminder+COM-B Message intervention was feasible, acceptable, and subjectively engaging to participants. Our pilot RCT results suggest that this intervention had stabilizing effects on daily adherence compared to a Reminder-only Message active control group. Qualitative interviews provided insight into the Reminder+COM-B Message intervention's mechanisms of behavior change and avenues for improving its acceptability.

Our feasibility, acceptability, and engagement hypotheses were largely supported. Specifically, few AYAs were ineligible due to not owning a mobile device, we did not encounter major technical or AYA-reported problems with our electronic pill bottles or our low-cost text messaging approach, and our retention rate for enrolled participants was high. The majority of AYAs in the Reminder+COM-B Message intervention group reportedly read the messages or acknowledged receiving and noticing the messages (even if they did not read the content). Further, no AYAs in the Reminder+COM-B Message intervention group reported iatrogenic effects of the messages reducing their desire to take medicine. These findings suggest that using an electronic pill bottle to monitor adherence while receiving daily text messages with content designed to promote adherence is a feasible and acceptable intervention approach for AYAs with CKD taking daily antihypertensive medicine. Our qualitative results highlighted ways to improve message engagement, which may help enhance content interest for AYAs who reportedly did not read the majority of the Reminder+COM-B messages. Future researchers may consider adapting our approach to designing and testing messaging-based mHealth interventions for AYA medication adherence and incorporating objective engagement indices to bolster our subjective engagement findings from the current study.

Our Phase 2 pilot results showed that our primary outcome, daily postrandomization adherence, declined at a significantly faster rate in the Reminder-only Message active control group compared to the Reminder+COM-B Message intervention group.

No significant group by time differences were observed for our secondary outcome, overall mean adherence, which likely reflected our sample size and statistical power for these analyses. Longitudinal research indicates that antihypertensive medication adherence is expected to decrease over time in AYAs with CKD [20], which was similarly observed in the Reminder-only Message active control group over the 4-week postrandomization phase. In contrast, the Reminder+COM-B Message intervention, which involved sending a bundled daily reminder message with an alternating capability or motivation message, appeared to stabilize daily adherence with some gradual improvements observed. Of note, during the first week of the postrandomization phase, the Reminder-only Message active control group demonstrated higher adherence than their baseline adherence and the Reminder-COM-B Message intervention group, which likely contributed to the significant group by time interaction observed during this monitoring period. It is possible that the Reminder-only Message active control group's decline in adherence represented regression to their earlier baseline. Given that this was a pilot study, results of our pilot RCT should be interpreted with caution and understood as an important step in refining our new intervention and informing whether further testing is warranted, rather than evidence of its efficacy [32-34]. Our feasibility, acceptability, and preliminary Phase 2 results suggest that more rigorous future investigations of our Reminder+COM-B Message intervention could be a productive next step in developing efficacious mHealth approaches for improving AYA adherence.

There are several hypothesized reasons for why the Reminder+COM-B Message intervention group may have shown stable and gradually improved adherence during postrandomization. AYA qualitative interview results suggest the different daily Reminder+COM-B Message content may have been interesting to AYAs and helped them attend to messages as a cue to take their medicine. It is possible that the higher intensity of messaging compared to the Reminder-only Message active control group enhanced the impact on adherence. Further investigation with larger samples is needed to clarify the Reminder+COM-B Message intervention's mechanistic effects on daily adherence in contrast to a daily simple reminder.

Unexpectedly, pre-post changes in AYAs' perceptions of their adherence motivation and capability (survey responses) did not differ significantly by group and within-group changes were small in magnitude. The sample size was small, which limited statistical power to detect effects in these analyses. During qualitative interviews, some AYAs discussed how the Reminder+COM-B messages led to tangible behavioral and motivational changes that enhanced their adherence, but most viewed the message bundles as adherence cues. Considering the COM-B model upon which this intervention was designed [19], an alternative hypothesis is that the mechanism of change in daily adherence may be linked to improving *opportunity* (cue) to take medicine rather than changing AYAs' perceptions of their adherence *capability* or *motivation*. This hypothesis is further supported by qualitative interviews suggesting that the Reminder+COM-B messages reminded AYAs to take their medicine regardless of message content. Further investigation is needed to elucidate how AYA perceptions of adherence capability and motivation can be modified via mHealth approaches to facilitate adherence behavior change.

In qualitative interviews about the Reminder+COM-B Message intervention, AYAs discussed suggested improvements to tailor message intensity, content, and timing to match adherence behavior and barriers, as well as AYA availability, receptivity, and engagement in the intervention. These major themes focused on behavioral components of the intervention and adherence and are suggestive of a just-in-time adaptive intervention (JITAI), which provides in-the-moment intervention exactly when AYAs need it most [35]. Adapting message intensity to match adherence (eg, send messages when medicine is missed) may enhance AYAs' engagement by increasing message novelty, while more clearly demonstrating the link between intervention delivery and adherence behavior. Moving toward an adaptive format for message timing (eg, send messages when an AYA is available to read them) and content (eg, match message content to current barriers that could lead to nonadherence) may increase AYA receptivity to the intervention and enhance the likelihood that attending to message content will lead to adherence or implementing strategies to prevent nonadherence. Providing praise when medicine is taken may positively reinforce adherence behavior and enhance the probability that AYAs will take subsequent doses. These hypotheses await further testing in a modified version of our mHealth messaging intervention built within a JITAI framework.

Limitations

The Reminder+COM-B messages were bundled and sent at the same frequency and in the same order to all AYAs randomized to this condition; hence, it is unknown how individual components (eg, motivation messages only), different message intensities (sending bundled vs single messages), or ordering may have differential effects on adherence. Future investigations that adhere to the Multiphase Optimization Strategy framework [36] to systemically and iteratively conduct optimization trials using novel experimental designs (eg, microrandomized trials [37]) may help identify specific intervention components delivered at particular times and intensities that maximize intervention acceptability and efficacy for an individual and form the basis for decision rules applied in a JITAI-based version of our intervention.

Our technology was limited in that we were unable to objectively evaluate AYA engagement by confirming if text messages were read. A key future direction is to incorporate objective engagement measures in evaluating our intervention approach. The sample size was small, which limited statistical power. However, we rigorously measured daily adherence data over 8 weeks, which yielded a larger number of data points for our analyses involving changes in daily adherence over time. AYAs were recruited from a single site and we monitored one prescribed antihypertensive medication per AYA, which may limit generalizability of the findings. AYAs were only offered electronic pill bottles, although some people use pillboxes to manage multiple medications. Researchers should consider offering the option of electronic pill bottles or pillboxes as a strategy for improving the acceptability and sustainability of objectively measuring daily adherence. Although we probed

for major problems using the electronic pill bottles after Phase 2 and identified no issues in transmitting data from the bottles to our data collection portal, we were unable to personally observe every participant at each expected medication dose to obtain additional verification of adherence behavior. We included AYAs irrespective of their baseline adherence levels, which may have limited opportunities to observe adherence improvements. Further, we did not assess the impact of removing text messages on adherence to see if there was a lasting protective effect of text messaging on adherence, which should be examined in future studies.

We included AYAs within a larger age range; a more limited age range could provide greater insight into the Reminder+COM-B Message intervention's acceptability within specific developmental phases (eg, young adulthood). A higher number of AYAs declined to enroll in the study due to lack of interest in participating. Although AYA enrollment challenges are common in behavioral intervention trials, our approach to introducing the study may benefit from adjustments to improve acceptability and engagement. Specifically, recruitment may benefit from deeper discussion of the altruistic reasons for study participation to enhance AYA interest [38], given that the primary reason for declining participation was lack of interest in research in general. Only AYAs randomized to the Reminder+COM-B Message intervention group completed qualitative interviews; future investigators should obtain AYA perspectives on receiving daily reminders only, as in the Reminder-only Message active control group (eg, to probe whether declines in adherence reflected fatigue from receiving the same messages). Additionally, the 8-week monitoring period limited the ability to examine associations between group allocation and clinical outcomes (eg, changes in estimated glomerular filtration rate), which should be evaluated in future studies designed to follow AYAs for longer periods of time.

Conclusions

Our theoretically informed Reminder+COM-B mHealth messaging intervention appears to be feasible, acceptable, and promising for promoting objectively measured antihypertensive medication adherence in AYAs with CKD. Future research using similarly rigorous adherence outcome measures is needed to test refined versions of this intervention that incorporate AYA feedback and use study designs aimed at determining the most efficacious intervention components (eg, microrandomized trials [37]) to maximize the positive impact on AYAs' medication adherence.

Acknowledgments

This study was supported by the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health Ruth L. Kirschstein Postdoctoral Individual National Research Service Award (F32DK118988) and the Society of Pediatric Psychology/American Psychological Association Division 54 Drotar-Crawford Postdoctoral Fellowship Research Grant (both awarded to CE).The authors thank Tumasi Agyapong, Eric Fei, Merry Qian, You Wang, and Marin Langlieb for their contributions to the conduct of this investigation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1850 KB - jmir_v22i8e19861_app1.pdf]

References

1. Fathallah-Shaykh SA. Proteinuria and progression of pediatric chronic kidney disease: lessons from recent clinical studies. Pediatr Nephrol 2017 May;32(5):743-751. [doi: 10.1007/s00467-016-3448-8] [Medline: 27350622]
2. Wühl E, Schaefer F. Therapeutic strategies to slow chronic kidney disease progression. Pediatr Nephrol 2008 May;23(5):705-716 [FREE Full text] [doi: 10.1007/s00467-008-0789-y] [Medline: 18335252]
3. Pruette CS, Coburn SS, Eaton CK, Brady TM, Tuchman S, Mendley S, et al. Does a multimethod approach improve identification of medication nonadherence in adolescents with chronic kidney disease? Pediatr Nephrol 2019 Jan;34(1):97-105 [FREE Full text] [doi: 10.1007/s00467-018-4044-x] [Medline: 30116892]
4. Hommel K, Ramsey R, Rich K, Ryan J. Adherence to pediatric treatment regimens. In: Roberts MC, Steele RG, editors. Handbook of pediatric psychology, 5th edition. New York: Guilford Press; 2017:119-133.
5. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014 Nov 20(11):CD000011 [FREE Full text] [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
6. Teens, social media & technology 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2020-08-09]
7. Lee Y, Cui Y, Tu M, Chen Y, Chang P. Mobile Health to Maintain Continuity of Patient-Centered Care for Chronic Kidney Disease: Content Analysis of Apps. JMIR Mhealth Uhealth 2018 Apr 20;6(4):e10173 [FREE Full text] [doi: 10.2196/10173] [Medline: 29678805]

8. Singh K, Diamantidis CJ, Ramani S, Bhavsar NA, Mara P, Warner J, et al. Patients' and Nephrologists' Evaluation of Patient-Facing Smartphone Apps for CKD. Clin J Am Soc Nephrol 2019 Apr 05;14(4):523-529 [FREE Full text] [doi: 10.2215/CJN.10370818] [Medline: 30898873]
9. Nguyen E, Bugno L, Kandah C, Plevinsky J, Poulopoulos N, Wojtowicz A, et al. Is There a Good App for That? Evaluating m-Health Apps for Strategies That Promote Pediatric Medication Adherence. Telemed J E Health 2016 Nov;22(11):929-937. [doi: 10.1089/tmj.2015.0211] [Medline: 27070837]
10. Kosse RC, Bouvy ML, de Vries TW, Koster ES. Effect of a mHealth intervention on adherence in adolescents with asthma: A randomized controlled trial. Respir Med 2019 Mar;149:45-51. [doi: 10.1016/j.rmed.2019.02.009] [Medline: 30803885]
11. Ramsey RR, Holbein CE, Powers SW, Hershey AD, Kabbouche MA, O'Brien HL, et al. A pilot investigation of a mobile phone application and progressive reminder system to improve adherence to daily prevention treatment in adolescents and young adults with migraine. Cephalalgia 2018 Dec;38(14):2035-2044. [doi: 10.1177/0333102418756864] [Medline: 29635937]
12. Fenerty SD, West C, Davis SA, Kaplan SG, Feldman SR. The effect of reminder systems on patients' adherence to treatment. Patient Prefer Adherence 2012;6:127-135. [doi: 10.2147/PPA.S26314] [Medline: 22379363]
13. Vervloet M, Linn AJ, van Weert JCM, de Bakker DH, Bouvy ML, van Dijk L. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature. J Am Med Inform Assoc 2012;19(5):696-704 [FREE Full text] [doi: 10.1136/amiajnl-2011-000748] [Medline: 22534082]
14. Anderson LM, Leonard S, Jonassaint J, Lunyera J, Bonner M, Shah N. Mobile health intervention for youth with sickle cell disease: Impact on adherence, disease knowledge, and quality of life. Pediatr Blood Cancer 2018 Aug;65(8):e27081. [doi: 10.1002/pbc.27081] [Medline: 29693797]
15. Creary S, Chisolm D, Stanek J, Hankins J, O'Brien SH. A Multidimensional Electronic Hydroxyurea Adherence Intervention for Children With Sickle Cell Disease: Single-Arm Before-After Study. JMIR Mhealth Uhealth 2019 Aug 08;7(8):e13452 [FREE Full text] [doi: 10.2196/13452] [Medline: 31397291]
16. Zhang S, Hamburger E, Kahanda S, Lyttle M, Williams R, Jaser SS. Engagement with a Text-Messaging Intervention Improves Adherence in Adolescents with Type 1 Diabetes: Brief Report. Diabetes Technol Ther 2018 May;20(5):386-389 [FREE Full text] [doi: 10.1089/dia.2018.0015] [Medline: 29792749]
17. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text Messaging and Mobile Phone Apps as Interventions to Improve Adherence in Adolescents With Chronic Health Conditions: A Systematic Review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
18. Badawy SM, Thompson AA, Kuhns LM. Medication Adherence and Technology-Based Interventions for Adolescents With Chronic Health Conditions: A Few Key Considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
19. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
20. Eaton CK, Eakin MN, Coburn S, Pruette CS, Brady TM, Fivush BA, et al. Patient Health Beliefs and Characteristics Predict Longitudinal Antihypertensive Medication Adherence in Adolescents With CKD. J Pediatr Psychol 2019 Jan 01;44(1):40-51 [FREE Full text] [doi: 10.1093/jpepsy/jsy073] [Medline: 30252092]
21. Latimer AE, Rench TA, Rivers SE, Katulak NA, Materese SA, Cadmus L, et al. Promoting participation in physical activity using framed messages: an application of prospect theory. Br J Health Psychol 2008 Nov;13(Pt 4):659-681. [doi: 10.1348/135910707X246186] [Medline: 17925060]
22. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]
23. Noar SM. A 10-year retrospective of research in health mass media campaigns: where do we go from here? J Health Commun 2006;11(1):21-42. [doi: 10.1080/10810730500461059] [Medline: 16546917]
24. Zhao X, Villagran MM, Kreps GL, McHorney C. Gain versus loss framing in adherence-promoting communication targeting patients with chronic diseases: the moderating effect of individual time perspective. Health Commun 2012;27(1):75-85. [doi: 10.1080/10410236.2011.569002] [Medline: 21745039]
25. Cho H, Salmon C. Unintended effects of health communication campaigns. J Commun 2007 Jun;57(2):293-317. [doi: 10.1111/j.1460-2466.2007.00344.x]
26. Simons LE, Blount RL. Identifying barriers to medication adherence in adolescent transplant recipients. J Pediatr Psychol 2007 Aug;32(7):831-844. [doi: 10.1093/jpepsy/jsm030] [Medline: 17522111]
27. Riekert K, Drotar D. The Beliefs About Medications Scale: Development, reliability, and validity. J Clin Psychol Med Settings 2002;9:177-184. [doi: 10.1023/A:1014900328444]
28. Riekert K, Rand C. Electronic monitoring of medication adherence: When is high-tech best? J Clin Psychol Med S. J Clin Psychol Med S 2002;9:25-34. [doi: 10.1023/A:1014131928789]
29. Riekert K. Health beliefs and medication adherence among adolescents with chronic health conditions. Doctoral dissertation, Case Western Reserve University, Cleveland, OH. 2000. URL: https://search.proquest.com/docview/304587464 [accessed 2020-09-09]

30. Eakin MN, Chung S, Hoehn J, Borrelli B, Rand-Giovannetti D, Riekert KA. Development and validation of CF-Medication Beliefs Questionnaire: A mixed-methods approach. J Cyst Fibros 2017 Sep;16(5):637-644 [FREE Full text] [doi: 10.1016/j.jcf.2017.05.001] [Medline: 28549609]
31. Hilbe J. Negative binomial regression. 2nd ed. Cambridge, UK: Cambridge University Press; 2011.
32. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
33. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
34. Kistin C, Silverstein M. Pilot Studies: A Critical but Potentially Misused Component of Interventional Research. JAMA 2015 Oct 20;314(15):1561-1562 [FREE Full text] [doi: 10.1001/jama.2015.10962] [Medline: 26501530]
35. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
36. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
37. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
38. Hendricks-Ferguson VL, Cherven BO, Burns DS, Docherty SL, Phillips-Salimi CR, Roll L, et al. Recruitment strategies and rates of a multi-site behavioral intervention for adolescents and young adults with cancer. J Pediatr Health Care 2013;27(6):434-442 [FREE Full text] [doi: 10.1016/j.pedhc.2012.04.010] [Medline: 22658379]

Abbreviations

AIC: Akaike information criterion
AMBS: Adolescent Medication Barriers Scale
AYA: adolescent/young adult
CKD: chronic kidney disease
JITAI: just-in-time adaptive intervention
mHealth: mobile health
NOE: Negative Outcome Expectancies
POE: Positive Outcome Expectancies
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 18.05.20; peer-reviewed by A Psihogios, C Kenyon, JK Low; comments to author 09.06.20; revised version received 10.07.20; accepted 26.07.20; published 14.08.20.

Please cite as:
Eaton C, Comer M, Pruette C, Psoter K, Riekert K
Text Messaging Adherence Intervention for Adolescents and Young Adults with Chronic Kidney Disease: Pilot Randomized Controlled Trial and Stakeholder Interviews
J Med Internet Res 2020;22(8):e19861
URL: http://www.jmir.org/2020/8/e19861/
doi:10.2196/19861
PMID:32795983

Original Paper

Adherence of Female Health Care Workers to the Use a Web-Based Tool for Improving and Modifying Lifestyle: Prospective Target Group Pilot Study

Tomislav Jukic[1], PhD; Alojz Ihan[2], PhD; Marija Petek Šter[3], PhD; Vojko Strojnik[4], PhD; David Stubljar[2], BSc; Andrej Starc[5], PhD

[1]Faculty of Medicine, Josip Juraj Strossmayer, Osijek, Croatia
[2]Institute of Microbiology and Immunology, Ljubljana, Slovenia
[3]Department of Family Medicine, Faculty of Medicine Ljubljana, Ljubljana, Slovenia
[4]Faculty of Sport Ljubljana, Ljubljana, Slovenia
[5]Faculty of Health Sciences of Ljubljana, Ljubljana, Slovenia

Corresponding Author:
David Stubljar, BSc
Institute of Microbiology and Immunology
Ljubljana
Slovenia
Phone: 386 40842593
Email: d.stubljar@gmail.com

Abstract

Background: Health care professionals are exposed to the psychological and physiological effects of stress, which is a well-known risk factor for various mental and physical health problems.

Objective: The aims of this study were to assess the adherence of female health care workers to use a web-based tool for improving and modifying lifestyle and to identify the potential factors influencing their adherence.

Methods: A prospective, observational study was performed. A total of 80 female health care workers (physicians and gradated nurses) from 2 university medical centers and female members of a family medicine society participated. Participants completed a questionnaire that inquired about their basic demographic data and physical fitness. Physical fitness was assessed by the Rockport Fitness Walking Test. Adherence to a web-based application (24@life) was followed for 3 months and the number of log-ins into the application was counted.

Results: The study was conducted from March to October 2019. Significantly high workload has been detected in all groups (P<.05), except in the general practitioner with normal workload group. The graduated nurse working in the surgery room group showed chronic stress with elevated S-cortisol levels (>690 nmol/L); activated cellular immune system with elevated concentrations of lymphocytes (reference 1.1-2.5 × 10^9 cells/L), CD3 cells (reference 0.7-1.9 × 10^9 cells/L), CD8 cells (reference 0.2-0.7 × 10^9 cells/L), and HLA-DR/CD3 cells (reference 0.04-0.2 × 10^9 cells/L); and the worst quality of sleep (mean 2.8 [SD 1.2]). Only 32 of 80 participants (40%) were adherent to the web-based application. Participants most frequently viewed web pages on areas of physical activity (497 times) and nutrition (332 times). No factors or participant's characteristics such as weight (odds ratio [OR] 1.026, 95% CI 0.977-1.078), BMI (OR 0.993, 95% CI 0.834-1.184), age (OR 0.970, 95% CI 0.910-1.034), or stress level (OR 0.997, 95% CI 0.995-1.000) were identified to affect the adherence rates.

Conclusions: Female health care workers exposed to high workload did not find the web-based application useful for improving and modifying their lifestyle. Therefore, other strategies that might help health care workers facing stress and improve their lifestyle should be identified.

(J Med Internet Res 2020;22(8):e19500) doi:10.2196/19500

KEYWORDS

mHealth; eHealth; health care workers; occupational stress; burnout; adherence; web-based tool

Introduction

Occupational stress affects not only job satisfaction, but also personal and social life domains such as family, well-being, health, and friendships. Stress is a major health problem for both employees and employers, and can lead to illness or psychological distress, burnout, and labor turnover and absenteeism [1,2]. Work environments of health care professionals, especially those of nurses, have received increasing attention due to high absenteeism and staff shortages, augmented by dramatic cutbacks in funding and restructuring of health care services [3]. It was reported that 86% of nurses found their workplaces stressful and understaffed, 88% were under-resourced at work, and 91% experienced heavy workloads [4]. Healthy work environments are vital to the retention and recruitment of health care professionals and the sustainability of health systems [5]. Therefore, quality of work life (QWL) and the management of stress are extremely important for nurses [6].

It is well recognized that a healthy lifestyle can reduce the risk of morbidity and increase well-being and wellness. The key factors associated with this are healthy nutrition, exercise, and emotional well-being. Positive changes in lifestyle can have a wide range of health benefits, but ultimately change can only be implemented by the individual. Implementation of behavioral treatment and lifestyle management tools into the existing professional/health care system can ensure positive changes in lifestyle management. Many health care professionals describe lack of time and training as barriers to providing behavioral treatment. However, recent evidence suggests that intensive lifestyle interventions may not need to be performed in-person. Telephone- and internet-based treatment produced weight loss comparable to an in-person intervention [7]. Hybrid interventions that use technology and remote intervention components to the existing in-person treatment programs could prove even more effective. A promising way to promote physical activity and healthy diets worldwide is to implement interventions using electronic health (eHealth) and mobile health (mHealth) solutions. These interventions are primarily delivered via modern information and communication technologies such as the internet, mobile phones, and other wireless devices. In 2015, 90% of people owned a smartphone and various mobile apps are becoming part of our lives in every part of the world [7,8].

It is crucial that administrators and health policy makers foster environments that appreciate the realities of health care practice and uniquely affect the quality of work environments for health care professionals. Effective health policy should incorporate health, social contexts of occupational stress, and its relationship to QWL. Focusing on occupational stress followed by stress management interventions and their implications for QWL, we need to evaluate workplace interventions associated with healthy work environments and QWL, targeting the management of occupational stress and burnout [6].

As health care professionals are exposed to the psychological and physiological effects of stress, which is a well-known risk factor for a host of mental and physical health problems, the aims of this study were to assess the adherence of female health care workers to use a web-based tool for improving and modifying lifestyle and to identify the potential factors influencing their adherence.

Methods

Study Design

The was a prospective, observational study. Baseline data, data about the psychological and physical fitness, satisfaction with life, data about workload in 1 week (7 consecutive days), and participant's 3-month adherence to a web-based tool for improving lifestyle were collected from March to October 2019. The intervention included 2 major components: a mobile device that provided feedback/coaching about the diet and activity and web-based software which allows information entered via mobile phone to be visible and also tracked on the web-based platform when users log in to their personal account that monitored the progression of participant's selected program.

Settings and Participants

Female health care workers (physicians and graduated nurses) from 2 university medical centers and female members of the Slovenian Family Medicine Society were selected as participants. All participants were invited individually. Selected participants were needed to fulfill all the eligible criteria based on a first come, first served basis.

Inclusion criteria were female gender; working as an anesthesiologist, as a graduated nurse in an operating room, or as a family physician; and aged between 29 and 64 years. Exclusion criteria were pregnancy, physical conditions that prevent walking to a longer distance, unable to understand national language, and working only part time. Each participant signed an informed consent to participate in the study. Participants did not get any financial reimbursement for their participation. The study procedure was conducted according to the Declaration of Helsinki and approved by the departmental committee.

Blood was withdrawn from all participants in the morning. Immune status based on concentrations of immune cells from whole blood with EDTA (ethylenediaminetetraacetic acid) as anticoagulant was measured using flow cytometry. Serum samples were collected to measure serum levels of cortisol during the morning hours.

Data Collection

Data were collected between March and October 2019. Participants completed a questionnaire which collected the following data: basic demographics, satisfaction with life, and status of psychological fitness. Physical fitness was assessed using the Rockport test. Participants were then introduced to a web-based application/tool (24@life; RC IKTS). After registration to the application, a short training about the use of the application was provided. One week after the educational workshop, participants received a diary, on which they were needed to fill in data about their working hours, leisure time, and sleeping.

The number of log-ins into the web-based application and type of instruction selected were detected by the web-based software automatically. Based on the number of log-ins into the web-based application, the participants were divided into two groups: (1) adherent group, which logged into the application at least once a week and (2) nonadherent group, which logged into the application on average less than once a week over the period of 3 months.

Instruments and Definitions of the Variables

Several instruments were used to assess the psychological and physical fitness, quality of sleep, and overall satisfaction with life.

Satisfaction With Life Scale (SWLS)

The satisfaction with life was measured by accessing global cognitive judgment of participants' view of their life on a 5-item scale (Satisfaction With Life Scale or SWLS [9]). There is a 7-point scale from *strongly disagree* to *strongly agree* with a score range of 5–35. A score of 20 represents the neutral point on the scale. Scores between 31 and 35 indicate extremely satisfied, 26-30 indicate satisfied, 21-25 slightly satisfied, 15-19 slightly dissatisfied, 10-14 dissatisfied, and 5-9 indicate extremely dissatisfied. The scale has strong internal reliability and moderate temporal stability.

State-Trait Anxiety Inventory (STAI-X)

The State-Trait Anxiety Inventory (STAI-X) measures anxiety as a stable personality trait, a persons' disposition to be nervous, instead of the more prominent use of the term assessing an emotional state characterized by subjective feelings of tension, apprehension, nervousness, and worry, and by activation or arousal of the autonomic nervous system [10,11]. Form X of the STAI contains 20 state anxiety items and 20 trait anxiety items. The state anxiety items are each rated on a 4-point intensity scale, from 1 for *not at all* to 4 for *very much so*. The trait anxiety items are rated on a 4-point frequency scale (from *almost never* to *almost always*). Respondents are asked to indicate how they generally feel. Scoring is reversed for anxiety-absent items (eg, "I feel calm"). STAI was developed as a unidimensional self-report measure; 10 items are positively worded, and 10 items are negatively worded. Score range is from 20 to 80, with higher scores indicating greater levels of anxiety.

Sleep Quality

This is the self-report measure, in which participants rate their sleep quality on the 5-stage Likert scale: 1 being very bad and 5 very good in either last month or last night. We asked participants about the quality of sleep for the night before, in the last 3 days, and in the last month.

Physical Ability Test

Each participant underwent an aerobic power testing at our first meeting. The Rockport Fitness Walking Test is a maximal paced 1-mile walk test used to evaluate cardiorespiratory fitness through the estimation or prediction of maximal oxygen consumption (VO_{2max}) in adults [12]. This test requires the participant to walk 1 mile (1609 m) as fast as possible. We measured time needed and heart rate at the time the participants finished their walk. Oxygen consumption was calculated based on the formula for VO_{2max} (Multimedia Appendix 1).

Workload

We assessed workload from the number of working hours in an observed week based on the self-reports from the participant's diaries. Then, we interpolated the number of working hours in an observed week to a 4-week period (ie, 1 month). We defined the required number of hours per month as 160. Considering the Slovenian law, which states that health care workers should be included in out-of-hour's service, we accepted higher number of working hours as *standard* for workload. The highest number of out of hours which does not need the explicit permission of the workers is 32. The highest acceptable level of working hours was assessed to be 192 hours per week, and more than 192 hours per week was assessed as high workload. According to the law, 4 subgroups have been defined: (1) 169 or less hours per month, (2) between 170 and 188 hours per month, (3) between 189 and 201 hours per month; and (4) more than 201 hours per month.

Adherence

Adherence to drug treatment is generally defined as the extent to which patients take medications as prescribed by their health care providers. Good adherence to drug treatment is defined as taking at least 80% of drugs as prescribed. We defined adherence based on the number of log-ins into the web-based application. Good adherence was defined as log-in(s) into a web-based application at least one a week on average, meaning more than 12 times throughout the observation period. We divided participants into 2 groups: the adherent group, which includes those who logged into the application at least once a week; and the nonadherent group, which includes those who logged into the application on average less than once a week.

Statistical Analyses

Data analyses were performed using SPSS version 21 (IBM). A descriptive data analysis was performed. Distribution of the variables was measured by the Kolmogorov–Smirnov test. Correlations between the dependent variables were performed by one-way analysis of variance, whereas for estimating qualitative correlations between variables a Pearson chi-square test was used. In case of abnormally distributed variables a Kruskal–Wallis test was performed. Logistic regression was used to predict the adherence to the web-based application. The level of significance was set to $P<.05$ for all the tests.

Results

The basic characteristics of all participants at the time of their enrollment into the study are presented in the Table 1. The participants were divided into 5 groups: (1) general practitioner with normal workload, (2) general practitioner with high workload, (3) anesthesiologist, (4) graduated nurse working with an anesthesiologist, and (5) graduated nurse working in the surgery room. The characteristics of participants did not differ significantly. A statistical difference in age was detected between the graduated nurse working in the surgery room and general practitioner with normal workload ($P=.025$) and between the graduated nurse working in the surgery room and general practitioner with high workload ($P=.031$). A high workload was

detected in all groups, except in the general practitioner with normal workload group, which differed from other groups with a statistical significance (P<.05). The graduated nurse working with an anesthesiologist and graduated nurse working in the surgery room groups also showed a statistically significant difference (P=.028 for both) when comparing workload, despite both working overtime. The worst quality of sleep was associated with the graduated nurse working in the surgery room group, which also showed elevated levels of serum cortisol (>690 nmol/L).

Besides elevated S-cortisol levels, the graduated nurse working in the surgery room group showed elevated concentrations of lymphocytes (reference 1.1-2.5 × 10^9 cells/L), CD3 cells (reference 0.7-1.9 × 10^9 cells/L), CD8 cells (reference 0.2-0.7 × 10^9 cells/L), and HLA-DR/CD3 cells (reference 0.04-0.2 × 10^9 cells/L), which imply an activated cellular immune system (Multimedia Appendix 2). HLA-DR/CD3 cells were elevated in the general practitioner with high workload group.

Participants logged into the application in an observational period of 3 months from 0 to 235 times (median 9.5 times), which is, in general, less than once per week. Based on our definition of adherence to the tool, only 32 of 80 participants (40%) were adherent to the web-based application. The number of log-ins into the web-based application is presented in Multimedia Appendix 3.

On average, the most logins were recorded by the anesthesiologist and general practitioner with high workload groups. These two groups also showed the highest adherence rates. Meanwhile, the graduated nurse working in the surgery room group had the least logins, and the general practitioner with normal workload group had the lowest adherence rates (Table 2).

Participants most frequently logged in to the application to get recommendations for improving their lifestyle in the areas of physical activity (497 times) and nutrition (332 times) (Multimedia Appendix 4). There were no differences in adherence to the web-based application between groups (F_2=1.744, dfn=4, dfd=75, P=.184).

As the general practitioner with normal workload group showed the lowest rates of adherence, this group was chosen as the reference for comparison with other groups. Different workplaces of our study groups did not show any correlation with adherence to the web-based application. Except for factors listed in Table 3, no other participant's characteristics affected the adherence rates. Moreover, a negative correlation with VO_{2max} and S-cortisol was found (F_2=–0.238, dfn=4, dfd=75, P=.044).

Table 1. Basic characteristics from the enrolled participants.

Characteristics	GP[a] with normal workload (N=22)	GP with high workload (N=19)	Anesthesiologist (N=13)	Graduated nurse (N=26)		*P* value
				Working with an anesthesiologist (N=11)	Working in the surgery room (N=15)	
Age (years), mean (SD)	47.1 (6.6)	47.2 (7.5)	45.6 (9.0)	42.6 (8.4)	39.3 (6.5)	.020
Weight (kg), mean (SD)	67.8 (12.5)	64.4 (9.6)	66.8 (12.9)	60.2 (6.9)	66.4 (13.2)	.389
BMI (kg/m^2), mean (SD)	24.8 (4.2)	23.8 (3.7)	23.0 (2.7)	22.1 (1.9)	24.8 (4.0)	.177
Working hours/month, mean (SD)	166.7 (16.5)	232.4 (43.0)	228.4 (62.2)	260.0 (33.9)	209.6 (45.4)	<.001
Overworking time (hours), mean (SD)	8.7 (12.7)	72.4 (43.0)	70.2 (57.0)	100.0 (33.9)	49.6 (45.4)	<.001
Workload law						<.001
≤169 hours, n (%)	15 (68)	0 (0)	2 (15)	0 (0)	4 (27)	
170-188 hours, n (%)	6 (27)	0 (0)	0 (0)	0 (0)	2 (13)	
189-201 hours, n (%)	1 (5)	8 (42)	2 (15)	0 (0)	2 (13)	
>201 hours, n (%)	0 (0)	11 (58)	9 (69)	11 (100)	7 (47)	
STAI-X[b], mean (SD)	46.0 (3.5)	45.8 (4.4)	46.4 (6.0)	43.8 (5.0)	42.9 (6.3)	.149
SWLS[c], mean (SD)	24.9 (6.5)	27.2 (5.3)	26.8 (4.1)	26.5 (5.6)	20.9 (6.8)	.056
QoS[d] in the last night, mean (SD)	3.6 (1.1)	4.1 (0.7)	3.7 (1.3)	2.9 (1.1)	2.8 (1.2)	.011
QoS 3 days ago, mean (SD)	3.6 (1.0)	4.1 (0.7)	3.4 (1.0)	3.6 (0.8)	2.9 (0.8)	.009
QoS in the last month, mean (SD)	3.2 (4.8)	2.1 (4.2)	3.4 (0.9)	3.7 (0.8)	3.4 (0.8)	.369
Physical fitness: VO_{2max}[e] (mL/kg/min), mean (SD)	37.9 (4.7)	36.5 (7.0)	39.2 (9.1)	38.5 (6.4)	36.3 (7.8)	.819
Time: Rockport test (min:s), mean (SD)	14:32 (01:23)	14:10 (01:15)	13:43 (01:04)	13:46 (00:26)	14:58 (00:56)	.027
HR[f]: Rockport test, mean (SD)	130.6 (20.9)	128.8 (23.4)	135.0 (18.8)	136.1 (26.0)	128.7 (25.2)	.933
HR/HR_{max} (%), mean (SD)	75.3 (11.9)	74.3 (13.8)	78.1 (10.8)	76.9 (14.6)	71.1 (13.0)	.674
S-cortisol (nmol/L), mean (SD)	545.9 (130.6)	654.5 (234.8)	547.8 (145.2)	492.7 (165.2)	700.4 (268.4)	.017

[a]GP: general practitioner.

[b]STAI: State-Trait Anxiety Inventory.

[c]SWLS: Satisfaction with Life Scale.

[d]QoS: quality of sleep.

[e]VO_{2max}: maximal oxygen consumption.

[f]HR: heart rate.

Table 2. Adherence to web-based application according to workplace.

Characteristics	GP[a] with normal workload (N=22)	GP with high workload (N=19)	Anesthesiologist (N=13)	Graduated nurse (N=26)		*P* value
				Working with anesthesiologist (N=11)	Working in the surgery room (N=15)	
Log-ins, mean (SD)	20.1 (25.5)	38.5 (59.3)	35.7 (53.2)	15.0 (13.2)	11.9 (9.4)	.301
Adherence, n (%)	7 (32)	9 (47)	7 (54)	4 (36)	5 (33)	.751

[a]GP: general practitioner.

Table 3. Factors that affected the adherence to the web-based application.

Factor	B value	Odds ratio (95% CI)	*P* value
Weight (kg)	0.026	1.026 (0.977-1.078)	.301
BMI (kg/m^2)	–0.007	0.993 (0.834-1.184)	.941
Age (years)	–0.030	0.970 0.910-1.034)	.351
SWLS[a]	–0.043	0.958 (0.883-1.040)	.306
STAI-X[b]	0.043	1.044 (0.951-1.146)	.365
QoS[c] in the last night	–0.033	0.967 (0.643-1.455)	.873
QoS 3 days ago	0.372	1.450 (0.849-2.475)	.173
QoS in the last month	–0.121	0.886 (0.495-1.586)	.683
VO_{2max}[d] (mL/kg/min)	0.033	1.034 (0.962-1.111)	.368
HR_{max} (beats/min)	0.030	1.031 (0.967-1.099)	.351
Rockport	0.123	1.131 (0.760-1.682)	.545
Heart rate	–0.022	0.979 (0.955-1.002)	.076
Workplace			
General practitioner with normal workload	Reference	Reference	Reference
General practitioner with high workload	0.588	1.800 (0.396-8.182)	.447
Anesthesiologist	0.993	2.700 (0.600-12.150)	.196
Graduated nurse who works with an anesthesiologist	0.742	2.100 (0.448-9.836)	.346
Graduated nurse working in the surgery room	0.365	1.440 (0.260-7.961)	.676
S-cortisol	–0.003	0.997 (0.995-1.000)	.066

[a]SWLS: Satisfaction with Life Scale.

[b]STAI: X form of the State-Trait Anxiety Inventory.

[c]QoS: quality of sleep.

[d]VO_{2max}: maximal oxygen consumption.

Discussion

Principal Findings

Occupational stress leads to the burnout of individuals. The prevalence of burnout has been found to be extremely high (up to 75%) among medical staff [13-15] and it has been associated with poor quality of personal relationships, individual well-being, and poorer patient care [16-18]. This research has focused on an intervention to assist health care professionals in developing new personal habits needed to manage occupational stress and increase personal resiliency. The aims of this study were to assess the adherence of female health care workers to use a web-based tool for improving and modifying lifestyle and to explore the potential factors influencing their adherence to use such a tool and demonstrate the outcomes of this program's adherence to organizational stakeholders or educators. Overall, the adherence to the web-based tool for improving and modifying lifestyle in our study was relatively poor (32/80, 40%). Although the application offers a personalized goal setting, a large majority of health care workers were mainly interested in physical activity and nutrition and found other aspects of the app not applicable to them. Moreover, none of the investigated factors and participant characteristics could

explain the difference between the adherers and nonadherers (Table 3).

Our intervention incorporated a mobile device that provided feedback about diet and activity (Multimedia Appendix 4), both of which were self-reported. Such a technology offers new channels to reconfigure the provision of effective components of behavior and lifestyle change (ie, self-monitoring, goal setting, lifestyle change, and in-person sessions for resiliency). By enabling a remote personalized training, we can reduce the burden of both cost and time on participants, and thus connective technology systems can help ease the burden of stress.

We included 5 different groups of female health care professionals in the study. Among these groups, only 1 group (graduated nurse working in the surgery room) showed positive biological markers for chronic stress. Participants in this group had an activated immune response and elevated S-cortisol levels at the time of study enrollment. Meanwhile, this group also experienced the worst quality of sleep. Although other groups worked significantly over time compared with the group with normal workload (ie, general practitioner with normal workload), no clear sign of chronic stress has been discovered.

Prior to study execution, participants showed great interest in the intervention (data not shown), as technologies such as the one used in our study have the potential to overcome time limitations, especially among health care professionals, who generally work overtime. Besides this trial, few technology-supported weight loss interventions have been tested. For example, Rao et al. [8] reviewed randomized clinical trials of internet-based weight loss and weight maintenance programs including diet, physical activity, weight loss, self-monitoring and weight-loss outcomes, goal setting for behavior change, and motivation. One of the trials, performed by Patrick et al. [19], described participant adherence in a 16-week period. The adherence was 100% in the first week, but it declined to 66% by week 16. This shows that digital solutions such as eHealth interventions are confronted with the challenge of loss of participants' enthusiasm after the initial periods, which is also the case with our study. However, this was not the case in a trial by Burke et al. [20], who conducted a 24-month trial of a behavioral intervention for weight loss including paper records and personal digital assistant with dietary and exercise software, with or without a feedback message. After 6 months, adherence was statistically significantly higher in the 2 personal digital assistant groups than in the paper group and was maintained. These findings suggest that mobile devices still may be a useful tool for self-monitoring as part of standard behavioral interventions.

Despite not within the scope of this study, interventions do reduce occupational stress among health care professionals, who have a focus on changing their lifestyle [21]. Nonetheless, hospital management is not focusing enough on individual approach, but instead trying to reduce occupational stress with organizational policies and procedures, such as reducing working hours, workload, or increasing additional staff resources [22-24]. Stress is likely a result of both individual- and organizational-level processes, so such efforts do not really work long-term and have restricted potential [25]. Furthermore, the nature of work with continuous exposure to situations which require medical care that saves lives cannot be changed. When searching through the literature, a lack of studies focusing more on investigating interventions to reduce individual occupational stress and burnout was found [14].

In general, web-based education and self-care programs have been shown to be effective in managing lifestyle changes [26,27]. However, on the downside a common feature for these internet-based interventions is the low rate of user adherence. Although most participants in internet-based trials volunteer, and thus must have some intention to use the program, a substantial proportion of participants never use the program or use it only once [28]. In a review of internet-based interventions for physical activity, rates at 3-month follow-up ranged from 10% to 50% [26]. These rates coincide with those from our study (40%). In our study, the bar between the adherent and nonadherent groups was set quite low, as adherent participants were those who logged into the program at least once a week in a 90-day period. The reasons for low adherence might be attributed to less-focused intervention strategies. The programs administered need to meet the expectations of users in order to have a greater chance of uptake. While potential problems of adherence to interventions are not necessarily unique to programs targeting health care professionals, the focus of these interventions should be on reducing occupational stress while not adding burden to an individual's workload. It is, therefore, vital to ensure such interventions are integrated into the workplace or the lifestyle in a nonintrusive manner [21]. Participants in our study worked significantly over time, which might be a reason for bad adherence; however, the lowest adherence was observed for the general practitioner with normal workload group, who did not work overtime. Unfortunately, we did not collect the exact reasons as to why the users did not utilize the selected program, so further conclusions cannot be drawn. Another reason might be the design of the application. In this context, design and user-friendliness play an important role in improving adherence. In terms of personalization and application control, numerous studies have shown that people differ in the extent to which they seek and appreciate control [29]. People desiring high control generally seek better control over their events, and hence react with a negative emotional response when opportunities for control are lacking. Thus, in an online context, this may lead to nonadherence. Therefore, the software needs to be flexible enough to provide the end user with some control over the dedicated/selected programs [30]. In addition to the lack of control of activities, automated tailored messages, which are an integral part of and supported by the selected solution/program, might provoke an annoying feeling among users and this again might cause them to log-in less. This was confirmed by Tate et al. [31], who found that participants that received automated tailored messages had the lowest number of log-ins. Therefore, in our opinion, reminders could bring more benefit to the user, as they represent a trigger.

Limitations

Our study had some limitations, such as small sample size, which limited the trial to study factors related to adherence. Based on the inclusion criteria, only the most motivated health care workers participated in the study, who wanted to change

their lifestyle and represented only 3 different profile groups, namely, general practitioner, anesthesiologist, and graduated nurses. A majority of participants assessed had experienced/experiencing high workload. Thus, our results have a limited generalizability to all health care workers. In addition, subjective data about working hours of participant might overestimate their workload. Although adherence to the web-based tool was assessed, the influence of the web-based tool on lifestyle changes was not assessed in this study. Therefore, additional research is needed to further examine the relationship between QWL and occupational stressors. These may ultimately lead to the development and implementation of programs that are tailored to meet the needs of health care employees.

Conclusions

Adherence of health care workers to personalized web-based tools for improving their lifestyle in the current sociocultural environment remains relatively poor. Thus, it is necessary to use a qualitative approach to identify which interventions might be and which ways of implementing these interventions would be appropriate based on the needs of health care workers. In the future, it will be also necessary to not just follow the adherence to the web-based tool, but also to follow the effect of the use of such a tool on lifestyle changes and well-being of health care workers. This study also found that the benefits of behavioral interventions for occupational stress in health care professionals might be less optimal. The challenge for educators is thus to identify programs that are effective in improving QWL in health care workers.

Conflicts of Interest

None declared.

Multimedia Appendix 1
VO2 max formula.
[DOCX File , 14 KB - jmir_v22i8e19500_app1.docx]

Multimedia Appendix 2
Immunity status of participants at enrolment into the study.
[DOCX File , 18 KB - jmir_v22i8e19500_app2.docx]

Multimedia Appendix 3
Number of entries to the web-based application in the observational period of 90 days.
[DOCX File , 15 KB - jmir_v22i8e19500_app3.docx]

Multimedia Appendix 4
Interest of entries to the web-based application.
[DOCX File , 15 KB - jmir_v22i8e19500_app4.docx]

References

1. Bartram T, Joiner TA, Stanton P. Factors affecting the job stress and job satisfaction of Australian nurses: implications for recruitment and retention. Contemp Nurse 2004 Oct;17(3):293-304. [doi: 10.5172/conu.17.3.293] [Medline: 15551680]
2. Edwards D, Burnard P. A systematic review of stress and stress management interventions for mental health nurses. J Adv Nurs 2003 Apr;42(2):169-200. [doi: 10.1046/j.1365-2648.2003.02600.x] [Medline: 12670386]
3. Schalk DM, Bijl ML, Halfens RJ, Hollands L, Cummings GG. Interventions aimed at improving the nursing work environment: a systematic review. Implement Sci 2010 Apr 27;5:34 [FREE Full text] [doi: 10.1186/1748-5908-5-34] [Medline: 20423492]
4. Rajbhandary S, Basu K. Working conditions of nurses and absenteeism: is there a relationship? An empirical analysis using National Survey of the Work and Health of Nurses. Health Policy 2010 Oct;97(2-3):152-159. [doi: 10.1016/j.healthpol.2010.04.010] [Medline: 20493577]
5. Pino O, Rossini G. Perceived Organizational Stressors and Interpersonal Relationships as Predictors of Job Satisfaction and Well-Being among Hospital Nurses. IJPBS 2012 Nov 1;2(6):196-207. [doi: 10.5923/j.ijpbs.20120206.02]
6. Nowrouzi B, Lightfoot N, Larivière M, Carter L, Rukholm E, Schinke R, et al. Occupational Stress Management and Burnout Interventions in Nursing and Their Implications for Healthy Work Environments: A Literature Review. Workplace Health Saf 2015 Jul;63(7):308-315. [doi: 10.1177/2165079915576931] [Medline: 26084675]
7. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
8. Rao G, Burke LE, Spring BJ, Ewing LJ, Turk M, Lichtenstein AH, et al. New and emerging weight management strategies for busy ambulatory settings: a scientific statement from the American Heart Association endorsed by the Society of

Behavioral Medicine. Circulation 2011 Sep 6;124(10):1182-1203 [FREE Full text] [doi: 10.1161/CIR.0b013e31822b9543] [Medline: 21824925]
9. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
10. Spielberger C. Anxiety as an emotional state. Inpielberger CD. ed. Anxiety: current trends in theory and research. San Diego (CA): Academic Press; 1972:A.
11. Spielberger C, Gorssuch R, Lushene P, Vagg P, Jacobs G. Manual for the State-Trait Anxiety Inventory. 1983. Consulting Psychologists Press 1983:1-42. [doi: 10.1002/9780470479216.corpsy0943]
12. Rintala P, Dunn JM, McCubbin JA, Quinn C. Validity of a cardiorespiratory fitness test for men with mental retardation. Med Sci Sports Exerc 1992 Aug;24(8):941-945. [Medline: 1406181]
13. Willcock SM, Daly MG, Tennant CC, Allard BJ. Burnout and psychiatric morbidity in new medical graduates. Med J Aust 2004 Oct 04;181(7):357-360. [Medline: 15462649]
14. McCray LW, Cronholm PF, Bogner HR, Gallo JJ, Neill RA. Resident physician burnout: is there hope? Fam Med 2008 Oct;40(9):626-632 [FREE Full text] [Medline: 18830837]
15. Shanafelt TD, Boone S, Tan L, Dyrbye LN, Sotile W, Satele D, et al. Burnout and satisfaction with work-life balance among US physicians relative to the general US population. Arch Intern Med 2012 Oct 08;172(18):1377-1385. [doi: 10.1001/archinternmed.2012.3199] [Medline: 22911330]
16. Shanafelt TD, Balch CM, Bechamps G, Russell T, Dyrbye L, Satele D, et al. Burnout and medical errors among American surgeons. Ann Surg 2010 Jun;251(6):995-1000. [doi: 10.1097/SLA.0b013e3181bfdab3] [Medline: 19934755]
17. Shanafelt TD, Balch CM, Bechamps GJ, Russell T, Dyrbye L, Satele D, et al. Burnout and career satisfaction among American surgeons. Ann Surg 2009 Sep;250(3):463-471. [doi: 10.1097/SLA.0b013e3181ac4dfd] [Medline: 19730177]
18. Hall LH, Johnson J, Watt I, Tsipa A, O'Connor DB. Healthcare Staff Wellbeing, Burnout, and Patient Safety: A Systematic Review. PLoS One 2016;11(7):e0159015 [FREE Full text] [doi: 10.1371/journal.pone.0159015] [Medline: 27391946]
19. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL, et al. A text message-based intervention for weight loss: randomized controlled trial. J Med Internet Res 2009;11(1):e1 [FREE Full text] [doi: 10.2196/jmir.1100] [Medline: 19141433]
20. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]
21. Clough BA, March S, Chan RJ, Casey LM, Phillips R, Ireland MJ. Psychosocial interventions for managing occupational stress and burnout among medical doctors: a systematic review. Syst Rev 2017 Jul 17;6(1):144 [FREE Full text] [doi: 10.1186/s13643-017-0526-3] [Medline: 28716112]
22. Martini S, Arfken CL, Balon R. Comparison of burnout among medical residents before and after the implementation of work hours limits. Acad Psychiatry 2006;30(4):352-355. [doi: 10.1176/appi.ap.30.4.352] [Medline: 16908615]
23. Gelfand DV, Podnos YD, Carmichael JC, Saltzman DJ, Wilson SE, Williams RA. Effect of the 80-hour workweek on resident burnout. Arch Surg 2004 Sep;139(9):933-8; discussion 938. [doi: 10.1001/archsurg.139.9.933] [Medline: 15381609]
24. Hutter MM, Kellogg KC, Ferguson CM, Abbott WM, Warshaw AL. The impact of the 80-hour resident workweek on surgical residents and attending surgeons. Ann Surg 2006 Jun;243(6):864-71; discussion 871. [doi: 10.1097/01.sla.0000220042.48310.66] [Medline: 16772790]
25. Gopal R, Glasheen JJ, Miyoshi TJ, Prochazka AV. Burnout and internal medicine resident work-hour restrictions. Arch Intern Med 2005;165(22):2595-2600. [doi: 10.1001/archinte.165.22.2595] [Medline: 16344416]
26. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
27. Walters ST, Wright JA, Shegog R. A review of computer and Internet-based interventions for smoking behavior. Addict Behav 2006 Feb;31(2):264-277. [doi: 10.1016/j.addbeh.2005.05.002] [Medline: 15950392]
28. Wangberg SC, Bergmo TS, Johnsen JK. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]
29. Gebhardt WA, Brosschot JF. Desirability of control: psychometric properties and relationships with locus of control, personality, coping, and mental and somatic complaints in three Dutch samples. Eur. J. Pers 2002 Nov;16(6):423-438. [doi: 10.1002/per.463]
30. Ludden GD, van RTJ, Kelders SM, van GJE. How to Increase Reach and Adherence of Web-Based Interventions: A Design Research Viewpoint. J Med Internet Res 2015;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
31. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]

Abbreviations

EDTA: ethylenediaminetetraacetic acid

GP: general practitioner
OR: odds ratio
QWL: quality of work life
STAI-X: State-Trait Anxiety Inventory
SWLS: Satisfaction With Life Scale
VO2max: maximal oxygen consumption

Edited by G Eysenbach; submitted 20.04.20; peer-reviewed by N Haliti, M Skvarc; comments to author 12.06.20; revised version received 14.06.20; accepted 14.06.20; published 14.08.20.

Please cite as:
Jukic T, Ihan A, Petek Šter M, Strojnik V, Stubljar D, Starc A
Adherence of Female Health Care Workers to the Use a Web-Based Tool for Improving and Modifying Lifestyle: Prospective Target Group Pilot Study
J Med Internet Res 2020;22(8):e19500
URL: https://www.jmir.org/2020/8/e19500
doi:10.2196/19500
PMID:32687475

Original Paper

#BlackBreastsMatter: Process Evaluation of Recruitment and Engagement of Pregnant African American Women for a Social Media Intervention Study to Increase Breastfeeding

Cassy Dauphin[1], BA; Nikia Clark[1*], BS; Renee Cadzow[2*], PhD; Frances Saad-Harfouche[1*], MSW; Elisa Rodriguez[1*], PhD; Kathryn Glaser[1*], PhD; Marc Kiviniemi[3*], PhD; Maria Keller[4*], MS; Deborah Erwin[1*], PhD

[1]Cancer Prevention and Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY, United States

[2]Department of Health Services Administration, D'Youville College, Buffalo, NY, United States

[3]Department of Health, Behavior and Society, University of Kentucky, Lexington, KY, United States

[4]Department of Community Health and Health Behavior, State University of New York at Buffalo, Buffalo, NY, United States

[*]these authors contributed equally

Corresponding Author:
Deborah Erwin, PhD
Cancer Prevention and Control
Roswell Park Comprehensive Cancer Center
Elm and Carlton Streets
Buffalo, NY, 14263
United States
Phone: 1 716 845 2927
Fax: 1 716 845 8487
Email: deborah.erwin@roswellpark.org

Abstract

Background: In the United States, there are lower rates of breastfeeding among African American mothers, particularly those who are younger women. Recent epidemiological studies have shown a strong association of more aggressive types of breast cancer (estrogen receptor negative) among African American women, with a higher risk in African American women who did not breastfeed their children.

Objective: This study aims to describe the process evaluation of recruitment and educational strategies to engage pregnant African American participants for a pilot study designed to determine whether social media messaging about breast cancer risk reduction through breastfeeding may positively influence breastfeeding rates.

Methods: This pilot study is conducted in collaboration with a local Women, Infants, and Children (WIC) organization and hospital and prenatal clinics of a local health care network. To engage African American women to enroll in the study, several methods and monitoring processes were explored, including WIC electronic text-based messages sent out to all phones of current WIC recipients (referred to as *e-blasts*); keyword responses to texts from flyers and posters in local community-based organizations, hospitals, and prenatal clinics; keyword responses using electronic links posted in established Facebook groups; and *snowball* recruitment of other pregnant women by current participants through Facebook. Once enrolled, participants were randomized to 2 study conditions: (1) an intervention group receiving messages about breast cancer risk reduction and breastfeeding or (2) a control group receiving breastfeeding-only messages. Data were obtained through electronic monitoring, SurveyMonkey, qualitative responses on Facebook, focus groups, and interviews.

Results: More than 3000 text messages were sent and received through WIC e-blasts and keyword responses from flyers. A total of 472 women were recruited through WIC e-blast, and 161 responded to flyers and contacts through the local health care network, community-based organizations, Facebook, and friend referrals. A total of 633 women were assessed for eligibility to participate in the study. A total of 288 pregnant African American women were enrolled, consented, and completed presurvey assessments (102.8% of the goal), and 22 participants attended focus groups or interviews reporting on their experiences with Facebook and the educational messages.

Conclusions: This process evaluation suggests that using electronic, smartphone apps with social media holds promise for both recruitment and conduct of health education intervention studies for pregnant African American women. Providing messaging

and resources through social media to reinforce and educate women about breastfeeding and potentially provide lactation support is intriguing. Convenience (for researchers and participants) is an attribute of social media for this demographic of women and worthy of further research as an educational tool.

Trial Registration: ClinicalTrials.gov NCT03680235; https://clinicaltrials.gov/ct2/show/NCT03680235

(J Med Internet Res 2020;22(8):e16239) doi:10.2196/16239

KEYWORDS

breastfeeding; breast cancer education; African American mothers; Facebook; mobile phone, social media

Introduction

Background

Over the past 25 years, African American (AA) mothers have had consistently lower breastfeeding rates than any other group of women in the United States [1]. Only 74.0% of AA infants are ever breastfed compared with 86.6% of non-Hispanic White infants and 82.9% of Hispanic infants [2]. Epidemiological evidence shows that this disparity may negatively impact health and increase breast cancer risk in AA women [3]. Given the combination of evidence for racial disparities in lactation patterns and reduced breast cancer risk with breastfeeding for mothers, AA women may be disproportionately impacted by multiple negative health problems—with a strong opportunity for intervention. Although developing an educational intervention to impact this behavior change is challenging, a recent systematic review of interventions to enhance breastfeeding rates suggests that multilevel interventions can be successful. There is a significant need to explore social media communication, text messaging, and the internet to provide tools and methods for reaching and engaging AA mothers [1]. Given the evidence of health benefits to mothers and infants from breastfeeding, exploring novel interventions for reducing these socially patterned disparities is a significant scientific goal and the focus of this paper.

Breastfeeding and Breast Cancer Risks

"Breastfeeding rates in the United States (US) are considered socially patterned. Previous research has documented startling racial and socioeconomic disparities in infant feeding practices [4]." Nationally, AA infants and mothers are approximately 12% less likely to have the benefit of breastfeeding compared with White infants and mothers [2,5,6]. The most recent data as of 2017 from the New York State Department of Health indicate that breastfeeding initiation (*ever breastfed*) has increased to 84% in the AA population compared with 80.1% of White counterparts, but this does not include exclusive breastfeeding [7]. Moreover, by 6 months, the breastfeeding rates for babies were only 38% for AA and 44.5% for whites [7].

Recent epidemiological research suggests that choosing formula feeding over breastfeeding may have a significant impact on the risk of developing aggressive, triple-negative breast cancer (TNBC) in AA women [3]. Research findings reported by Palmer et al [3] showed that parity—formerly considered a risk reduction factor—actually *increases the risk* of TNBC in AA women. However, these breast cancer risks can be ameliorated in AA women if they initiate breastfeeding, and further research on the role of duration continues to be studied [3]. These results suggest that promoting breastfeeding in AA women may be an effective tool for reducing the occurrence of TNBC, which disproportionately contributes to breast cancer mortality for these women [3]. Therefore, encouraging breastfeeding among AA mothers offers a strong opportunity for intervention, given the combination of evidence for reduced cancer risk with breastfeeding and the overall national racial disparities in lactation patterns.

Barriers to Breastfeeding

Multiple barriers to breastfeeding among AA mothers have been reported in literature [1]. Primary issues include the burden of breastfeeding by women who work in jobs with little flexibility, unsupportive employers, and having limited maternity leave [4,8]. In addition, the social context for breastfeeding is weakened because AA women may not discuss the benefits of breastfeeding because of a lack of understanding and awareness [9,10]. Some women may have cultural discomfort regarding negative historical images as a result of the legacy of slavery [1,11,12]. Survey data collected from 225 middle and high school youth aged 13 to 18 years in Erie County, New York, showed that 62% of youth agreed or strongly agreed that "Breast milk is the best food for baby," but only 28% agreed or strongly agreed that "Breasts are meant for breastfeeding," and 36% reported this behavior as annoying. Young people exposed to formula feeding still outnumber those exposed to breastfeeding. These data demonstrate the challenges and limitations of changing cultural contexts to support breastfeeding. Personal experience and public or media exposure are statistically correlated with future intent to breastfeed [13]. Therefore, the educational intervention messages for this study included images aimed at positively impacting feelings and intent to breastfeed, and qualitative assessments included family members and partners of the mothers to explore social context.

Social Media and Smartphone Use for Reaching AA Women

To reach pregnant AA women, investigators chose to explore social media platforms for both study recruitment and intervention engagement, as smartphone and social media use are prevalent in this demographic. Research data show that the rate of social media use by all Americans is currently 69%, which is the same use rate among AAs [14]. The most frequently used social media platforms are Facebook and Instagram. The usage of these two social media platforms varies by age. Research indicates that 81% of young adult women aged between 18 and 29 years use Facebook, and 64% of this same demographic use Instagram [15]. For many users, social media

has become part of their daily routine, especially as most people have access to smartphones. The frequency of use by young adults aged between 18 and 29 years who check their Facebook or Instagram account daily is 72% [15]. Nearly three-fourth of AAs use smartphones to access social media compared with approximately 66% using desktop or laptop computers [14]. Facebook reported that the majority of users use the platform to post personal ideas, opinions, life events, and milestones (ie, graduations, vacations, employment, weddings, pictures of family, friends, or selfies) [15]. Moreover, social media communication methods may help address the reported lack of personal support [16] and low self-efficacy related to breastfeeding [1].

It is notable that income levels no longer present a major challenge in owning a smartphone, as 63% of low-income (<US $30,000 a year) AAs are owners of smartphones. However, low-income smartphone users who are AA are about twice as likely as whites to have their smartphone use canceled, or service interrupted because of the expense [14]. These interruptions with cell phones may create challenges when delivering services to low-income groups, although web-based social media platforms such as Facebook provide a unique mechanism of communication that continues despite cell service interruption.

Findings from prior research demonstrated similar rationales for the use of the technology in study recruitment for first-time AA mothers and for the use of social media to disseminate health education information on breastfeeding [17]. Social media access and use as an effective social support mechanism in pregnancy and postpartum have been tested by others [18]. Baker and Yang [18] found that social media present the opportunity to educate new mothers from the comfort of their home 24 hours a day.

This paper offers a process evaluation focused on the recruitment and study implementation techniques for a pilot breastfeeding intervention via smartphones and social media platforms among AA mothers participating in Western New York. As part of this larger pilot study, it was important to explore the acceptance and process of using social media and electronic apps to engage pregnant AA women with specific health messaging. This study focused on describing the methodological processes, benefits, and limitations for recruiting and engaging women through a social media conduit regarding breastfeeding and breast cancer risk reduction.

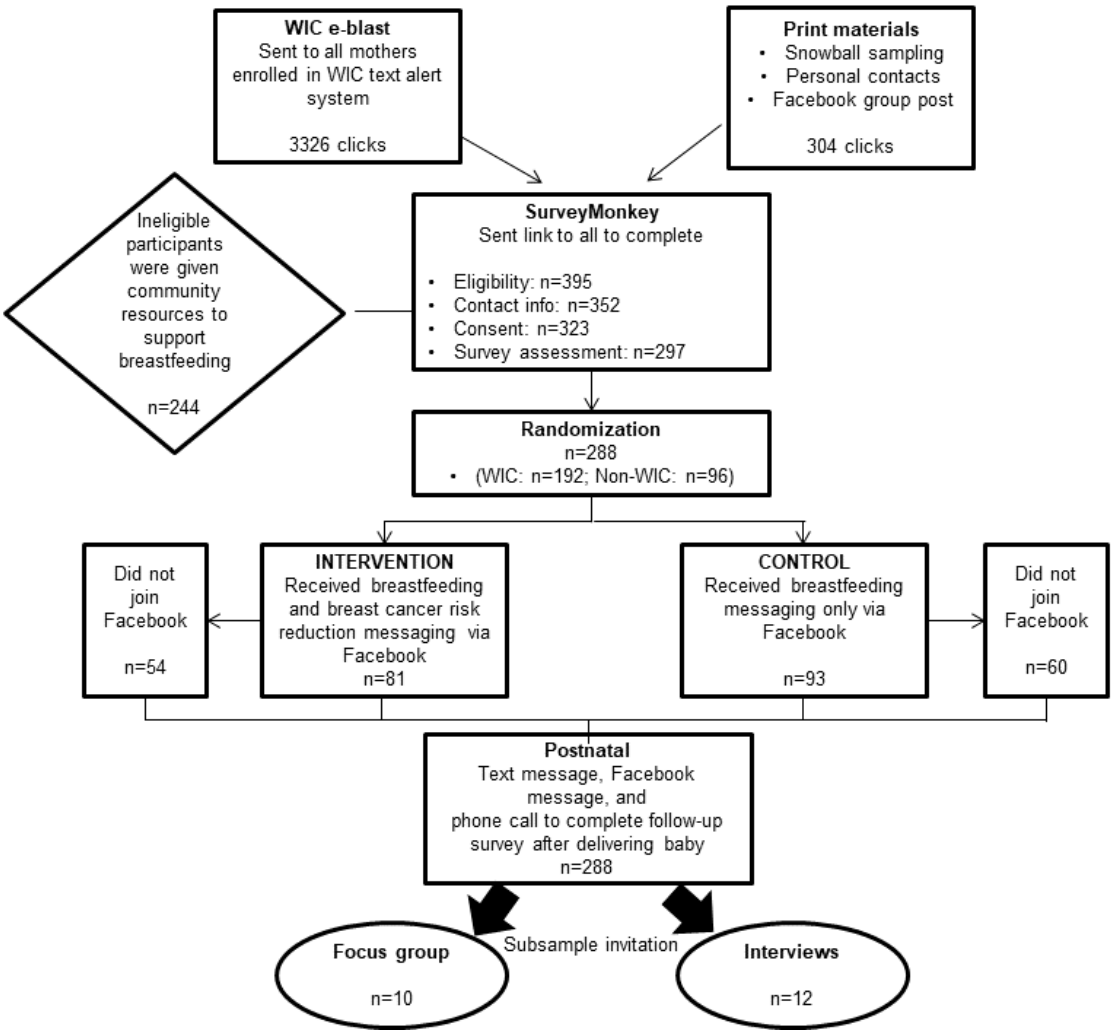
Methods

Intervention

The pilot study intervention was centered on enrolling eligible participants (AA, pregnant, and English speaking) into a private Facebook platform, and participants were randomized into either a control arm in which they received breastfeeding-only messaging or an intervention arm that received breastfeeding and breast cancer risk reduction messaging (Figure 1). Participants were asked to complete pre- and postbirth assessments and a smaller subgroup of mothers and their family members were recruited to participate in either focus groups or semistructured interviews postpartum. The goal was to obtain opinions from the mothers on the content presented in the Facebook group as well as the feasibility and effectiveness of using smartphone technology and Facebook for education and promotion of breastfeeding, with a specific aim of determining any outcome differences by specific message content (breastfeeding vs breast cancer risk reduction and breastfeeding).

Family members and partners of the mothers were included in the qualitative assessments to obtain pertinent social influences in the context of mothers' experiences. Focus groups and interviews included mothers and their support persons (family, friends, or partners) to discuss the content presented in the Facebook group. This study was registered in ClinicalTrials.gov, under registration number NCT03680235.

Figure 1. Study schema. WIC: Women, Infants, and Children.



Study Participants

The study population consisted of pregnant AA mothers residing in Erie and Niagara County, especially those served by Women, Infants, and Children (WIC). As the WIC program in Erie County serves a significant proportion of the pregnant AA women in the area, we partnered with WIC to inform their clients about the study. When the study was designed (2016), the Erie/Niagara WIC assisted women with approximately 2000 births per year; however, data are not available by race [19].

Participant Demographics

Of the study participants, approximately 25.3% (73/288) completed high school or received a general educational development (GED) certificate, 33.3% (96/288) completed some college, and 12.2% (35/288) were college graduates. In addition, 55.2% (159/288) of the mothers reported that they were employed, approximately 56.6% (163/288) reported making less than US $20,000 per year, 13.5% (39/288) earned US $20,000 to US $29,999, and 13.5% (39/288) earned US $30,000 to US $49,999, with 1.0% (3/288) earning more than US $50,000, and 15.3% (44/288) declined to answer. Moreover, 70.8% (204/288) of the mothers were never married; 24.0% (69/288) were married or partnered; and 4.9% (14/288) were divorced, widowed, or separated. A total of 22.6% (65/288) of women reported being breastfed as a child. Overall, 26.7% (77/288) of women reported this as their first pregnancy, 20.8% (60/288) reported this as their second pregnancy, 21.2% (61/288) as their third pregnancy, and 31.3% (90/288) as their fourth or more pregnancy. Approximately, 53.8% (155/288) of the women reported the age of their first pregnancy being 19 years or younger, 36.5% (105/288) reported first pregnancy between the ages of 20 and 25 years, 5.6% (16/288) reported first pregnancy between the ages of 26 and 29 years, and 2.1% (6/288) reported first pregnancy at 30 years and older (Table 1).

Table 1. Demographics.

Demographic categories	Values, n (%)
Age (years)	
14-17	3 (1.0)
18-24	101 (35.1)
25-34	149 (51.4)
35-44	35 (12.2)
Currently employed	
Yes	159 (55.2)
No	127 (44.1)
Not reported	2 (0.7)
Marital status	
Single or never married	204 (70.8)
Widowed, divorced, or separated	14 (4.9)
Married	69 (24.0)
Not reported	1 (0.3)
Education	
Less than high school	29 (10.1)
High school or general educational development certificate	73 (25.3)
Some college	96 (33.3)
College degree	35 (12.2)
Not reported	55 (19.1)
Income (US $)	
<20,000	163 (56.6)
20,000-29,999	39 (13.5)
30,000-39,999	26 (9.0)
40,000-49,999	13 (4.5)
>50,000	3 (1.0)
Not reported	44 (15.2)
Age of first pregnancy (years)	
<19	155 (53.8)
20-25	105 (36.5)
26-29	16 (5.6)
>30	6 (2.1)
Not reported	6 (2.1)
Breastfed as a child	
Yes	65 (22.6)
No	153 (53.1)
Not reported	70(24.3)

Recruitment and Assessment Procedures

To engage AA women to enroll in the study, the following methods and monitoring processes were explored: (1) WIC text-based *e-blast* messages were sent to all WIC enrollees at specified clinic sites, (2) in-person recruitment was conducted at prenatal clinics and community-based organization events using keyword response giving participants web links to eligibility criteria and consent documents, (3) keyword response text messages were monitored through Telerivet (Telerivet, Inc) based on flyers and posters in the hospital and WIC office, (4) Bitly (Bitly.com) was used to track *clicks* on the link navigating

women to the web-based SurveyMonkey (SVMK Inc) surveys, (5) brochures and posters in prenatal clinics were distributed with electronic links described, (6) recruitment of pregnant women from established support groups through Facebook, and (7) recruitment of other pregnant women by current participants through Facebook was encouraged. Once women consented and completed the prenatal survey, they were randomized to the 2 study conditions and 2 private Facebook groups. Data were obtained through electronic monitoring, SurveyMonkey, qualitative responses on Facebook, focus groups, and interviews.

These recruitment processes used 4 computer-based mobile messaging platforms. The first mobile messaging platform was operated by WIC and was used to disseminate the initial enrollment text through bulk messaging referred to as an *e-blast*. The second app was used to receive enrollment text messages referred to as *keyword responses* and was monitored through Telerivet by study staff. Telerivet is a web-based mobile text messaging app that serves as a messaging command center and allows program owners to send *text blasts* or bulk messaging to multiple participants who would receive the text individually (ie, not group texts). Through this app, study staff were able to streamline communication by using features such as message scheduling or staggering, which allows the scheduling of messages for times or recurring intervals. This app allowed staff to schedule polls via text messages in which participants could respond either by numeric or alphabetic responses, eliminating the need for manual data entry. This app also had an automatic stop function that allowed participants to *opt out* and stop receiving messages.

Bitly is a link management platform that was used to track the amount of interest in the study through *clicks* on the link that navigated women to the web-based survey. Participants could click the e-blast–delivered link multiple times to be able to access and return to the survey. SurveyMonkey was the fourth web-based software used to collect survey responses from the women through their mobile devices and build the study database.

The initial recruitment request was sent out through an *e-blast* to all WIC participants who served and registered at urban Buffalo WIC clinic sites that included the highest proportion of AA clients (Figure 1). The monthly e-blast was sent through the WIC-owned and operated mobile messaging platform to all mothers aged ≥18 years and enrolled at the specified sites who were currently pregnant and due to deliver within 3 months, with the intent to enroll women and allow them to receive Facebook messaging for at least two to three months before delivery. The clients' mobile phone numbers were an existing part of the WIC client database and used by WIC for various communication processes, so the study team did not have to have clients' permission to contact individuals. The e-blast message briefly introduced the study to the clients and gave them a link to explore it in more depth (including eligibility criteria and consent). As WIC does not report data by race, e-blast messages were sent to all clients from the selected clinics regardless of race, and eligibility was assessed after they responded. Recruitment text messages were delivered monthly to cohorts of women based on their expected delivery month over 3 months (ie, May-July). Each woman received up to 3 recruitment text messages from the WIC's mobile messaging platform.

On receiving the text on their smartphones, WIC participants could opt to click on a link to connect them to the study eligibility criteria questions (eg, pregnancy status: due date, WIC enrollment status, race: AA, primary language: English). WIC clients who did not meet the study eligibility criteria received the following message: "Thank you but you are not eligible for this study" and were directed to other links with information about healthy pregnancy practices and community resources such as the Durham Baby Café, a local Center for Teens, Moms & Kids, and WIC Breastfeeding Partners. Eligible participants were directed to the electronic consent form, which required one click at the bottom of the form for *Agree*. The study was approved by the Roswell Park Cancer Institute institutional review board (IRB) as minimal risk.

To be more inclusive of pregnant women aged <18 years (New York State WIC did not allow recruitment of WIC clients aged <18 years for research), the study team partnered with additional obstetrics and gynecology (OB/GYN) offices in the urban area and community-based organizations focused on delivering services to pregnant women. Posters and flyers to promote the study were displayed and disseminated at all participating recruitment sites as well as in-person recruitment by study staff visiting the clinics. The poster and flyers contained a keyword response, "OurBreastsMatters," which participants were requested to text to a long-code 10-digit local phone number. For consistency, posters, flyers and original e-blasts from WIC used a race-neutral study name because WIC was not able to send messages to only AA clients. After meeting eligibility criteria, participants were directed to the BlackBreastsMatter site for further surveys and Facebook intervention. Similar to the WIC enrollees, participants received an automated response that contained the link to opt in to the study and connect them to the eligibility criteria. These methods allowed participants to initiate the process for enrolling in the study while having the ability to complete the survey at their convenience. In addition, a final strategy consisting of a snowball sampling approach was also used to recruit additional participants by offering this study's participants an incentive to recruit other pregnant friends and relatives using the keyword response approach.

All study participants received both weekly and monthly scheduled text messages regarding their status in the project as well as their current use of the resources allocated by our partners. Telerivet allowed study staff to streamline participant contact and vary the amount of contact for participants at different intervals in the study.

Data Collection Procedures

Using SurveyMonkey, participants were asked to complete eligibility criteria, social media use, demographics, study consent, and pre- and postpartum surveys from smartphones. Once the participants consented, they were asked to complete a 73-question preintervention baseline survey on their smartphone. Participants could return to the survey if they did not choose to complete it immediately. Participants were incentivized with a US $40 gift card to local retail shops to

complete the survey. Women who completed the baseline survey (N=288) were randomized to either the intervention arm (n=135) or control arm (n=153). Participants were asked to sign on to Facebook to receive messages through a *private* Facebook group that was invitation only and did not publicly appear as an open group nor available through the search bar menu on Facebook. The intervention and control groups received targeted weekly or biweekly Facebook posts. Electronic engagement was encouraged by a variety of polls, discussions, and raffles.

Unlike commercial Facebook pages, private pages were not eligible for tracking activities (eg, likes, comments) at the time this study was conducted; therefore, methods were limited to manual review of comments, responses, and questions. A manual data collection spreadsheet was initiated and monitored for 2 months but proved too time-consuming to continue with the limited staff supported by the pilot (NIH/NCI R21) grant funding. It was determined to be more important to engage and interact with the participants on Facebook, continuing regular posting of educational messages, and answering questions and comments.

Approximately 4 to 6 weeks postpartum, participants received text invitations and Facebook reminders to complete the postpartum survey. The expected due dates were collected through baseline surveys. Study staff often discovered births, particularly preterm births, through Facebook postings. After delivery and completion of the postpartum assessment, a smaller sample (n=23) of women, family members, and partners were recruited for focus groups or semistructured interviews to qualitatively assess the use of technology as well as the message content.

Intervention Delivery

The Facebook group breastfeeding messages centered on 5 themes of support: bonding, health and wellness for mother, health and wellness for child, financial impact, and social or lifestyle impact. Eligible participants were randomized into a control Facebook group page that received only these breastfeeding messages or an intervention Facebook group page that received both breastfeeding messaging and breast cancer risk reduction (as it relates to breastfeeding) messaging, including a video presenting this information (Multimedia Appendix 1). All messaging was culturally appropriate for AA women (Multimedia Appendix 2). Participation in the Facebook group pages was intended to engage participants in the topics and viewers were able to *like* posts, comment, ask questions, and contact study staff for more information or if they had any issues or concerns.

As new participants joined the respective Facebook group throughout the study, the 5 topics periodically rotated so that participants engaged in Facebook would be exposed to all 5 topics over approximately 5 weeks. As the topics were presented, the specific content and photos offered new or revised information so that the participants did not see the same content repeatedly.

Focus Groups and Interviews

In addition to quantitative surveys (for measuring intervention outcomes), focus groups were hosted to discuss participants' feelings, perceptions, and thoughts on the topic of breastfeeding, messages about breast cancer risk, and participation in the intervention in general. Participants who had completed the study and delivered their babies were invited via text messages and phone calls to attend a focus group. In total, 4 focus groups were hosted at the Roswell Park Comprehensive Cancer Center (2 groups included mothers who participated in the study, and 2 groups included individuals identified as support; eg, family members and partners) by the mothers. These groups were hosted on a Saturday for the convenience of the participants. Although family members and partners did not directly receive Facebook posts, the published research clearly demonstrates that breastfeeding occurs or fails to occur within a larger social context that is directly impacted by the views and experiences of the mothers' family members and partners. Therefore, the pilot study was designed to explore the social context of the use of Facebook and the messaging with the mothers and their significant others. For example, did the mothers and their supportive family members share the information they learned and how was this received?

All sessions lasted no more than 60 min with the first session starting in the morning and the second session hosted in the afternoon to give participants varying options in which group they would like to participate. The discussion included questions concerning personal perception and relationship with breastfeeding, risk perception of breast cancer, benefits and barriers to breastfeeding, and the use of social media and the messages they received. The groups, including other family members (ie, support), covered the same question content, included an explanation of what the study and intervention addressed, and offered examples of messaging shared with mothers involved in the study. Refreshments were served, and all participants received an incentive of US $30 for participating in the focus group as well as the ability to participate in a free community holiday photography shoot hosted at the cancer center.

Semistructured interviews were used to specifically cover topics that were triggered by the focus group participants in a grounded theory approach and to allow data gathering from male partners and more family members who did not choose to attend focus groups. As some issues related to breastfeeding may be sensitive and mothers and family members may not want to discuss them in a group setting, additional individual interviews were offered with participating mothers and their selected support persons. A total of 12 interviews were conducted, 9 by phone (8 mothers and 1 support participant), and 3 male support member interviews completed in person. The interviews consisted of topics similar to those addressed in the focus groups, such as personal perception and relationship with breastfeeding, risk perception of breast cancer, and benefits and barriers to breastfeeding. Phone interviews averaged 15-20 min and 5-10 min for the in-person interviews. All interviews were conducted by trained study staff and were race concordant. Participants who completed the interviews also received gift card incentives.

All focus groups and interviews were recorded and transcribed. These transcriptions were reviewed and coded by 4 members of the study team. Discrepancies were discussed, and final categorical analyses were determined by consensus. Text coding

was sorted into thematic codes and subcodes. Text analysis was validated by discussions with the project staff who were directly interacting with the participants in focus groups, through Facebook and texts.

Results

Recruitment

In total, WIC delivered more than 3000 text messages to women enrolled in their mobile messaging platform from which there were 1113 responsive *clicks* on the link to the survey. More than 66.7% (192/288) of the participants were recruited through the WIC e-blasts (Table 2).

Owing to the structure of the delivery of recruitment text through WIC, it was expected to receive fewer *clicks* compared with the text messages delivered, as women could receive the recruitment text from WIC up to 3 times. Through our partnerships with the local health care network, including hospital and prenatal clinics, offices and community-based organizations, Facebook support groups and snowball referrals we received more than 200 incoming text messages initiating the keyword response (from posters, flyers, and personal contacts) to receive the link to begin the survey from which there were 304 *clicks*. It was expected to see more *clicks* compared with incoming text messages from this group, as women engaged and recruited in person would often return to the survey link later to complete the enrollment information, so the conversion to participation rate was higher (78.9% [45/57] compared with 40.7% [192/472] for WIC), although the reach was greater through WIC. Through these combined recruitment efforts, 633 women were assessed for eligibility, with 288 pregnant AA women (103% of the original goal, n=280) successfully enrolled and consented (Table 2). Exclusions included 345 women not meeting inclusion criteria based on race, 29 women declined, 37 women did not provide follow-up contact information for enrollment, 26 women did not complete presurvey assessments, and 9 women gave birth before enrolling in the Facebook group. Of the 288 consented participants, 135 were randomized into the intervention group and 153 were randomized into the control group.

Interestingly, of the women enrolled in the study through local OB/GYN offices and community-based organizations (n=96), 90% (n=86) were currently enrolled in or planning to enroll in WIC services, indicating that the study sample can be expected to be relatively similar from both recruitment sources. Times when women (from all recruitment sources) clicked on the study e-blast invitations, enrolled, and completed surveys on their smartphones were widely distributed: 41.7% (120/288) of surveys were completed between the hours of 7 AM and 3 PM, 54.5% (157/288) of surveys were completed between the hours of 3 PM and 11 PM, and 3.8% (11/288) were completed between 11 PM and 7 AM. Of the 288 participants, 20.5% (n=59) responded to recruitment and enrollment messages on their smartphones *after working hours* between 6 PM and 9 AM.

Overall, 92.4% (266/288) of the mothers stated that they used or knew how to use Facebook, and 99.7% (287/288) used or knew how to text. A total of 96.5% (278/288) of women stated that they were willing to join and use Facebook and text messaging for involvement in the study. Although almost all participants reported knowledge, use, and willingness to engage in Facebook, 39.9% (115/288) of consented participants randomized to the intervention and control arms failed to ever join the assigned private Facebook group. This inadvertently created *true controls* for the study, as 40.0% (54/135) of the breastfeeding plus breast cancer risk (intervention) and 39.2% (60/153) of the breastfeeding-only (control) message groups did not receive any of the educational messages. All participants completed prebirth baseline surveys, and 74.0% (213/288) completed postnatal surveys.

An anecdotal finding as women had their babies was that 60 new mothers from both Facebook groups became more active and engaged with the topics of breastfeeding and newborn care and requested to be able to continue to communicate on their Facebook page (study participants were considered *complete* and removed from the study protocol and follow-up once they completed their postnatal survey about 4- 8 weeks after the birth of their infants). In response to this, the study staff set up a third Facebook group for women after they had completed the postnatal survey and study protocol, but it was just an open group without designed study messages. As this was not part of the original study design and budget, no further data collection or postnatal messaging was planned.

Table 2. Recruitment sources for participants.

Sources	Recruited, n (%)	Enrolled and randomized, n (%)	Conversion rate, %
WIC[a] e-blast	472 (74.6)	192 (66.7)	40.7
Obstetrics and gynecology office	57 (9.0)	45 (15.6)	78.9
Hospital recruitment	13 (2.1)	6 (2.1)	46.2
Facebook recruitment	44 (7.0)	16 (5.6)	36.4
WIC office recruitment	14 (2.2)	8 (2.8)	57.1
Community-based organization	24 (3.8)	14 (4.9)	58.3
Friend referral	9 (1.4)	7 (2.4)	77.8
Total	633 (100.0)	288 (100.0)	45.5

[a]WIC: Women, Infants, and Children.

Focus Group and Interview Participants

Of the 10 female participants who attended the focus groups, 6 participant mothers and 4 family or friend support individuals (2 mothers and 2 friends), all mothers reported breastfeeding for at least three weeks. The participants' average age was 33.3 years (median 29.0, SD 10.4), 4 participants reported working full time or part time, 3 participants completed high school or had a GED certificate, 4 participants completed some or all college, 7 participants reported earning US $20,000 or less in household income, and all participants reported being insured. Of the 9 female participants who completed interviews over the phone, the mean age of mothers was 28.4 years, 6 had never married, 9 reported completing high school or more, 3 reported working full time, and 6 earned less than US $20,000. We did not collect demographics for the 3 male partners.

Of the female participants in focus groups and interviews, 9 participants were aged >25 years and were often more experienced with pregnancy and breastfeeding than the total study sample. Participants with prior breastfeeding experience were very positive about breastfeeding, social media messages, and supportive of any or all efforts to increase breastfeeding. Other less-experienced participants reported knowing very little before the intervention, such as, "Didn't know much before…," "Learned about how it helped the baby," and "Was just going to see how it went." Another inexperienced mother demonstrated the limitations of education alone, stating, "What I saw made it look easy, and that wasn't true—not for me." The participants did report a litany of the benefits they learned about breastfeeding (eg, brain development, saving money, benefits to the mother, bonding, and lowering the risk of breast cancer). It was notable that several mothers reported memorable messages about breastfeeding rights and issues of "to cover-up or not cover-up when breastfeeding in public."

In response to the use of Facebook, almost all focus groups and interview participants would have liked to stay in the group longer, possibly share posts, or to have had an ongoing Facebook support group postpartum. There was a general sentiment that "I would have liked to invite friends into the group;" "I wanted to share posts with friends in other groups;" be allowed "to share content with other groups and friends;" "I would have loved to invite others!" The participants also would like to have had access to more videos on the Facebook pages. All the women expressed support for the title and logo of the Facebook group, "Black Breasts Matter." Inquiries about the logo on Facebook also received positive responses. "It was wonderful because you're supporting Black women;" "Rings a bell;" "Loved it!" Several participants stated that they do not see very much media or information specific to Black women and breastfeeding: "I think that Caucasian people actually do it [breastfeed] more than black women, so that gets us motivated to do something that we're not so familiar doing" and "I liked it because it had to do with blacks…you know it was more for like black mothers and it made me feel comfortable." (Notably, the tag to initially enroll in the study was "Our Breasts Matter" in deference to the New York State Health Department concerns showing any racial exclusion with the study. It was not until a participant met qualifications and was consented that they saw the "Black Breasts Matter" logo. Ineligible White participants were routed to other breastfeeding resources.) Interestingly, one participant shared the perspective that Black women's breast milk is superior to other mothers' milk and believed this could be empowering other Black women to breastfeed. Breast cancer risk information and messaging about potential risk reduction from breastfeeding for those in the intervention group was commented on as being *impactful* and new information for most of the women, and they wanted to share this information with family and others.

Discussion

Principal Findings

This study describes the use of social media and electronic platforms to effectively recruit and engage pregnant AA women into an intervention study focused on educating about the benefits of breastfeeding and the potential to reduce breast cancer risks. The specifics of how smartphone and social media apps can create a social and educational space as well as a novel approach for discussing breastfeeding as a form of breast cancer risk reduction, specifically for AA women, are described. The process evaluation illustrates various strengths and limitations of methods and recruitment processes for this demographic of participants and suggests ideas for future studies incorporating smartphone and social media use.

The overall strengths of the study and the methods employed included the successful (18 months) recruitment of 288 pregnant AA women into the study and the effectiveness of electronic consent forms and surveys. More than 58.3% (168/288) of participants explored study participation and eligibility and consented and completed surveys during evening hours (after 3 PM). Almost 20.5% (59/288) of participants completed study tasks during nontraditional working hours for study staff (ie, 6 PM-9 AM). The ability of participants to complete surveys and read messages by smartphone at any time of the day or night, depending on their schedules and needs, was an advantage for the study staff and participants and was less challenging than recruiting for and scheduling the in-person focus groups or interviews for this study. This recruitment process was also more effective than attempts in a previous study in Buffalo in collaboration with Planned Parenthood that had very poor results collecting in-person survey data during daytime working hours from this demographic of younger women regarding tobacco use and cervical cancer risk [20].

Of the recruitment sources explored, WIC e-blast had the greatest reach and proportion of participants enrolled (192/288, 66.7%) as well as being able to contact potential participants multiple times in a cost-effective way. In-person recruitment using keyword responses at prenatal clinics and community-based organizations resulted in a lower proportion of enrollments (approximately 30%), but the conversion rate for these in-person and especially the snowball friend referrals (7/9, 77.8%) was high. These more direct methods were also more time-consuming and labor-intensive (Table 2).

It was notable that the participants strongly endorsed and welcomed the racial and cultural tailoring of the messages, photos, and issues presented in the Facebook intervention.

Qualitative responses included pride and opportunities for the promotion of breastfeeding for enhancing Black infants' and mothers' health and appreciation for focusing on their needs and issues at the policy level. As the lowest rates of breastfeeding are among AA women and many of the barriers, including racially biased health care [1] and negative social contextual issues disproportionately impact AA women, it could be important to address these issues in future intervention approaches. Moreover, incorporating racially specific messages may enhance self-efficacy by AA mothers to address policies and feel greater support through social media.

Future Studies and Limitations

An important consideration specifically for studies incorporating Facebook is that there are operational differences for a private Facebook group versus more commercial, *business*, or even personal Facebook accounts. A private Facebook page operates in a much more limited fashion than an organic, individually based, or business-based Facebook page that includes individuals familiar with one another or with some type of shared experiences. The fact that all participants were pregnant helped to focus and engage some women, but they were not naturally familiar with one another, and the general trust and sharing activities in the study Facebook pages were more limited. As time progressed, women seemed to become more familiar with one another, especially as they gave birth and were faced with issues that they wanted to share, and this reinforces that support and evaluation of social media activity postnatally should be included in future research. The fact that 60 mothers were interested in having more education and interaction after their babies were born and they had completed the study demonstrates that addressing the issues and challenges of new mothers, including those of breastfeeding, often gain much greater awareness with the immediacy of needs. This further reinforces the theory that *teachable moments* and *need to know* timing is important for adult learners. The comments by participants also indicate limitations in the nature of the private Facebook groups, as the information on their group pages was limited to sharing only with study participants. Participants could not electronically *share* with other family and friends or Facebook pages. This may have limited activity, ability to influence their own social network, as well as the responses and conversations about the postings.

One limitation of the study recruitment process in collaboration with New York State WIC was the fact that the New York State Department of Health limited all research to be done with individuals aged ≥18 years. This was not known by the study team before the initiation of the study. As up to 2% of births among AA women are in women aged <18 years and the younger the age at first birth, the more births they are likely to have, resulting in potentially higher risks for TNBC, it was important to include younger women [21]. The study IRB approval process allowed recruitment of women aged ≥14 years; however, it was challenging to reach and recruit these younger women without the WIC recruitment process, even within the local OB/GYN practices.

National corporate changes in technology security issues with Facebook are issues that may impact participation in a study such as this one. With the use of social media comes the constant alteration, upgrades, and changes to the operational use of the app as well as features that are accessible through the app that may leave users vulnerable. In 2018, as the study was winding down, Facebook publicly faced two major scandals concerning the use of private information of its users. It was publicized that Facebook faced a breach and attack on its computer network, which in turn left the personal information of nearly 50 million users available as well as access to their accounts. Although discussions in community settings occurred about the safety of Facebook, we did not see a decline in recruitment after the announcement of the breach, but we did have additional challenges.

Challenges and limitations of the social media methodology included the fact that about one-third (n=114) of the participants never chose to sign on to and receive any messaging from the private Facebook groups, basically creating a *true* control with no messaging about breastfeeding. It was unclear why participants failed to sign on, as survey data demonstrated a high proportion (>92%) of participants had knowledge of and used Facebook personally. This may have been the result of a lack of trust or unfamiliarity with the use and practices of a private Facebook account. This reinforces the challenges of using Facebook for this clientele, as it may be waning in popularity as a communication platform with this demographic.

At the time of the study, Facebook did not have a feature to be able to collect analytical metrics from Facebook groups but rather only activated this feature for business pages. One of the more recent advancements made is "Group Insights," which now allows the generation of reports from a created group including metrics such as the growth of total members, pending members, approved requests, engagement through posts, comments, reactions, activity status, and member contribution ranking members who contribute more than average members. Unfortunately, this advancement could not be applied to an already established group but is a current feature available for future intervention research.

Monitoring the use of Facebook was another limitation that we encountered. The culture of social media is based on immediate responses, new content that is updated frequently, and 24-hour access. The staff implementing this study was small (2 part-time staff); therefore, it was challenging to meet the demands for posting new messaging and responding to study participants' comments and questions quickly compared with other social media groups who have dedicated staff members posting and monitoring content. Feedback from study participants (eg, posting more videos) indicated that they would have liked more production and were more engaged within the Facebook group pages by having more access to study staff through Facebook Live and web-based question-and-answer sessions. In addition, when using a social media platform such as Facebook for research study purposes, there is a learning curve for what information can be accessed and tracked and best practices on how to do so. The study staff learned that there are restrictions put in place by Facebook that did not allow them to automatically track and extract a study participant's frequency of participation within the group. Some of these limitations are

different for private accounts versus commercial or personal accounts.

Conclusions

Educating pregnant AA women about the positive attributes of breastfeeding and potential benefits for breast cancer risk reduction is countered by strong circumstantial and time constraints for many who may still be in school and are still more engaged in their age cohort activities and interests. The options for providing messaging and resources through social media to reinforce breastfeeding and possibly provide lactation support hold promise. Convenience is a major attribute of social media for this demographic and worthy of further research as an educational tool.

The use of text messaging platforms such as Telerivet in combination with survey collection software holds promise as an effective tool for electronic recruitment for populations in which convenience is a big factor of participants into a study and allows for larger volumes of people to be reached instantly as well as data collection. Social networking platforms can be an effective tool for the dissemination and engagement of education information cohorts but require a group of moderators dedicated to the management of the tool, including timely response to comments and questions, consistency of posting as well as in-person support for a behavior such as lactation counselors for breastfeeding.

Methodologically, this social media approach holds promise for both recruitment and conduct of future breastfeeding intervention studies. Repeat studies may consider extending Facebook access pre- and postnatally, new metric analysis such as “Group Insight,” and to consider links to additional, live lactation support services, including racially concordant opportunities. Further analysis and outcomes may be a call to current providers of peer support such as WIC to implement a web-based system to support the mission of WIC and peer counselors.

Acknowledgments

This study was funded by the National Institutes of Health (NIH), National Cancer Institute (NCI) 1R21CA202263. The authors would like to acknowledge the support and collaboration from Katie Constantino and the Catholic Charities WIC program in Western New York; Kaleida Health System, Durham’s Baby Café in Buffalo, New York; Dr Ann Oluloro; and Dr Joseph Denagy. The authors also acknowledge support from the University of Kentucky, Lexington, Kentucky; D’Youville College, Buffalo, New York; and Roswell Park Comprehensive Cancer Center (NIH/NCI P30CA016056), Buffalo, New York.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Youtube video link for “African American women reduce risk of breast cancer by breastfeeding”.
[PNG File , 20 KB - jmir_v22i8e16239_app1.png]

Multimedia Appendix 2
Messaging by topic.
[PDF File (Adobe PDF File), 528 KB - jmir_v22i8e16239_app2.pdf]

References

1. Johnson A, Kirk R, Rosenblum K, Muzik M. Enhancing breastfeeding rates among African American women: a systematic review of current psychosocial interventions. Breastfeed Med 2015;10(1):45-62 [FREE Full text] [doi: 10.1089/bfm.2014.0023] [Medline: 25423601]
2. Breastfeeding Facts. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/breastfeeding/data/facts.html#disparities [accessed 2019-08-01]
3. Palmer JR, Viscidi E, Troester M, Hong C, Schedin P, Bethea T, et al. Parity, lactation, and breast cancer subtypes in African American women: results from the AMBER consortium. J Natl Cancer Inst 2014 Oct;106(10):237 [FREE Full text] [doi: 10.1093/jnci/dju237] [Medline: 25224496]
4. Colen CG, Ramey DM. Is breast truly best? Estimating the effects of breastfeeding on long-term child health and wellbeing in the United States using sibling comparisons. Soc Sci Med 2014 May;109:55-65 [FREE Full text] [doi: 10.1016/j.socscimed.2014.01.027] [Medline: 24698713]
5. Centers for Disease Control and Prevention (CDC). Progress in increasing breastfeeding and reducing racial/ethnic differences - United States, 2000-2008 births. MMWR Morb Mortal Wkly Rep 2013 Feb 8;62(5):77-80 [FREE Full text] [Medline: 23388550]
6. Forste R, Hoffmann JP. Are US mothers meeting the healthy people 2010 breastfeeding targets for initiation, duration, and exclusivity? The 2003 and 2004 national immunization surveys. J Hum Lact 2008 Aug;24(3):278-288. [doi: 10.1177/0890334408317617] [Medline: 18689715]

7. 2017 New York State Pediatric Nutrition Surveillance System Report. New York State Department of Health. 2017. URL: https://www.health.ny.gov/statistics/prevention/nutrition/pednss/2017/ [accessed 2019-08-01]
8. Guendelman S, Kosa J, Pearl M, Graham S, Goodman J, Kharrazi M. Juggling work and breastfeeding: effects of maternity leave and occupational characteristics. Pediatrics 2009 Jan;123(1):e38-e46. [doi: 10.1542/peds.2008-2244] [Medline: 19117845]
9. Cricco-Lizza R. Infant-feeding beliefs and experiences of black women enrolled in WIC in the New York metropolitan area. Qual Health Res 2004 Nov;14(9):1197-1210. [doi: 10.1177/1049732304268819] [Medline: 15448295]
10. Bentley ME, Caulfield L, Gross S, Bronner Y, Jensen J, Kessler L, et al. Sources of influence on intention to breastfeed among African-American women at entry to WIC. J Hum Lact 1999 Mar;15(1):27-34. [doi: 10.1177/089033449901500109] [Medline: 10578772]
11. Ogbuanu CA, Probst J, Laditka SB, Liu J, Baek J, Glover S. Reasons why women do not initiate breastfeeding: a southeastern state study. Womens Health Issues 2009;19(4):268-278 [FREE Full text] [doi: 10.1016/j.whi.2009.03.005] [Medline: 19589476]
12. Kaufman L, Deenadayalan S, Karpati A. Breastfeeding ambivalence among low-income African American and Puerto Rican women in north and central Brooklyn. Matern Child Health J 2010 Sep;14(5):696-704. [doi: 10.1007/s10995-009-0499-5] [Medline: 19644744]
13. O'Brian A, Cadzow RB. Adolescent Perceptions of Breastfeeding and Future Infant Feeding Intentions. DocPlayer. 2015. URL: https://docplayer.net/15622112-Research-and-scholarly-activity-2014-2015.html [accessed 2020-06-26]
14. Perrin A, Turner E. Smartphones Help Blacks, Hispanics Bridge Some – But Not All – Digital Gaps With Whites. Pew Research Center. 2019. URL: https://www.pewresearch.org/fact-tank/2019/08/20/smartphones-help-blacks-hispanics-bridge-some-but-not-all-digital-gaps-with-whites/ [accessed 2019-06-01]
15. Social Media Fact Sheet. Pew Research Center. 2019. URL: https://www.pewinternet.org/fact-sheet/social-media/ [accessed 2019-09-10]
16. Persad MD, Mensinger JL. Maternal breastfeeding attitudes: association with breastfeeding intent and socio-demographics among urban primiparas. J Community Health 2008 Apr;33(2):53-60. [doi: 10.1007/s10900-007-9068-2] [Medline: 18060485]
17. Asiodu IV, Waters C, Dailey D, Lee K, Lyndon A. Breastfeeding and use of social media among first-time African American mothers. J Obstet Gynecol Neonatal Nurs 2015;44(2):268-278 [FREE Full text] [doi: 10.1111/1552-6909.12552] [Medline: 25712127]
18. Baker B, Yang I. Social media as social support in pregnancy and the postpartum. Sex Reprod Healthc 2018 Oct;17:31-34. [doi: 10.1016/j.srhc.2018.05.003] [Medline: 30193717]
19. New York State Community Health Indicator Reports (CHIRS). New York State Prevention Agenda Dashboard - State Level. 2018. URL: https://webbi1.health.ny.gov/SASStoredProcess/guest?_program=/EBI/PHIG/apps/chir_dashboard/chir_dashboard&p=it&ind_id=Ig62 [accessed 2018-09-12]
20. duPont NC, Mahoney M, Kahn L, Vest BM, Widman CA, Clark-Hargrave N, et al. Developing a smoking cessation intervention for low income and minority women. J Womens Health Care 2016 Apr;5(2):309 [FREE Full text] [doi: 10.4172/2167-0420.1000309] [Medline: 28824830]
21. Martin J, Hamilton B, Osterman M, Driscoll A, Drake P. Births: Final Data for 2017. National Vital Statistics Report: Center for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_08-508.pdf [accessed 2020-06-26]

Abbreviations

GED: general educational development
IRB: institutional review board
NCI: National Cancer Institute
NIH: National Institutes of Health
OB/GYN: obstetrics and gynecology
TNBC: triple-negative breast cancer
WIC: Women, Infants, and Children

Edited by G Eysenbach; submitted 16.09.19; peer-reviewed by C Calyx, N Alianmoghaddam, J Loss; comments to author 09.12.19; revised version received 30.01.20; accepted 22.02.20; published 10.08.20.

Please cite as:
Dauphin C, Clark N, Cadzow R, Saad-Harfouche F, Rodriguez E, Glaser K, Kiviniemi M, Keller M, Erwin D
#BlackBreastsMatter: Process Evaluation of Recruitment and Engagement of Pregnant African American Women for a Social Media Intervention Study to Increase Breastfeeding
J Med Internet Res 2020;22(8):e16239
URL: https://www.jmir.org/2020/8/e16239
doi:10.2196/16239
PMID:32773377

Original Paper

Smartphone-Enhanced Symptom Management In Psychosis: Open, Randomized Controlled Trial

Shon Lewis[1], MD; John Ainsworth[1], PhD; Caroline Sanders[1], PhD; Charlotte Stockton-Powdrell[1], MA; Matthew Machin[1], BEng; Pauline Whelan[1], PhD; Richard Hopkins[1], PhD; Zhimin He[2], PhD; Eve Applegate[1], PhD; Richard Drake[1], PhD; Charlie Bamford[1], PhD; Chris Roberts[1], PhD; Til Wykes[2,3], PhD

[1]Manchester Academic Health Sciences Centre, Greater Manchester Mental Health Foundation Trust, The University of Manchester, Manchester, United Kingdom

[2]Institute of Psychiatry, Psychology and Neurosciences, King's College London, London, United Kingdom

[3]South London and Maudsley NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Shon Lewis, MD
Manchester Academic Health Sciences Centre
Greater Manchester Mental Health Foundation Trust
The University of Manchester
Faculty of Biology, Medicine & Health
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 1613067944
Fax: 44 1613067945
Email: shon.lewis@manchester.ac.uk

Abstract

Background: Improving recovery from acute symptoms and preventing relapse are two significant challenges in severe mental illness. We developed a personalized smartphone-based app to monitor symptoms in real time and validated its acceptance, reliability, and validity.

Objective: To assess (i) acceptability of continuous monitoring to SMI patients and health professionals over 3 months; (ii) impact of active self-monitoring on positive psychotic symptoms assessed at 6 and 12 weeks; and (iii) the feasibility of detecting early warning signs of relapse.

Methods: The active symptom monitoring smartphone app was built into an end-to-end system in two NHS Trusts to enable real-time symptom self-monitoring and detection by the clinical team of early signs of relapse in people with severe mental illness. We conducted an open randomized controlled trial of active symptom monitoring compared to usual management to assess: (i) acceptability and safety of continuous monitoring over 3 months; (ii) impact of active self-monitoring on positive psychotic symptoms assessed at 6 and 12 weeks; (iii) feasibility of detecting early warning signs of relapse communicated to the healthcare staff via an app streaming data to the electronic health record. Eligible participants with a Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) diagnosis of schizophrenia and related disorders, and a history of relapse within the previous two years were enrolled from an early intervention team and a community mental health team.

Results: Of 181 eligible patients, 81 (45%) consented and were randomized to either active symptom monitoring or management as usual. At 12 weeks, 90% (33/36) of those in the active monitoring group continued to use the system and exhibited an adherence rate (defined as responding to >33% of alerts) of 84% (30/36}. Active symptom monitoring was associated with no difference on the empowerment scale in comparison to the usual management group at 12 weeks. The pre-planned intent-to-treat analysis of the primary outcome, a positive score on the Positive and Negative Syndrome Scale (PANSS) scale, showed a significant reduction in the active symptom monitoring group over 12 weeks in the early intervention center. Alerts for personalized early warning signs of relapse were built into the workflows of both NHS Trusts, and 100% of health professional staff used the system in a new digital workflow. Qualitative analyses supported the acceptability of the system to participants and staff.

Conclusions: The active smartphone monitoring system is feasible and was accepted by users in a 3-month study of people with severe mental illness, with surprisingly high levels of adherence. App use was associated with psychotic symptom improvement

in recent-onset participants, but not those with longstanding illness, supporting the notion of improved self-management. When built into clinical management workflows to enable personalized alerts of symptom deterioration, the app has demonstrated utility in promoting earlier intervention for relapse.

Trial Registration: ISRCTN Registry ISRCTN88145142; http://www.isrctn.com/ISRCTN88145142

(J Med Internet Res 2020;22(8):e17019) doi:10.2196/17019

KEYWORDS

digital; smartphone; m-health; psychosis; mental health; self management

Introduction

Severe mental illnesses such as schizophrenia often run a relapsing, lifelong course. Persons with severe mental illness have two primary goals: to improve the speed and quality of their recovery and to prevent future relapse. Following the first episode, 70% will have at least one relapse during the next five years [1]. Despite the rise of community care, 40% of the costs of care for a person with severe mental illness are on unplanned inpatient care for relapse [2]. In standard UK practice, contact with health professionals typically occurs only once every 2-6 weeks, so that early signs of relapse are usually picked up too late to enable prompt intervention. Early warning signs of relapse usually comprise the emergence of a mixture of dysphoric symptoms such as anxious mood, and then attenuated psychotic symptoms, appearing over 1-5 days, with insight usually retained until the day of relapse [3].

We developed a smartphone-based platform in 2010 (ClinTouch) to help persons with severe mental illness to manage their symptoms and prevent relapse. Randomized feasibility trials showed this method of active symptom monitoring to be safe, feasible, and acceptable to people with severe mental illness [4,5]. Users with severe mental illness preferred the smartphone app to an equivalent SMS-based version, which took longer to complete (mean 326 seconds versus 68 for the smartphone app [6]).

Having demonstrated the proof of concept, we integrated the standalone smartphone system (ClinTouch) via an application programming interface (API) into NHS Trust information and communication technology platforms. This integration enabled the streaming of summary information into electronic health records, enabling health professionals to track current symptoms on desktops at the team base and receive personalized alerts when symptoms exceeded a pre-agreed threshold.

This report describes an open randomized controlled trial of smartphone-based active symptom management versus usual care to assess the (i) acceptability and safety of continuous monitoring in persons with severe mental illness and health professionals over 3 months, (ii) impact of active self-monitoring on positive psychotic symptoms assessed at 6 and 12 weeks, and (iii) feasibility of detecting early warning signs of relapse.

Methods

Study Design

The trial of ClinTouch active symptom management versus management as usual was a two-center, open, randomized controlled trial at the NHS Mental Health Trusts in Manchester and South London. Software development, beta testing, and prior cohort and smaller randomized trials had used an experience-driven design process in which service users with severe mental illness were involved in all stages of the design and development of the app, its functionality, and its standard operating procedures. Health professionals were included in design issues where they related to the use of the system within routine practice and in the design of new, digitally-enabled workflows. In preparing for the current trial, 6 focus groups were conducted and audiotaped, including a total of 23 service users, 5 carers, and 30 healthcare staff. Qualitative in-depth interviews were conducted with 19 service users, 6 carers, and 17 staff. A Service User and Carer advisory group met quarterly throughout the project and provided advice on study design, information for participants, and related issues.

The personalized smartphone app triggers the user to rate their symptoms several times a day, and wirelessly uploads these in real time to a secure central server. An audio cue triggered semi-randomly 2-4 times a day reminds the user to complete a set of 12-14 branching items about current symptom severity using a touchscreen slider. A graphical summary of how symptoms fluctuate over time is assembled and displayed on the handset. By conducting face-to-face interview assessments using the gold standard Positive and Negative Syndrome Scale (PANSS) [7] before and after one week of 4 times daily ClinTouch assessment, we confirmed the validity of the self-reported items. Core psychotic symptom and mood items showed moderate to strong (r>0.6) correlations between the in-person and self-report methods [5]. Non-core, behaviorally assessed items such as negative symptoms showed weaker correlations.

The trial was approved by the South Birmingham NHS Research Ethics Committee (14/WM/0045). The trial was registered with the National Institute of Health Research CRN portfolio: 16361, and ISRCTN 88145142. The Medicines and Healthcare Regulatory Agency elected not to designate ClinTouch as a medical device as deployed for the trial.

Participants

One community clinical team from each Trust participated. In England, community mental health teams serve a geographically defined catchment area. All Trusts use electronic patient record systems. Management of individual service users is coordinated by a mandatory care coordinator, usually with a nursing background. Each team has one consultant psychiatrist working with psychologists and other mental health professionals. Teams typically have caseloads of 200-400 people with severe mental

illness, with 20-30 care coordinators. In the South London Trust, an Early Intervention for Psychosis (EIP) team was selected and in the Manchester Trust, a Community Mental Health Team (CMHT). NHS EIP teams are configured to manage care for people in the first three years after the first psychotic episode, after which care is transferred to a CMHT. Different types of teams were chosen a priori to investigate the effect of duration of illness on any response to digital treatment. Recruitment took place between February 2014 and May 2015.

Participant inclusion criteria were: (i) operational Diagnostic and Statistical Manual 8th Edition DSM-IV [8] diagnosis of schizophrenia and related disorders; (ii) aged 16-65; (iii) one or more psychotic episodes in the previous 2 years, including the first psychotic episode. Exclusion criteria were: (i) unable to speak English; (ii) unable to give informed consent. Patients who met these criteria were identified separately in the two clinical teams.

Randomization

Participants were allocated by computer using randomized, permuted blocks to one of two groups: active symptom monitoring plus management as usual, or management as usual alone, each for 12 weeks. No stratification was used.

Procedure/Intervention

There were two linked interventions. Active symptom monitoring with feedback to participants was aimed at encouraging self-management of symptoms. Alerts were fed back to the care coordinator when personalized early warning sign thresholds were exceeded, allowing very early intervention.

The ClinTouch active symptom management system was integrated into the electronic health record (EHR) platform of the Manchester NHS Trust via an application programming interface (API). The API instructed the EHR to retrieve data from the ClinTouch dashboard, including a list of participants using the ClinTouch system and any alerts associated with them. The EHR system displayed information relating to the ClinTouch data within the individual patient record. Care coordinators and clinicians were given secure individual logins to the ClinTouch system, enabling them to view the data on a desktop along with graphs of symptom changes over time. The EHR provider for the South London Trust denied API access. To compensate for this lack of access, an automated email was sent to the appropriate care coordinator whenever a patient alert was raised. The email solution was also employed in the Manchester Trust.

Frontline clinical staff (n=42) were trained in the use of ClinTouch. For patients randomized to the experimental group, the relevant care coordinator delivered training in the use of the handset. Either the Android app was installed on the participant's phone, or a preconfigured Samsung Galaxy smartphone was provided on loan for the duration of the study. The branching items covered positive psychotic symptoms, anxiety, and mood as validated against the PANSS scale in previous studies. Semi-random twice-daily auditory cues from the handset prompted symptom data collection and wireless upload. The system was then used for 12 weeks in the context of the preexisting care plan modified for the ClinTouch algorithms.

Care coordinators then determined the criteria for each participant's early warning signs, using previous electronic patient records for reference. The early warning sign threshold was set by assigning a score of 0-3 (low through high) to each symptom according to how relevant it is for that participant's relapse signature based on previous experience. The alert algorithm was constructed so that symptoms scored as 1 collectively comprised 20% of the total early warning sign score, those scoring 2 comprised 30%, and those scoring 3 comprised 50%. An alert was generated if the total score for a single datapoint rose to 40% higher than the baseline defined as the mean score for the first 3 days of recording, or 25% higher across two consecutive data points. Operationally, this was done as part of the standard crisis planning meeting.

Standard operating procedures were established for recording and handling adverse events. Technical measures to ensure data privacy and patient confidentiality followed industry-standard best practices, and all data communications between app and server used encrypted channels. Data were handled per the UK Data Protection Act 1998.

Follow-up and Assessments

Participants were assessed in an in-person interview at baseline, then 6 and 12 weeks after randomization. Research assistants trained to criterion inter-rater reliability undertook participant assessments.

Feasibility and acceptability outcomes in the experimental group were two-fold. The client-centered outcomes included the proportion of eligible clients consenting to a trial of ClinTouch active symptom management. We predicted that 50% would remain in follow-up for 12 weeks. We predicted that 50% of participants would complete >33% of all possible symptom self-ratings over the 12-week trial. The clinical team outcomes included the proportion of all care coordinators accessing patients' online symptom data. Adverse effects were routinely monitored during the weekly telephone support calls to participants.

Primary efficacy endpoints over 12 weeks included (i) Score on the positive symptom subscale of the PANSS, (ii) user empowerment from interviews, and the Empowerment Rating Scale [9]. Secondary efficacy outcomes were (i) Calgary Depression Scale [10], (ii) Global Assessment of Functioning scale (GAF) [8], and (iii) health-related quality of life, the EuroQol 5D (EQ5D) [11]. These face-to-face interviews were recorded in hard copy versions of the rating scales and the data stored securely in accord with Medical Research Council guidance.

In order to gain an estimate of how frequently clinical staff recorded episodes of possible early warning signs independently of the active symptom monitoring system, transcripts of electronic care records for the 12 weeks of the trial plus a further 4 weeks were anonymized and any reference to randomized treatment redacted. These were then rated independently by two experienced clinicians (SL, RH) for the documented occurrence

of emergent symptoms, which met early warning criteria of documented worsening of psychotic symptoms.

Qualitative interviews were conducted in a subsample of those declining to participate and those allocated to ClinTouch at exit.

Statistical Analysis

The effect of ClinTouch-enhanced monitoring on PANSS Positive Subscale totals at follow-up was examined using analysis of covariance (ANCOVA), including allocation group and site (Manchester or London) as cofactors and baseline scores as a covariate, using Stata 14.1 (College Station). The teams at each site were selected purposely so that differences in response between young, recent onset participants (London) and older, more chronically unwell participants (Manchester) could be examined. Sensitivity analyses examined the effect of demographic variables (covariates were sex, age, level of qualifications, ethnic minority status, living independently, being single, unemployed, in current psychotherapy or abusing alcohol) using backward stepwise elimination of associations of $P>.20$. A comparison of individual general linear models for the two sites was pre-planned to examine the likely differences. Finally, secondary analyses of other PANSS subtotals and total were conducted in the same way as the primary analysis.

The sample size was calculated based on a 50% reduction in early warning signs in the experimental treatment arm over 12 weeks, from 40% to 20%. Assuming a 10% drop out rate, a sample size of 72 would have 80% power to detect this difference with a one-sided alpha of 0.2, as recommended for a feasibility trial. The analysis was by intent to treat ANCOVA using STATA, with data at baseline, then 6 and 12 weeks.

Results

Recruitment and Feasibility

Of 181 eligible service users approached, 81 (46%) consented to participate and were randomized to either ClinTouch-enhanced management or management as usual (see Tables 1 and 2, Figure 1). There were substantial demographic differences between sites (see Table 2), as intended and expected. The CMHT participants (Manchester) were older and chronically unwell (mean 46 years; median 2.5 hospital admissions, IQR 1 to 4) than the EIT participants (mean 26 years; London: median 1 admission, IQR 0 to 1). Of those 40 who were randomized to the ClinTouch-enhanced management arm, 38 (95%) stayed in the trial for 12 weeks. Of these 38, acceptable adherence as defined by responding to at least 33% of beep alerts (four-item sets per day) was 84%, good adherence (greater than 50% of alerts) was 60%. Healthcare professionals (care coordinators) used ClinTouch-enhanced management in 100% of cases, accessing ClinTouch data an average of 24 times per patient.

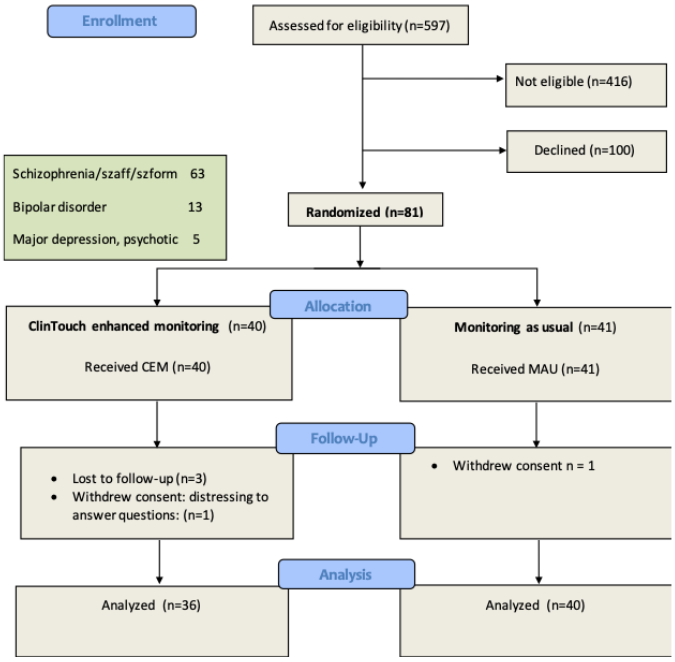
Table 1. Demographic data by treatment group.

Descriptor	ClinTouch enhanced monitoring plus standard care	Standard care
Number	40	41
Age (years), mean (range)	33.7 (21-61)	35.3 (20-68)
Sex - female	11	16
Ethnicity - white	20	23
Ethnicity - black/black British	17	15
Ethnicity - other	3	3

Table 2. Demographic data by site.

Descriptor	Community mental health (Manchester)	Early psychosis (London)
Number	37	44
Active symptom monitoring treatment arm	18	22
Age (years), mean (range)	46.1 (21-68)	26.1 (19-36)
Sex - female	14	13
Ethnicity - white	31	12
Ethnicity - black/black British	6	26
Ethnicity - other	0	6

Figure 1. Randomized controlled trial flow diagram.



Safety

Adverse effects were routinely monitored during weekly telephone support calls. Of 38 participants who completed 12 weeks of the trial, three (8%) reported significant events: 1 reported increased anxiety prompted by questions; 1 reported increased irritation due to the alert beeps, and 1 had their charger explode. All 3 continued to complete the 12 weeks in the trial.

Clinical efficacy

There were no substantial differences in symptom severity at the point of randomization between those allocated to ClinTouch-enhanced monitoring or standard care (Table 3). On the primary efficacy outcomes, there was no significant difference between groups in PANSS Positive total after 6 or 12 weeks, nor were there significant differences in secondary outcomes (Table 3). Sensitivity analysis showed that including demographic variables made no substantial difference to the allocation group's coefficient or significance.

The planned analyses of each site separately demonstrated different outcomes in the different services. Although there were no significant differences between ClinTouch-enhanced monitoring and control participants in Manchester (apart from a difference in depression scores identifiable at baseline and persisting without significant alteration during the trial; Table 4), findings in London were different (Table 5). There was a significant reduction in positive symptoms after 12 weeks of ClinTouch-enhanced monitoring in the early psychosis subsample (adjusted mean difference –3.04; CI –5.49, –0.59; *P*=.016. Although there was a significant site-by-group interaction for PANSS total (Supplementary Table 1; *P*=.003), indicating a significantly lower PANSS total after 12 weeks of ClinTouch-enhanced monitoring in the early psychosis center, this benefit was not in itself significant (adjusted mean difference –5.83; CI –14.14, 2.48; *P*=.164 2-tailed). There were no other significant site-by-group differences. In addition to the conventional rating scales, the ClinTouch device provided real-time individual active symptom data, which indicated that over 12 weeks, all symptoms except one declined in mean severity. Severity of hallucinations decreased by 29%.

The frequency of early warning signs, as documented in electronic patient records, was 33% in the CEM group and 46% in the control group over 12 weeks, after excluding 8 cases where records were too scant to be rated. The actual performance of this early prototype of the Early Warning Signs algorithm was suboptimal, in terms of the accuracy of ClinTouch alerts versus early warning signs as contemporaneously documented in the electronic patient record. Sensitivity was 75%, specificity 8%, giving a positive predictive value of 29%.

Table 3. Clinical Measures at baseline, 6, and 12 weeks by allocation group.

Scale and visit	CareLoop enhanced monitoring, mean (SD)	Management as Usual, mean (SD)	Adjusted mean difference[a]	95% CI	*P* value 2-tailed	*P* value 1-tailed	Intercept site*trial arm *P*
PANSS[b] Total							
Baseline	72.9 (14.8)	76.8 (17.4)					
Weeks 6	70.7 (17.0)	73.9 (20.7)	–0.47	–6.47 to 5.53	.874	.44	.46
Week 12	64.5 (15.7)	69.3 (20.7)	–1.93	–7.50 to 3.64	.492	.25	.003
PANSS Positive							
Baseline	18.8 (5.4)	18.3 (5.7)					
Weeks 6	17.3 (6.2)	17 (6.2)	–0.37	–2.35 to 1.60	.708	.35	.34
Week 12	16 (5.3)	16.7 (6.2)	–1.13	–3.12 to .87	.264	.13	.057
PANSS Negative							
Baseline	18.8 (4.3)	18.3 (5.5)					
Weeks 6	16.1 (4.4)	18.2 (5.7)	–0.44	–2.18 to 1.30	.616	.31	.75
Week 12	15 (4.4)	17.1 (5.6)	–0.69	–2.51 to 1.15	.462	.23	.53
PANSS General							
Baseline	38.2 (8.7)	40 (9.2)					
Weeks 6	37.4 (9.5)	38.7 (10.9)	–0.17	–3.70 to 3.37	.994	.47	.64
Week 12	33.5 (8.6)	35.5 (10.7)	–0.79	–3.86 to 2.29	.611	.31	.38
ERS[c] Total							
Baseline	86.3 (7.4)	81.4 (7.8)					
Weeks 6	85.4 (7.6)	81.6 (10.3)	0.58	–2.99 to 4.15	.748	.37	.47
Week 12	86.5 (11.9)	83.6 (8.1)	–0.05	–4.35 to 4.25	.983	.49	.32
EQ5D[d] Total							
Baseline	8.8 (3.1)	9.6 (4.1)					
Weeks 6	9.2 (3.4)	8.8 (4.2)	–0.29	–2.43 to 1.85	.286	.14	.29
Week 12	8.0 to 4.1	8.4 (3.8)	0.15	–1.23 to 1.53	.812	.41	.15
CDS[e] Total							
Baseline	5.8 to 4.6	8.1 (5.6)					
Weeks 6	6 to 4.5	7.3 (5.2)	0.29	–1.30 to 1.90	.712	.36	.03
Week 12	4.6 to 3.7	6.5 (4.8)	–0.67	–2.24 to 0.90	.4	.20	.83
GAF[f]							
Baseline	49.7 to 14.9	49.3 (11.8)					
Weeks 6	49.2 to 14.5	47.7 (16.7)	–0.62	–6.45 to 5.21	.595	.23	.90
Week 12	51.8 to 13.7	52.2 (16.2)	–2.65	–8.38 to 3.07	.850	.43	.30

[a]Follow-up differences adjusted for baseline scores and the main effect of site.

[b]PANSS: Positive and Negative Syndrome Scale

[c]ERS: Empowerment Rating Scale

[d]EQ5D: EuroQol-5D

[e]CDS: Calgary Depression Scale

[f]GAF: Global Assessment of Functioning

Table 4. Clinical measures at baseline, 6 weeks, and 12 weeks: community team sample.

Scale and visit	CareLoop enhanced monitoring, mean (SD)	Management as usual, mean (SD)	Adjusted mean difference[a]	95% CI	*P* value, 2-tailed
PANSS[b] Total					
Baseline	72.73 (11.71)	78.32 (19.02)			
Weeks 6	69.47 (17.43)	72.68 (22.53)	1.30	–8.98 to 11.58	.80
Week 12	57.84 (14.23)	68.59 (22.22)	–5.83	–14.14 to 2.48	.16
PANSS Positive					
Baseline	19 (4.22)	19.36 (6.12)			
Weeks 6	16.42 (5.67)	17.41 (6.65)	–0.95	–3.91 to 2.01	.52
Week 12	14.11 (4.10)	17.18 (6.27)	–3.04	–5.49 to –0.59	.02
PANSS Negative					
Baseline	16.09 (4.21)	18.77 (5.52)			
Weeks 6	15.42 (4.50)	17.32 (6.12)	0.004	–2.83 to 2.92	.98
Week 12	13.47 (4.58)	16.45 (6.10)	–0.76	–3.36 to 1.85	.56
PANSS General					
Baseline	37.64 (7.49)	40.18 (10.13)			
Weeks 6	37.63 (10.12)	37.95 (11.46)	1.43	–4.51 to 7.36	.63
Week 12	30.26 (8.55)	34.95 (11.15)	–2.67	–7.46 to 2.13	.27
ERS[c] Total					
Baseline	86.73 (5.59)	83.27 (6.53)			
Weeks 6	85.74 (5.69)	83.68 (8.16)	0.13	–4.18 to 4.43	.95
Week 12	86.26 (6.15)	85.91 (8.07)	–1.26	–5.80 to 3.29	.58
EQ5D[d] Total					
Baseline	8.09 (2.81)	8.64 (3.09)			
Weeks 6	6.91 (3.53)	8.23 (3.16)	1.00	–0.78 to 2.79	.26
Week 12	6.45 (3.49)	7.00 (2.31)	0.33	-1.35 to 0.83	.20
CDS[e] Total					
Baseline	5.82 (4.16)	8.5 (5.86)			
Weeks 6	6.63 (4.21)	6.73 (4.46)	1.60	–0.486 to 3.67	.13
Week 12	4.21 (3.03)	6.27 (4.38)	–0.85	–2.94 to 1.25	.42
GAF[f]					
Baseline	42.38 (12.04)	38.4 (6.42)			
Weeks 6	44.47 (13.29)	41.91 (16.25)	–0.11	–9.40 to 9.17	.98
Week 12	48.47 (11.75)	45.14 (18.93)	–1.09	–10.05 to 7.00	.80

[a]Follow-up differences adjusted for baseline scores.

[b]PANSS: Positive and Negative Syndrome Scale

[c]ERS: Empowerment Rating Scale

[d]EQ5D: EuroQol-5D

[e]CDS: Calgary Depression Scale

[f]GAF: Global Assessment of Functioning

Table 5. Clinical measures at baseline, 6, and 12 weeks: First episode psychosis sample.

Scale and visit	CareLoop enhanced monitoring, mean (SD)	Management as usual, mean (SD)	Adjusted mean difference[a]	95% CI	*P* value, 2 tailed
PANSS[b] Total					
Baseline	73.11 (18.16)	75.00 (15.72)			
Weeks 6	72.06 (16.86)	75.44 (18.77)	–2.84	–8.81 to 3.12	.34
Week 12	71.50 (14.31)	70.11 (19.24)	2.69	–4.81 to 10.19	.47
PANSS Positive					
Baseline	18.44 (6.66)	17.16 (5.1)			
Weeks 6	18.24 (6.82)	16.56 (5.79)	2.67	–2.48 to 3.01	.84
Week 12	18.06 (5.86)	16.06 (6.35)	1.10	–2.16 to 4.35	.498
PANSS Negative					
Baseline	15.83 (4.48)	17.95 (5.69)			
Weeks 6	16.76 (4.28)	19.28 (5.04)	–1.03	–3.01 to .94	.30
Week 12	16.5 (3.73)	18.06 (4.83)	–0.50	–3.09 to 2.10	.70
PANSS General					
Baseline	38.83 (10.17)	39.84 (8.34)			
Weeks 6	37.06 (9.13)	39.61 (10.32)	–2.32	–5.85 to 1.23	.19
Week 12	36.94 (8.91)	36.11 (10.44)	1.33	–2.60 to 5.26	.495
ERS[c] Total					
Baseline	85.83 (9.32)	79.26 (8.84)			
Weeks 6	85 (9.52)	78.94 (12.25)	1.40	–4.73 to 7.53	.65
Week 12	86.72 (16.05)	80.72 (7.43)	1.63	–6.25 to 9.50	.68
EQ5D[d] Total					
Baseline	9.5 (3.59)	11.21 (4.60)			
Weeks 6	9.22 (4.25)	10.42 (4.86)	0.10	–2.48 to 2.69	.94
Week 12	8.06 (7.38)	6.116 (5.17)	–1.40	–5.72 to 2.02	.51
CDS[e] Total					
Baseline	5.72 (5.21)	7.68 (5.23)			
Weeks 6	5.29 (4.81)	8.06 (6.00)	–1.36	–0.98 to 3.69	.25
Week 12	5.06 (4.41)	6.83 (5.48)	–0.63	–1.81 to 3.07	.61
GAF[f]					
Baseline	56.94 (13.12)	53.84 (12.30)			
Weeks 6	52.64 (14.52)	51.56 (16.79)	–1.19	–8.51 to 6.15	.74
Week 12	54.11 (14.65)	57.17 (10.78)	–5.09	–12.38 to 2.20	.17

[a]Follow-up differences adjusted for baseline scores.

[b]PANSS: Positive and Negative Syndrome Scale

[c]ERS: Empowerment Rating Scale

[d]EQ5D: EuroQol-5D

[e]CDS: Calgary Depression Scale

[f]GAF: Global Assessment of Functioning

Discussion

We conducted an open randomized controlled trial of active symptom monitoring compared to usual management in people with serious mental illness to assess over 12 weeks the (i) acceptability and safety of continuous monitoring, (ii) impact of active self-monitoring on positive psychotic symptoms, and (iii) feasibility of detecting early warning signs of relapse

communicated to the healthcare staff via an API allowing data to be streamed into the EHR.

A systematic review has suggested that smartphone apps may be helpful in the management of mental health disorders such as depression [12]. Nonetheless, almost none of the publicly available mental health apps have good quality data concerning safety and efficacy [13]. The real-time digital approach used in this study holds several advantages over routine clinical assessment. It reduces the confounding effects of retrospective recall bias, forgetting, and averaging in symptom appraisal. It allows the context of symptom changes to be assessed and increases patient involvement in continuing care through participation in symptom and progress monitoring. It may also enable a degree of symptom self-management via a trusted and ubiquitous, ever-present personal device.

The trial demonstrated several things. The active symptom monitoring intervention was safe and acceptable: 45% of the eligible sample agreed to enter the trial. Furthermore, and importantly, of those using the ClinTouch-enhanced monitoring system, 90% continued to use it regularly at 3 months. In these patients, adequate adherence was 84%, defined as responding to >33% of item prompts. On pre-planned intent-to-treat analysis, the primary outcome of positive symptom score on the PANSS scale showed a significant reduction in the ClinTouch group over 12 weeks only in the early intervention center. The larger therapeutic effect in the early psychosis participants was not due to the severity or adherence differences between the two subsamples. It may be that, as has been shown with pharmacological and psychological treatments for psychosis, the therapeutic effect is larger earlier in the course of the disorder.

We have demonstrated from a software perspective that we can build an algorithm into the ClinTouch app to provide an alert when symptoms start to worsen. An API allowed this to be built into the electronic patient record system in one Trust. With symptom data streamed into the EHR system, health professionals could view it on a secure desktop at the team base. Alerts for early warning signs were built into the workflows of the two NHS Trusts, and 100% of health professional staff used the system to access symptom data and alerts in a new digital workflow. Qualitative analyses supported the acceptability of the system to participants and staff.

There were limitations to the trial. In the second Trust, the commercial provider of the EHR did not comply with the study, indicating a potential barrier to full scale roll out in the NHS where Trusts have a range of different commercially provided EHR platforms. Another limitation was that, at the time of the trial (2014-2016), the ClinTouch app was only available for the Android operating system. In addition, the accuracy of the early prototype in detecting EWS was limited by our focus being mainly on operability. Case record documentation of EWS was often scanty, proving to be an inadequate gold standard. Artifacts in functionality were identified for improvement, such as alerts being mistimed if the user was temporarily in an area without a wireless network. Subsequent versions are proving more refined. Further work is now taking place to refine the alert algorithm through robust risk prediction modeling in order to increase its sensitivity and specificity and improve the effectiveness of promoting early intervention by clinical teams to improve patient outcomes.

In conclusion, the active smartphone monitoring system is feasible and acceptable over three months to users with severe mental illness, with surprisingly high levels of adherence both from users and health professionals. It was associated with psychotic symptom improvement in patients with recent-onset psychosis, and supports the notion of improved self-management in those with first episode psychosis. In terms of implications for clinical practice, digital health interventions appear to hold considerable promise in the management of people with psychosis. Smartphone-based active symptom monitoring can be built into EHR systems and regular clinical workflows and allow preventive, personalized care, especially if combined in with added digital functionality such as medication management and physical health monitoring.

Acknowledgments

This study was funded by Medical Research Council Developmental Clinical Studies grant MR/K015516. The Medical Research Council had no role in the study design, data collection, analysis, interpretation, or writing of the report. The corresponding author had full access to all data and had final responsibility for the decision to submit for publication.

Conflicts of Interest

SL is the Director of Affigo, a not for profit social enterprise digital company and the Medical Director of Xenzone (remunerated), a digital counselling company. JA, CS-P, and PW are Directors of Affigo CIC, which is a community interest company set-up in 2015 to make ClinTouch more widely available to mental health Trusts. The IP for ClinTouch has been assigned to Affigo. Our company registration number is 09928775.

Multimedia Appendix 1
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 957 KB - jmir_v22i8e17019_app1.pdf]

References

1. Robinson D, Woerner MG, Alvir JM, Bilder R, Goldman R, Geisler S, et al. Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. Arch Gen Psychiatry 1999 Mar;56(3):241-247. [doi: 10.1001/archpsyc.56.3.241] [Medline: 10078501]
2. Mangalore R, Knapp M. Cost of schizophrenia in England. J Ment Health Policy Econ 2007 Mar;10(1):23-41. [Medline: 17417045]
3. Birchwood M, Spencer E, McGovern D. Schizophrenia: early warning signs. Adv. psychiatr. treat 2018 Jan 02;6(2):93-101. [doi: 10.1192/apt.6.2.93]
4. Palmier-Claus JE, Rogers A, Ainsworth J, Machin M, Barrowclough C, Laverty L, et al. Integrating mobile-phone based assessment for psychosis into people's everyday lives and clinical care: a qualitative study. BMC Psychiatry 2013 Jan 23;13:34 [FREE Full text] [doi: 10.1186/1471-244X-13-34] [Medline: 23343329]
5. Palmier-Claus JE, Ainsworth J, Machin M, Barrowclough C, Dunn G, Barkus E, et al. The feasibility and validity of ambulatory self-report of psychotic symptoms using a smartphone software application. BMC Psychiatry 2012 Oct 17;12:172 [FREE Full text] [doi: 10.1186/1471-244X-12-172] [Medline: 23075387]
6. Ainsworth J, Palmier-Claus JE, Machin M, Barrowclough C, Dunn G, Rogers A, et al. A comparison of two delivery modalities of a mobile phone-based assessment for serious mental illness: native smartphone application vs text-messaging only implementations. J Med Internet Res 2013 Apr 05;15(4):e60 [FREE Full text] [doi: 10.2196/jmir.2328] [Medline: 23563184]
7. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987;13(2):261-276. [doi: 10.1093/schbul/13.2.261] [Medline: 3616518]
8. American Psychiatric Association. Diagnostic and Statistical Manual - 4th Edition. Washington, DC: American Psychiatric Association; 2000.
9. Rogers ES, Ralph RO, Salzer MS. Validating the empowerment scale with a multisite sample of consumers of mental health services. Psychiatr Serv 2010 Sep;61(9):933-936. [doi: 10.1176/ps.2010.61.9.933] [Medline: 20810594]
10. Addington D, Addington J, Maticka-Tyndale E. Assessing depression in schizophrenia: the Calgary Depression Scale. Br J Psychiatry Suppl 1993 Dec(22):39-44. [Medline: 8110442]
11. The EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9]
12. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
13. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: Evaluation of mental health app store quality claims. NPJ Digit Med 2019;2:18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]

Abbreviations

ANCOVA: analysis of covariance
API: application programming interface
CDS: Calgary Depression Scale
CMHT: Community Mental Health Team
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition
EHR: electronic health record
EIP: Early Intervention for Psychosis
EQ5D: EuroQol 5D
GAF: Global Assessment of Functioning
PANSS: positive and negative syndrome scale

Edited by G Eysenbach; submitted 13.11.19; peer-reviewed by H Mehdizadeh, T Timmers; comments to author 27.12.19; revised version received 04.02.20; accepted 16.02.20; published 13.08.20.

Please cite as:
Lewis S, Ainsworth J, Sanders C, Stockton-Powdrell C, Machin M, Whelan P, Hopkins R, He Z, Applegate E, Drake R, Bamford C, Roberts C, Wykes T
Smartphone-Enhanced Symptom Management In Psychosis: Open, Randomized Controlled Trial
J Med Internet Res 2020;22(8):e17019
URL: https://www.jmir.org/2020/8/e17019
doi:10.2196/17019
PMID:32788150

Original Paper

Examining Responsiveness to an Incentive-Based Mobile Health App: Longitudinal Observational Study

Jacob Brower[1], PhD; Monica C LaBarge[1], PhD; Lauren White[2], MSc; Marc S Mitchell[3], PhD

[1]Smith School of Business, Queen's University, Kingston, ON, Canada

[2]Carrot Insights Inc, Toronto, ON, Canada

[3]School of Kinesiology, Western University, London, ON, Canada

Corresponding Author:
Jacob Brower, PhD
Smith School of Business
Queen's University
805 Johnson St
Kingston, ON, K7L2B6
Canada
Phone: 1 6132174726
Email: jbrower@queensu.ca

Abstract

Background: The Carrot Rewards app was developed as part of a public-private partnership to reward Canadians with loyalty points for downloading the app, referring friends, completing educational health quizzes, and health-related behaviors with long-term objectives of increasing health knowledge and encouraging healthy behaviors. During the first 3 months after program rollout in British Columbia, a number of program design elements were adjusted, creating observed differences between groups of users with respect to the potential impact of program features on user engagement levels.

Objective: This study examines the impact of reducing reward size over time and explored the influence of other program features such as quiz timing, health intervention content, and type of reward program on user engagement with a mobile health (mHealth) app.

Methods: Participants in this longitudinal, nonexperimental observational study included British Columbia citizens who downloaded the app between March and July 2016. A regression methodology was used to examine the impact of changes to several program design features on quiz offer acceptance and engagement with this mHealth app.

Results: Our results, based on the longitudinal app use of 54,917 users (mean age 35, SD 13.2 years; 65.03% [35,647/54,917] female), indicated that the key drivers of the likelihood of continued user engagement, in order of greatest to least impact, were (1) type of rewards earned by users (eg, movies [+355%; P<.001], air travel [+210%; P<.001], and grocery [+140%; P<.001] relative to gas), (2) time delay between early offers (−64%; P<.001), (3) the content of the health intervention (eg, healthy eating [−10%; P<.001] vs exercise [+20%, P<.001] relative to health risk assessments), and (4) changes in the number of points offered. Our results demonstrate that reducing the number of points associated with a particular quiz by 10% only led to a 1% decrease in the likelihood of offer response (P<.001) and that each of the other design features had larger impacts on participant retention than did changes in the number of points.

Conclusions: The results of this study demonstrate that this program, built around the principles of behavioral economics in the form of the ongoing awarding of a small number of reward points instantly following the completion of health interventions, was able to drive significantly higher engagement levels than those demonstrated in previous literature exploring the intersection of mHealth apps and financial incentives. Previous studies have demonstrated the presence of incentive matters to user engagement; however, our results indicate that the number of points offered for these reward point–based health interventions is less important than other program design features such as the type of reward points being offered, the timing of intervention and reward offers, and the content of the health interventions in driving continued engagement by users.

(J Med Internet Res 2020;22(8):e16797) doi:10.2196/16797

KEYWORDS

mHealth; behavioral economics; public health; incentives; mobile apps; mobile phone

Introduction

If modifiable chronic disease risk factors (eg, physical inactivity, unhealthy eating) were eliminated, 80% of both ischemic heart disease and type 2 diabetes and 40% of cancers could be prevented [1]; consequently, modest population-level improvements can make a big difference. For instance, a 1% reduction in the proportion of Canadians accumulating less than 5000 daily steps would yield Can $2.1 billion (US $1.615 billion) per year in health care system savings [2]. Such health behaviors, though, are notoriously difficult to stimulate and sustain, with persistent global overweight and obesity rates providing cases in point [1]. The World Health Organization [3] and others [4] suggest that broader socioecological solutions with interventions delivered at multiple levels (eg, individual, community, societal) are needed to address this issue. At the individual level, smartphones have revolutionized health promotion [5]. Their pervasiveness (eg, 1 billion global smartphone subscriptions expected by 2022) [6] and rapidly evolving functionalities (eg, built-in accelerometers, geolocating capabilities, machine learning techniques) have made it easier to deliver more timely and personalized health interventions on a mass scale.

Mobile Health Apps

The smartphone-based mobile health (mHealth) app market has grown steadily in recent years. In 2017, for example, there were 325,000 mHealth apps available on all major app stores, up by 32% from the previous year [7]. The number of mHealth app downloads also increased by 16% from 2016 to 2017 (3.2 to 3.7 billion) [7]. Although both supply (apps published in stores) and demand (app downloads) is growing, low engagement (with engagement measured as repeated usage, consistent with behavioral science approaches [8,9]) has resulted in small effect sizes and presented hurdles for financial sustainability continues to be a challenge for the industry [10-13]. For instance, 90% of all mHealth apps are uninstalled within 30 days, and 83% of mHealth app companies have fewer than 10,000 monthly active users, a standard industry engagement metric [14]. Systematic reviews of controlled studies on this topic suggest that tailoring content to individual characteristics, regularly updating apps, and incorporating a range of behavior change techniques, for example, may boost engagement [10-13]. To date, however, evaluations of only a few mHealth apps out of the thousands in the app stores have been published in peer-reviewed scientific journals [12]. To better elucidate the conditions under which mHealth app interventions are likely to succeed in real-world settings, more applied research is needed [5]. Traditional randomized controlled trials can be difficult to conduct in a fast-paced digital health environment, but mHealth has benefited from innovative approaches which attempt to determine causal mechanisms for outcomes of intervention effectiveness through methods such as microrandomized trials, factorial designs, and quasiexperimental designs such as pre-post, inverse roll-out, and interrupted time series [15]. Given the widespread proliferation of mHealth apps, evaluation methods that examine what does or does not work in the field, even when not employing methodologies that may generate interpretations of causality, may still contribute to our understanding of the contextual (eg, population characteristics) and program (eg, intervention design) factors that impact engagement, which ultimately influence the effectiveness both in terms of the financial cost of interventions as well as measurable impacts on health. This study, therefore, responds to a call in the literature to create a more comprehensive understanding of contextual and program factors that may impact engagement.

Study Context

The Carrot Rewards app (Carrot), created by a private company with support from the Public Health Agency of Canada [16], presents a unique research opportunity to explore the effects of some of these factors. Carrot was a Canadian app (ie, 1 million downloads, 500,000 monthly active users) grounded in behavioral economics, an offshoot of traditional economic theory complemented by insights from psychology [17]. Briefly, behavioral economics has demonstrated how small changes in the decision environment, particularly those that align with so-called economic *rationality* by cueing individuals' financial goals, can have powerful effects on behavioral change at both individual and societal levels [17]. In exchange for engaging in short educational quizzes on a range of public health topics, the app rewarded users with loyalty points from 4 major Canadian loyalty program providers, which can be redeemed for popular consumer products such as groceries, air travel, movies, or gas. Although monetary health incentives (eg, paying people to walk more) have shown promise with evidence of short- and long-term effects [18], only a limited amount of research has examined alternative types of financial incentives [19]. Research in consumer psychology and decision making on individuals' responsiveness to loyalty points in particular suggests that they are overvalued by consumers in general [20] and that the way individuals behave with respect to accumulating and spending points is nonlinear [21,22]. This behavior is idiosyncratic to factors such as the effort required to earn the reward [23] and the computational ease with which individuals are able to translate points to equivalent dollar values [24]. In public health campaigns, when large financial incentives are unlikely to be suitable or sustainable [25], opportunities for more financially feasible types of incentives are worthy of further study. In addition, a robust understanding of the likely effectiveness of such programs requires a more nuanced examination of program factors such as which individuals are likely to respond to these types of interventions as well as what program design features (eg, size and timing of rewards) [26] are influential in maintaining engagement with the platform.

In the first few weeks after its launch in 2016, Carrot underwent several important program changes, resulting in a nonexperimental observational study [27] that can be examined to shed light on factors influencing engagement. Our primary objective was to examine the impact of reducing reward size on engagement to tackle competing predictions: although previous research has suggested that the size of a financial incentive is important for sustained engagement and thus behavior change [28,29], which is consistent with principles of economic rationality, other research [30] and theory in consumer psychology suggest that the magnitude of the incentive may be somewhat inelastic to size [30-32]. The secondary objectives

were to examine whether reward timing, type of reward, and health intervention content influence engagement.

Methods

Background

Carrot Insights Inc was a private company that developed the Carrot Rewards app, in conjunction with a number of federal and provincial government partners (the federal-provincial funding arrangement is described elsewhere) [16,31]. British Columbia (BC) was the company's founding provincial partner, and Carrot Insights Inc also partnered with 4 Canadian health charities (ie, Heart & Stroke Foundation of Canada, Diabetes Canada, YMCA Canada, BC Alliance for Healthy Living), primarily for the purpose of reviewing/approving health content delivered by the app. The marketing assets of one charity and 4 loyalty partners were also leveraged in the initial weeks of the app launching in BC with, for example, 1.64 million emails sent to the members of 3 of the 5 partners [31].

App Registration

Carrot Rewards was made available on the Apple App Store and Google Play app stores on March 3, 2016, in both English and French (Canada's official languages). On downloading the app, users entered their age, gender, postal code, and loyalty program card number of 1 of 4 programs of their choice (ie, movie, gas, grocery, or airline). To successfully register, users had to have entered a valid BC postal code and be aged 13 years or older (the age cutoff of the participating loyalty programs). British Columbians could download the app in 1 of 3 ways: organically (ie, finding it in the app store on their own), via an email invitation from a partner, or by using the promotional code friend referral mechanism [31].

Intervention Overview

Once the app was downloaded and registration was completed, users were offered 1 to 2 educational health quizzes per week over the first 5 weeks after registration, each containing 5 to 7 questions related to public health priorities identified by the BC Ministry of Health—healthy eating and physical activity/sedentary behavior (3 quizzes each)—and 2 separate health risk assessments that included items from national health surveys (regarding physical activity, eating and smoking habits, alcohol consumption, mental health, and overall well-being as well as the frequency of influenza immunization). The timing, content, and order of quizzes were the same for all individuals, other than the initial quiz timing, which will be discussed in greater detail in the *StudyDesign* section. Quizzes were developed to inform and familiarize users about self-regulatory health skills [32] or stepping stone behaviors (ie, goal setting, tracking, action planning, and barrier identification), skills that have been demonstrated in the past to promote health behaviors [31]. After completing a health quiz or health risk assessment and immediately earning incentives (US $0.04 to US $1.48 depending on the length, timing, and date of completion of the quiz), users could view relevant health information on partner websites. Each health quiz or assessment was designed to take approximately 1 to 3 min to complete.

Study Design and Participants

During the roll-out of Carrot in BC, there were 3 notable changes in the program introduced by its administrators, which provided the basis for the program variance that we explored in this study. These changes were driven largely by economic necessity rather than by theory or hypothesis testing but also presented the opportunity for a longitudinal nonexperimental observational study [27].

The first 2 changes were related to the number of points that participants could earn for completing quizzes. Specifically, during the study period following the launch of the app, there were (1) differences in the number of points offered across quizzes to compensate for differing quiz duration and timing (as demonstrated in the columns for each participant in Table 1) and (2) reductions in reward magnitude offered for the same quizzes over time (as demonstrated across in the rows for each quiz in Table 1). Owing to the unforeseen popularity of the platform and the need to manage costs within a finite budget financed by Carrot's public sector partners, the number of reward points awarded for the completion of each quiz was reduced over time. This meant that early subscribers received more points for the initial quizzes than those who were enrolled later in the study window we examined, as demonstrated by comparing the point profiles for each sample participant in Table 1. Of note, the content of the quizzes remained invariant across time and, as mentioned previously, was informed by behavioral theories in self-regulation and habit formation [33].

The third change introduced during the evaluation period was in the number of quizzes that a participant received on the day that they registered for the app: participants who self-registered between March 3 and March 17 were awarded points for registering and were immediately offered 2 quizzes (and thus opportunities to earn reward points) on the day of registration, whereas participants who registered after March 17 were also awarded points for registering but offered only one quiz on the day of registration (as demonstrated in Table 1 by comparing the offer day for participants A and B with those noted for participants C, D, and E). This created variance in terms of both (1) the number of quizzes offered at the time of enrollment and (2) the number of opportunities to earn reward points before participants faced their first 5-day waiting period between quizzes.

Taking advantage of these program-level changes to examine their impact on app engagement and attrition, we examined quiz acceptance rates over the first 5 weeks after registration for BC residents who received the first 9 offers (initial registration plus 8 quizzes) under the launch campaign and who registered for the Carrot app between the launch (March 3, 2016) and July 21, 2016 (n=54,817), with within-subjects repeated measures for each quiz, yielding 383,719 participant-level observations.

Table 1. Sample profiles of points awarded for quiz completion by initial registration date (Movie Reward Program).

Quiz number and names	Participant A (joined March 9)	Participant B (joined March 16)	Participant C (joined March 23)	Participant D (joined March 30)	Participant E (joined April 6)
1. Welcome to Carrot					
Day	Day 1	Day 1	Day 1	Day 1	Day 1
Points	100	38	25	25	17
2. What Does Eating a Rainbow Taste Like?					
Day	Day 1	Day 1	Day 1	Day 1	Day 1
Points	98	33	33	33	17
3. No Gym or Equipment Needed					
Day	Day 1	Day 1	Day 5	Day 5	Day 5
Points	165	101	58	58	33
4. Stand Up for Your Health					
Day	Day 5	Day 5	Day 10	Day 10	Day 10
Points	53	40	40	16	16
5. Carrot Health Survey 1					
Day	Day 10	Day 10	Day 15	Day 15	Day 15
Points	23	23	16	16	16
6. Rethink Sugary Drinks					
Day	Day 15	Day 15	Day 20	Day 20	Day 20
Points	18	18	16	16	16
7. The 2 Colours You Shouldn't Eat Without					
Day	Day 20	Day 20	Day 25	Day 25	Day 25
Points	17	17	17	17	17
8. Is Exercise Really Like Medicine?					
Day	Day 25	Day 25	Day 30	Day 30	Day 30
Points	17	15	15	15	15
9. Carrot Health Survey 2					
Day	Day 30	Day 30	Day 35	Day 35	Day 35
Points	14	14	14	14	14

Outcome Measures

For the purpose of this research, we explored the extent to which the 2 sources of program-level variance influenced the likelihood that a participant chose to engage with a given quiz. Thus, our outcome measure was a binary measure of whether a participant chose to complete each of the 8 quizzes during the initial 5 weeks postregistration.

Data Analyses

Independent Variables: Natural Experiment Factors

Although we observed the number of points earned for those who completed each quiz offer, including the change in the level of reward points offered (see Reward Points Schedule (first table) in Multimedia Appendix 1 for averages across waves), one limitation of our data is that they do not contain the number of points that were offered to participants who did not choose to complete a particular quiz offer; that is, because points awarded were based on the date of completion of a particular quiz, we only observed how many points a participant earned if they completed a particular quiz. Therefore, to explore the impact of these point changes on participants' probability of quiz acceptance, it was necessary to impute the number of points that participants who did not complete a quiz would have been offered for the completion of a particular quiz. In this case, there were 4 important pieces of observable information that inform this data imputation: (1) the app was designed such that when it was opened, participants were shown the number of points they could earn by completing each quiz on a given date; (2) the date when a particular quiz was made available to a participant; (3) the schedule of how many points an individual *could* have earned to complete a quiz on a given date; and (4) the date of quiz completion for all participants who *completed* a particular quiz. On the basis of the assumption that participants were choosing to either complete or not complete a particular quiz and to ensure the robustness of our results, we imputed the

missing observations for reward points that would have been offered to noncompleting participants in 2 ways, which are detailed in Multimedia Appendix 1. As the regression results for the 2 imputation approaches were consistent (as demonstrated in the third table in Multimedia Appendix 1–Regression Results by Imputation Methods), we reported only the results for the average days to completion imputation here.

Finally, as discussed previously, another important independent variable of interest was the combination of the differences in the timing of the quizzes and the number of reward points offered. To examine the impact of the timing structure on participants' probability of quiz acceptance, we created a dummy variable to act as the independent variable that indicates whether or not a participant was facing their first delay between quizzes.

Observed Variables: Individual, Intervention, and Reward Program Factors

We also examined the influence of 4 observed variables that differed across participants or quizzes but were independent of the program-level changes described previously, which are the primary focus of this study. These were (1) self-reported demographic variables of age and gender that participants logged at the time of registration, (2) dummy variables indicating the content of each quiz (healthy eating, exercise/physical activity, and health risk assessment), and (3) a set of indicator variables for each reward program under which participants could earn points via the Carrot app. Additionally, we controlled for whether a participant completed the quiz preceding the focal quiz. This accounts for the nature of the quizzes and the likely path dependence present in intervention programs of this type (ie, the propensity to respond to a quiz is likely correlated with the decision to respond to the previous quiz).

The reward program indicator variables were potentially important observed variables for 3 reasons, and all speak to our attempts to address alternative possible explanations for observed variations in quiz participation. First, each reward program has a different level of engagement with its participants and engaged in different levels of marketing efforts promoting their partnership with Carrot. As such, the inclusion of this variable may theoretically capture the potentially different levels of promotion of the Carrot app undertaken by each reward program. Second, the enrolled participant base of each reward program has varying demographic characteristics that may not be captured in the self-reported age and gender variables described earlier. Table 2 summarizes the census-level demographics (by forward sorting area—the first 3 characters of a Canadian postal code) for each program, and although generally the bases of the programs are similar, we observed some important differences between the gas program and others in terms of socioeconomic status and lifestyle characteristics (eg, commute method, urban vs extra urban). As such, we believe that this observed variable may also capture unobserved demographic differences at the individual level. Finally, each reward program varies with respect to the earnings and redemption mechanisms of the programs. One important difference between programs is that although the actual cash value of the points awarded for each completion of each quiz remained consistent across programs, the discrete number of points varied because the dollar-to-point conversion bases of the programs were not the same. For example, for the same quiz, participants would earn the same real monetary value in points for completion, the airline and grocery programs award might be 5 points, the movie reward program might offer 10 points, and the gas program might award 100 points. As such, there was a possible numeracy effect that was captured by this observed variable. Finally, some programs offer more utilitarian rewards on redemption (eg, gas, groceries), whereas others offer more experiential redeemed rewards (eg, movies, travel), which have been demonstrated to impact individual responses to reward programs [23]. Theoretically, the inclusion of this observed variable should capture the variance associated with psychological and behavioral factors. Descriptive statistics and a correlation matrix for all measures are summarized in Table 3.

Table 2. Demographic characteristics of the rewards program (census-level data at the forward sorting area level).

Demographic variables	Movies	Airline	Grocery	Gas
Age (years), mean (SD)	33 (0.06)	40 (0.21)	41 (0.16)	42 (0.30)
Male, n (%)	14503 (35.40)	1797 (39.98)	1742 (21.66)	1130 (44.68)
Household income[a] (US $), mean (SD)	52,975 (76.36)	53,600 (262.72)	52,822 (161.36)	50,616 (250.36)
Bachelor's degree or higher, mean % (SD)	21 (0.00)	25 (0.00)	18 (0.00)	17 (0.00)
Active transport for commute, mean % (SD)	9 (0.00)	11 (0.00)	8 (0.00)	8 (0.00)
Public transit for commute, mean % (SD)	13 (0.00)	14 (0.00)	9 (0.00)	9 (0.00)
Motor vehicle for commute, mean % (SD)	77 (0.00)	73 (0.00)	82 (0.00)	82 (0.00)
Immigrant population, mean % (SD)	33 (0.00)	35 (0.00)	26 (0.00)	27 (0.00)
Aboriginal population, mean % (SD)	3 (0.00)	3 (0.00)	5 (0.00)	5 (0.00)

[a]Can $ currency converted at US $1.3.

Table 3. Descriptive statistics and correlation matrix.

Variables	Mean (SD)	1	2	3	4	5	6	7	8	9
1. Point change since previous quiz	0.003 (0.503)	1	N/A[a]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2. First delay in quizzes	0.143 (0.350)	0.27	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
3. Movie rewards program	0.731 (0.443)	−0.02	0.00	1	N/A	N/A	N/A	N/A	N/A	N/A
4. Airline rewards program	0.080 (0.271)	0.04	0.00	−0.49	1	N/A	N/A	N/A	N/A	N/A
5. Grocery rewards program	0.143 (0.350)	−0.01	0.00	−0.68	−0.12	1	N/A	N/A	N/A	N/A
6. Eating knowledge quiz	0.286 (0.452)	−0.23	−0.26	0.00	0.00	0.00	1	N/A	N/A	N/A
7. Exercise knowledge quiz	0.429 (0.495)	0.41	0.47	0.00	0.00	0.00	−0.55	1	N/A	N/A
8. Completed previous study	0.806 (0.395)	0.09	0.16	0.12	−0.03	−0.09	−0.04	0.11	1	N/A
9. Gender (1=male)	0.350 (0.477)	0.00	0.00	0.04	0.04	−0.11	0.00	0.00	−0.02	1
10. Age	35.116 (13.195)	0.00	0.00	−0.29	0.11	0.21	0.00	0.00	−0.05	−0.04

[a]N/A: not applicable.

Statistical Methods

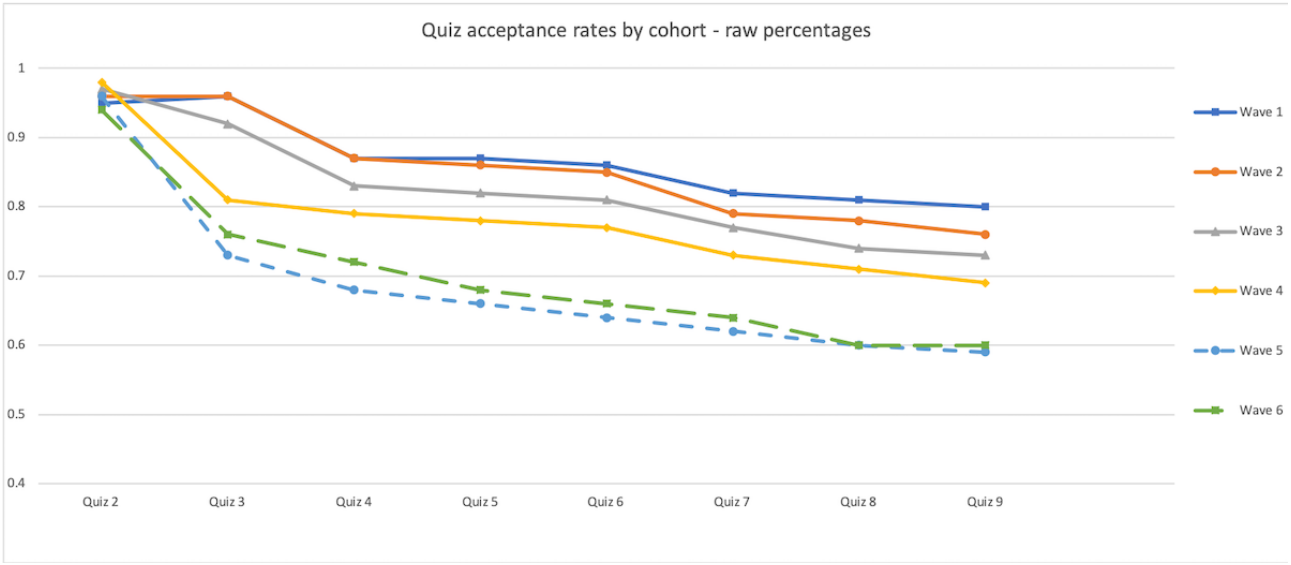
Data manipulation was conducted using SPSS version 24 (IBM Corporation), and statistical analysis was performed using the xtlogit procedure in STATA version 12.1 (Stata Corp). These random effects panel logit regression method was used to explore the impact of each of our program change variables, observed variables, and control variables on participants' probability of quiz acceptance across the 8 quiz offers received in the 5 weeks postregistration (the outcome measure). We ran our model using the average days to quiz completion imputation procedures for the data missing from individuals who did not complete a given quiz, as described in Multimedia Appendix 1. The results of this analysis are presented in the following section.

Results

Engagement

Although participants self-registered continuously throughout the study period, the date on which they registered placed them in different point schedules (see Table 1 or Reward Points Schedule (first table) in Multimedia Appendix 1 for more details). For the purpose of this study, we created analytical groups that we refer to as *waves*, which represent the point schedule in place at the time of their registration with the app. These groupings allow us to explore the effects of exogenous point changes on participant program engagement. Figure 1 demonstrates participant response rates for each of the first 8 quizzes across waves. Although not a conclusive analysis, we observed 4 key characteristics of these curves. First, the response rates to quiz 2 across the 6 waves were consistently high across all waves, ranging from 95% to 97%. This suggests that despite the significant difference in reward points offered for the completion of this quiz across waves (as demonstrated in Table 1), participants across waves are approximately equally likely to respond to this quiz. Second, response rates to the final quiz examined (quiz 9) varied, ranging from 80.5% (19,525/24,249) for wave 1 registrants to 60% (2930/4893) for wave 6 registrants, which suggests the possibility of intervening factors that predicted participant engagement between the time of initial registration and quiz 9. Third, a statistical test of the equality of the slopes from quiz 4 onward indicates that response rates across registrants in each of the waves did not differ significantly from one another after quiz 4. This suggests that the factors driving the differences in response rates appear to be independent of the underlying differences in responsiveness driven by the time of registration in the program. These results also provide support for the belief that participants were not fundamentally different from one another across the study period and indicate that it is unlikely that there were significant differences between registrants over the study period with respect to their attitudes to the underlying behavior of interest (ie, as in diffusion theories [34] where innovators, early adopters, and so on only emerge over an extended life cycle of a product). Finally, we also note that much of the decline in responsiveness across waves occurs around the time of quizzes 3 and 4, which coincides with the timing delay noted previously and which we explored in greater detail in our regression analysis.

Figure 1. Quiz acceptance rates across quizzes by wave.



Regression Results

Overall, from the regression results summarized in Table 4, the regression results with respect to the features of this observational study suggest that the impact of a change in points on the likelihood of response to a quiz is positive and statistically significant (P<.001); however, the magnitude of the coefficient suggests that the elasticity of this response is limited. Specifically, in our data, a 10% decrease in points offered from the previous quiz resulted in a 1% decrease in the likelihood of response to a given quiz. Second, we found that the impact of the first delay that participants face was negative and statistically significant (P<.001). This result demonstrates that the longer delay between quizzes 2 and 3 faced by those who registered after March 17 decreased their likelihood of response by 64%.

With respect to the observed variables, we first examined the impact of demographic characteristics on the likelihood of quiz acceptance. We observed no significant effect of age (P=.52) on the likelihood of response; however, gender did have a significant impact (P<.001), whereby male participants were 15% to 16% less likely to respond to a given quiz than female participants. The next feature we explored in the model was the impact of quiz content on response rates. We observed that relative to the health risk assessments, participants were 21% more likely to respond to physical activity–related quiz offers (P<.001) and 10% less likely to respond to healthy eating–related quiz offers (P<.001). Finally, the results indicate that the reward program under which a participant registers had a significant impact on the likelihood of a quiz response. Specifically, relative to the gas rewards program, we observed that the movie rewards program participants were 355% more likely to respond to a given quiz (P<.001), the airline rewards program participants were 210% more likely to respond to a given quiz (P<.001), and grocery rewards program participants were 140% more likely to respond to a given quiz (P<.001).

Table 4. Regression results (N=383,719 observations of 54,917 individuals, with 95% CIs in parentheses).

Dependent variable=probability of quiz acceptance	Estimated coefficients (95% CI)
Point change since previous quiz	1.100 (1.065-1.137)
First delay in quizzes	0.361 (0.345-0.378)
Gender (1=male)	0.845 (0.804-0.887)
Age	0.999 (0.998-1.001)
Eating knowledge quiz[a]	0.896 (0.861-0.932)
Exercise knowledge quiz[a]	1.206 (1.156-1.257)
Movie rewards program[b]	3.552 (3.176-3.971)
Airline rewards program[b]	2.096 (1.840-2.386)
Grocery rewards program[b]	1.402 (1.247-1.575)
Completed previous study	71.891 (67.261-76.839)
Model fit—McFadden pseudo-R^2	0.172

[a]Knowledge quiz estimates relative to health risk assessments.

[b]Rewards program estimates relative to the gasoline rewards program.

Discussion

This study responds to a call in the literature to create a more comprehensive understanding of how contextual and program factors relate to program effectiveness [35] and, specifically, to explore how the type, amount, and timing of incentives impact program outcomes [36]. It does so by exploring how the variance induced by changes in the features of health interventions under the Carrot Rewards program (eg, incentive size, variability, quiz timing) during its BC roll-out, as well as observed differences between intervention content, reward program, and characteristics of participants, all impact the likelihood of quiz response. Thus, it contributes to the broader literature on how financial incentives [35] (particularly loyalty points) [36] and/or mHealth apps [28,37,38] can be deployed to improve uptake and engagement with health education–based interventions [37,39,40] and encourage health-related behaviors such as physical activity [18,36], management of chronic health conditions [28,41,42], smoking cessation [43,44], weight loss [45], and medication adherence and clinical treatment plans [40].

Effectiveness

Our study demonstrates that the ongoing provision of a stream of small incentives (as low as US $0.05 per offer) can largely sustain the initial high levels of responsiveness generated by these programs. Interestingly, however, our results also demonstrate that the responsiveness to reward point decreases is relatively inelastic when compared with the other features of the interventions being offered via the Carrot app. Specifically, our results indicate that reducing the number of points associated with a particular quiz by 10% only leads to a 1% (P<.001) decrease in the likelihood of offer response. This relative inelasticity is consistent with the findings of Carrera et al [46], who found that US $60 for 9 gym visits in 6 weeks was no more effective than US $30 [46]. However, detailed subgroup meta-analyses by Mitchell et al [26] suggest that larger incentives can, in some cases, produce larger effect sizes in a physical activity context. As health-related interventions are often multifaceted and include incentives as just one of several components intended to produce positive health-related behavior changes, it is likely that other program features (eg, reward timing, salience of feedback, goal setting approach) can play equally important roles. Thus, when combined with previous findings that suggest that the *presence* of financial incentives matters for responsiveness [47], our results suggest that a relatively small number of points may be nearly as effective as larger numbers of points. This may be particularly true when the incentives are delivered immediately on completion of the focal behavior in question, as it leverages the formidable bias toward the present, which has been demonstrated in behavioral economics [48,49] as well as speaks to the power of technologies such as smartphones to leverage that bias for increased behavioral compliance. This nearly instantaneous connection between behavior and reward may support the potential for diminution of the size (and thus cost) of incentives necessary to produce the requisite level of behavioral change. Despite findings that suggest that there is a minimum threshold for the size of daily incentives to maintain engagement and produce the desired behavior change [25,29], recent health incentive studies appear to be using much smaller incentives in part because of the ability of mHealth apps to deliver daily rewards, consistent with behavioral theories [26].

Contribution to Theory

In the context of the consumer psychology literature on loyalty points discussed earlier, our results support this prior work, which has demonstrated that consumers do not behave rationally (in the behavioral economics sense of the term) in their response to the quizzes and the size of the financial incentives offered. On the basis of previous research in mHealth and on the findings presented here, quiz response and thus app engagement appear to rely more on the simple presence of points (as a form of financial incentive), as opposed to the absolute quantity, as well as on the continued opportunity to earn more points without a delay. This may be for a variety of reasons. First, loyalty points already require users to accumulate over time, so there may be a built-in progress element and a recognition that the ultimate reward will be sometime in the future [22,23]. This may insulate programs such as the one discussed here from the negative effects seen in other studies where incentives get smaller over time. In addition, point programs have been hypothesized to cue goals related to gamification, where it is the acquisition of points itself that generates the affective value, rather than the economic reward per se. Our results demonstrate that the continuous, timely provision of small numbers of reward points throughout the program significantly improved retention and engagement with the Carrot app relative to previous programs that used reward points only for recruitment purposes [37] and provides support for previous findings regarding the efficiency of small, variable rewards and a decreasing schedule in maintaining engagement with health-based interventions [39,50]. Our study also provides a novel context in which to examine the behavior of individuals in response to loyalty points, but in a health context rather than a commercial context, pointing to the opportunity to use an alternative currency from which individuals already derive personal, affective value [33] as a type of financial incentive for health-related behaviors.

Program Design Considerations

Perhaps most importantly, our results suggest that other elements of the design of these interventions need to be considered when attempting to increase response rates. Existing research has demonstrated the critical importance of including direct end user feedback in the development of such apps early in the design to ensure both short- and long-term engagement [8,40,51,52]. One important finding from this study is that consistent with research on habit formation [30], an increased focus is needed to retain individuals during the early engagement period. Indeed, this was the largest driver of participant attrition that we found in our study, with a differential delay between early offers being associated with a 64% decrease in response probability. This suggests that other interventions and communication initiatives should be considered during this waiting period to retain participants. The good news is that once this hurdle is surmounted (ie, once a participant re-engages after the first waiting period), retention rates between interventions are remarkably high and consistent across the registration period

studied, particularly considering the small magnitude of incentives that are being offered (Figure 1).

Additional findings of our study support the importance of nonincentive elements of this program, although by and large these results are much stronger in suggesting directions for further study than they are for creating concrete recommendations for program optimization. One important factor identified in our findings with respect to participant response rates is the content of the quiz being offered. Our results suggest that when the content of the quiz is focused on healthy eating, individuals are significantly less likely to respond (10% decrease in the likelihood of acceptance relative to health risk assessments), and when the content of the quiz focused on exercise or physical activity, individuals are more likely to respond (20% increase in the likelihood of acceptance relative to health risk assessments). These results suggest that more than just the simple time cost of each intervention should be considered when designing incentive programs, perhaps incorporating individual and/or psychological factors that may influence the attractiveness of a quiz, which may increase or decrease the perceived costs of quiz acceptance and ultimately affect the propensity for uptake of an intervention.

Similarly, among the observed variables that measured age and gender, quiz content, participation in previous quiz, and reward program, we find that the strongest predictor of the likelihood of quiz acceptance among all variables studied is the point program under which a participant is earning points for completing quiz offers via Carrot. Specifically, we find that individuals who choose to earn points in the movie (355%), airline (210%), and grocery (140%) programs all have a greater likelihood of response relative to the gasoline rewards program. We previously acknowledged that the differences between these reward programs may vary along multiple dimensions, including differential promotional efforts in support of Carrot, differences in the demographic composition of the participant base for each program, differences in individual responsiveness due to a numeracy effect (ie, holding actual economic value constant, the quantity of points earned per activity varies across programs, and people respond more to higher numbers of points earned), and differences in the earning and redemption mechanisms of each point program that may each influence the responsiveness of individuals. This facet of program design requires more exploration, but our results suggest that it is yet another factor that should be considered when designing optimal incentive-based intervention programs.

Limitations and Future Directions

One concern of the modeling approach used in this study is the possibility that the differential observed responsiveness to quizzes between what we have categorized as waves is due to underlying or inherent differences (whether demographic or psychographic) between participants who registered earlier versus later in the observed window of this study, rather than due to the previously described variations of the program. However, given (1) the very short time windows between the changes in the point scheme (as demonstrated in the first table in Multimedia Appendix 1–Reward Points Schedule), (2) the consistent slopes of the engagement curves described in the previous discussion of Figure 1, (3) the clustering approach at the participant level used in the estimation of the random effects regression model, and (4) the inclusion of a control for response to prior quizzes (which is likely to be autoregressive), we can be reasonably confident that program features drive the observed effects, rather than the unobservable characteristics of the participants. From observable data as well as theoretically, there is little to suggest that early registrants differ meaningfully from later registrants in terms of responsiveness; thus, the differences in ultimate outcomes are likely driven by features of the intervention rather than sampling considerations. Recall that participants were neither recruited nor assigned to waves; rather, those waves were our analytical construction to allow us to explore the exogenous changes to the program structure. Consequently, we would have no reason to believe that the presence of any particular individual in a given wave was anything other than random. Ideally, we would be able to more explicitly examine potential confounding factors such as individual-level differences at a psychological level (eg, promotion vs prevention orientation); regrettably, that information was not collected by the app as part of the registration process and therefore is not available.

Another limitation related to statistical analysis concerns the fact that we can only be certain of the number of points that participants saw for any given quiz for those individuals who completed the quiz in question. Although we have used conservative tests for imputation of the missing data (see Multimedia Appendix 1 for specific details), our ability to draw conclusions about individuals' engagement is limited to those who remained engaged with the app and completed the quizzes, not about individuals who skipped or ceased participating. It is also not possible, with this study and data, to explore the extent to which individuals were aware of or sensitive to the point changes that occurred over time (eg, for participant A in Table 1, a decrease from quiz 3 of 165 points to quiz 5, where only 23 points were offered). We do find that for the individuals we were able to observe who continued to complete quizzes, even if the declining number of points was salient, it did not appear to have a large effect on engagement with the app.

We proposed several theoretical mechanisms by which different reward programs may affect engagement with the app (eg, the hedonic vs utilitarian nature of the rewards, numeracy effects), but in this study, we do not have the data to be able to empirically substantiate those hypotheses. In addition, we are unable to retroactively document the promotion or advertising of Carrot by each of the 4 participating loyalty programs or to identify underlying psychographic differences between enrollees in one program versus another. The inclusion of these factors in the model would potentially allow us to further control the reward program–specific characteristics to be able to make more informed recommendations for practitioners, program administrators, or government stakeholders in terms of the need for or effectiveness of promotion efforts to drive enrollment in and engagement with the app. In general, there is a lack of data within the literature to date to support the economic basis for the use of mHealth behavioral interventions [40,53]. Although this study does not directly tackle that issue, it does suggest an alternative dimension of cost consideration that may not be

captured in existing frameworks such as the Consolidated Health Economic Evaluation Reporting Standards [53] and speaks to the need for ground effectiveness assessments in economic considerations as well as more traditional measurements such as reach. Finally, the realities of a program such as the one observed here in the variation of points offered confine us to a posthoc analysis of quizzes over a limited period. This constrains our ability to engage in repeated measures of a particular attitude or intention related to a specific context (eg, healthy eating, exercise) beyond the types of quizzes that happened to be administered within the period of observation.

Additional factors that could be integrated into the model include devising quasiexperimental studies to examine other contextual factors that may impact program engagement and effectiveness. These could include individual difference variables such as the impact of socioeconomic or health status on both app engagement and quiz effectiveness at driving attitudinal or behavioral change or program-centered factors, such as the comparative effectiveness of individual rewards versus team-based rewards, norm-based messaging, or differential offer timing. Future programming should use a modular approach such that content can be contained in a specific topic of interest to measure engagement in the quizzes in a more systematic manner and explore how the impacts of different program features may vary over time. Examining these modular programs using microrandomized controlled trial designs, similar to ongoing work by Kramer et al [54], would allow for a more targeted evaluation of how content, along with frequency, variations leads to greater and/or sustained engagement within the app.

Conclusions

The results of this study demonstrate that the Carrot program, built around an ongoing stream of a small number of reward points awarded instantly on completion of health interventions, was able to drive significantly higher engagement levels than demonstrated in previous literature exploring the intersection of mHealth apps and financial incentives. Furthermore, although previous studies have demonstrated that the presence of an incentive matters to user engagement, our study suggests that the number of points offered for these reward point–based health interventions is less important than other program design features such as the type of reward points being offered, the timing of intervention and reward offers, and the content of the health interventions for driving continued engagement by users.

Acknowledgments

The authors acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada for this project.

Conflicts of Interest

MM received consulting fees from Carrot Insights Inc. from 2015 to 2018 as well as travel reimbursement in January and March 2019. MM had stock options in the company as well but these are now void since Carrot Insights Inc. went out of business in June 2019. LW was employed by Carrot Insights Inc. from March 2016 to June 2019 and also had stock options which are now void. The other authors report no conflicts of interest.

Multimedia Appendix 1
Point schedule changes across study window and missing data imputation for points offered.
[DOCX File , 26 KB - jmir_v22i8e16797_app1.docx]

References

1. How Healthy Are Canadians? A Trend Analysis of the Health Of Canadians From a Healthy Living And Chronic Disease Perspective. Government of Canada. 2016. URL: https://www.canada.ca/content/dam/phac-aspc/documents/services/publications/healthy-living/how-healthy-canadians/pub1-eng.pdf [accessed 2019-04-16]
2. Krueger H, Turner D, Krueger J, Ready AE. The economic benefits of risk factor reduction in Canada: tobacco smoking, excess weight and physical inactivity. Can J Public Health 2014 Mar 18;105(1):e69-e78 [FREE Full text] [doi: 10.17269/cjph.105.4084] [Medline: 24735700]
3. Global Action Plan on Physical Activity 2018–2030: More Active People for a Healthier World. World Health Organization. 2018. URL: https://www.who.int/ncds/prevention/physical-activity/global-action-plan-2018-2030/en/ [accessed 2019-08-12]
4. Sallis JF, Bull F, Guthold R, Heath GW, Inoue S, Kelly P, Lancet Physical Activity Series 2 Executive Committee. Progress in physical activity over the olympic quadrennium. Lancet 2016 Sep 24;388(10051):1325-1336. [doi: 10.1016/S0140-6736(16)30581-5] [Medline: 27475270]
5. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. J Behav Med 2019 Feb;42(1):67-83 [FREE Full text] [doi: 10.1007/s10865-018-9966-z] [Medline: 30825090]
6. Heuveldop N. Ericsson Mobility Report. Ericsson Careers: Join our Global Team. 2017 Jun. URL: http://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-june-2017-raso.pdf [accessed 2019-05-03]
7. mHealth Economics 2017 – Current Status and Future Trends in Mobile Health. Research2Guidance. 2017. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2019-05-03]

8. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
9. McCallum C, Rooksby J, Gray CM. Evaluating the impact of physical activity apps and wearables: interdisciplinary review. JMIR Mhealth Uhealth 2018 Mar 23;6(3):e58 [FREE Full text] [doi: 10.2196/mhealth.9054] [Medline: 29572200]
10. Schoeppe S, Alley S, van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 7;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
11. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of ehealth and mhealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-28.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
12. Dounavi K, Tsoumani O. Mobile health applications in weight management: a systematic literature review. Am J Prev Med 2019 Jun;56(6):894-903 [FREE Full text] [doi: 10.1016/j.amepre.2018.12.005] [Medline: 31003801]
13. Feter N, dos Santos TS, Caputo EL, da Silva MC. What is the role of smartphones on physical activity promotion? A systematic review and meta-analysis. Int J Public Health 2019 Jun;64(5):679-690. [doi: 10.1007/s00038-019-01210-7] [Medline: 30758514]
14. How Uninstall Tracking Helps Improve App Store Ranking. Singular. 2016. URL: https://www.singular.net/ [accessed 2019-06-19]
15. Handley MA, Lyles CR, McCulloch C, Cattamanchi A. Selecting and improving quasi-experimental designs in effectiveness and implementation research. Annu Rev Public Health 2018 Apr 1;39:5-25. [doi: 10.1146/annurev-publhealth-040617-014128] [Medline: 29328873]
16. National Healthy Living Platform: 'Carrot Rewards' Targets Lifestyle Improvements. Government of Canada. URL: https://www.canada.ca/en/news/archive/2015/07/national-healthy-living-platform-carrot-rewards-targets-lifestyle-improvements.html [accessed 2019-05-30]
17. Scott RH. Nudge: improving decisions about health, wealth, and happiness. Choice Rev Online 2008 Oct 1;46(2):46-0977. [doi: 10.5860/choice.46-0977]
18. Mitchell MS, Orstad SL, Biswas A, Oh PI, Jay M, Pakosh MT, et al. Financial incentives for physical activity in adults: systematic review and meta-analysis. Br J Sports Med 2019 May 15:- epub ahead of print. [doi: 10.1136/bjsports-2019-100633] [Medline: 31092399]
19. Mitchell MS, Orstad SL, Biswas A, Oh PI, Jay M, Pakosh MT, et al. Financial incentives for physical activity in adults: systematic review and meta-analysis. Br J Sports Med 2019 May 15:-. [doi: 10.1136/bjsports-2019-100633] [Medline: 31092399]
20. van Osselaer SM, Alba JW, Manchanda P. Irrelevant information and mediated intertemporal choice. J Consum Psychol 2004 Jan;14(3):257-270. [doi: 10.1207/s15327663jcp1403_7]
21. Kivetz R, Urminsky O, Zheng Y. The goal-gradient hypothesis resurrected: purchase acceleration, illusionary goal progress, and customer retention. J Mark Res 2018 Oct 10;43(1):39-58. [doi: 10.1509/jmkr.43.1.39]
22. Bagchi R, Li X. Illusionary progress in loyalty programs: magnitudes, reward distances, and step-size ambiguity. J Consum Res 2011 Feb 1;37(5):888-901. [doi: 10.1086/656392]
23. Kivetz R, Simonson I. Earning the right to indulge: effort as a determinant of customer preferences toward frequency program rewards. J Mark Res 2018 Oct 10;39(2):155-170. [doi: 10.1509/jmkr.39.2.155.19084]
24. Kwong JY, Soman D, Ho CK. The role of computational ease on the decision to spend loyalty program points. J Consum Psychol 2011 Apr;21(2):146-156. [doi: 10.1016/j.jcps.2010.08.005]
25. Kramer J, Tinschert P, Scholz U, Fleisch E, Kowatsch T. A cluster-randomized trial on small incentives to promote physical activity. Am J Prev Med 2019 Feb;56(2):e45-e54 [FREE Full text] [doi: 10.1016/j.amepre.2018.09.018] [Medline: 30553693]
26. Mitchell M, White L, Lau E, Leahey T, Adams MA, Faulkner G. Evaluating the carrot rewards app, a population-level incentive-based intervention promoting step counts across two Canadian provinces: quasi-experimental study. JMIR Mhealth Uhealth 2018 Sep 20;6(9):e178 [FREE Full text] [doi: 10.2196/mhealth.9912] [Medline: 30148712]
27. Craig P, Katikireddi SV, Leyland A, Popham F. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. Annu Rev Public Health 2017 Mar 20;38:39-56 [FREE Full text] [doi: 10.1146/annurev-publhealth-031816-044327] [Medline: 28125392]
28. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered ehealth interventions for children, adolescents, and adults with sickle cell disease: systematic review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
29. Strohacker K, Galarraga O, Williams DM. The impact of incentives on exercise behavior: a systematic review of randomized controlled trials. Ann Behav Med 2014 Aug;48(1):92-99 [FREE Full text] [doi: 10.1007/s12160-013-9577-4] [Medline: 24307474]
30. Carrera M, Royer H, Stehr M, Sydnor J. Can financial incentives help people trying to establish new habits? Experimental evidence with new gym members. J Health Econ 2018 Mar;58:202-214 [FREE Full text] [doi: 10.1016/j.jhealeco.2018.02.010] [Medline: 29550665]

31. Mitchell M, White L, Oh P, Alter D, Leahey T, Kwan M, et al. Uptake of an incentive-based mhealth app: process evaluation of the carrot rewards app. JMIR Mhealth Uhealth 2017 May 30;5(5):e70 [FREE Full text] [doi: 10.2196/mhealth.7323] [Medline: 28559224]
32. Kowatsch T, Kramer J, Kehr F, Wahle F, Elser N, Fleisch E. Effects of charitable versus monetary incentives on the acceptance of and adherence to a pedometer-based health intervention: study protocol and baseline characteristics of a cluster-randomized controlled trial. JMIR Res Protoc 2016 Sep 13;5(3):e181 [FREE Full text] [doi: 10.2196/resprot.6089] [Medline: 27624645]
33. Mitchell M, White L, Oh P, Alter D, Leahey T, Kwan M, et al. Uptake of an incentive-based mhealth app: process evaluation of the carrot rewards app. JMIR Mhealth Uhealth 2017 May 30;5(5):e70 [FREE Full text] [doi: 10.2196/mhealth.7323] [Medline: 28559224]
34. Beal GM, Rogers EM, Bohlen JM. Validity of the concept of stages in the adoption process. Rural Sociol 1957;22(2):166-168 [FREE Full text]
35. Adams J, Giles EL, McColl E, Sniehotta FF. Carrots, sticks and health behaviours: a framework for documenting the complexity of financial incentive interventions to change health behaviours. Health Psychol Rev 2014;8(3):286-295. [doi: 10.1080/17437199.2013.848410] [Medline: 25053215]
36. Faulkner G, Dale LP, Lau E. Examining the use of loyalty point incentives to encourage health and fitness centre participation. Prev Med Rep 2019 Jun;14:100831 [FREE Full text] [doi: 10.1016/j.pmedr.2019.100831] [Medline: 30886815]
37. Goyal S, Morita PP, Picton P, Seto E, Zbib A, Cafazzo JA. Uptake of a consumer-focused mhealth application for the assessment and prevention of heart disease: the <30 days study. JMIR Mhealth Uhealth 2016 Mar 24;4(1):e32 [FREE Full text] [doi: 10.2196/mhealth.4730] [Medline: 27012937]
38. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
39. Leahey TM, Subak LL, Fava J, Schembri M, Thomas G, Xu X, et al. Benefits of adding small financial incentives or optional group meetings to a web-based statewide obesity initiative. Obesity (Silver Spring) 2015 Jan;23(1):70-76 [FREE Full text] [doi: 10.1002/oby.20937] [Medline: 25384463]
40. Badawy SM, Kuhns LM. Texting and mobile phone app interventions for improving adherence to preventive behavior in adolescents: a systematic review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
41. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
42. Barte JC, Wendel-Vos GC. A systematic review of financial incentives for physical activity: the effects on physical activity and related outcomes. Behav Med 2017;43(2):79-90. [doi: 10.1080/08964289.2015.1074880] [Medline: 26431076]
43. Radley A, Ballard P, Eadie D, MacAskill S, Donnelly L, Tappin D. Give It Up For Baby: outcomes and factors influencing uptake of a pilot smoking cessation incentive scheme for pregnant women. BMC Public Health 2013 Apr 15;13:343 [FREE Full text] [doi: 10.1186/1471-2458-13-343] [Medline: 23587161]
44. Giles EL, Robalino S, McColl E, Sniehotta FF, Adams J. The effectiveness of financial incentives for health behaviour change: systematic review and meta-analysis. PLoS One 2014;9(3):e90347 [FREE Full text] [doi: 10.1371/journal.pone.0090347] [Medline: 24618584]
45. Jeffery RW. Financial incentives and weight control. Prev Med 2012 Nov;55 Suppl:S61-S67 [FREE Full text] [doi: 10.1016/j.ypmed.2011.12.024] [Medline: 22244800]
46. Carrera M, Royer H, Stehr M, Sydnor J. Can financial incentives help people trying to establish new habits? Experimental evidence with new gym members. J Health Econ 2018 Mar;58:202-214 [FREE Full text] [doi: 10.1016/j.jhealeco.2018.02.010] [Medline: 29550665]
47. Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am J Prev Med 2013 Nov;45(5):658-667. [doi: 10.1016/j.amepre.2013.06.017] [Medline: 24139781]
48. Camerer CF, Loewenstein G. Behavioral economics: past, present, future. In: Advances in Behavioral Economics. Princeton, UK: Princeton University Press; 2003.
49. Volpp KG, Asch DA, Galvin R, Loewenstein G. Redesigning employee health incentives--lessons from behavioral economics. N Engl J Med 2011 Aug 4;365(5):388-390 [FREE Full text] [doi: 10.1056/NEJMp1105966] [Medline: 21812669]
50. Andrade LF, Barry D, Litt MD, Petry NM. Maintaining high activity levels in sedentary adults with a reinforcement-thinning schedule. J Appl Behav Anal 2014;47(3):523-536 [FREE Full text] [doi: 10.1002/jaba.147] [Medline: 25041789]
51. Badawy SM, Thompson AA, Liem RI. Technology access and smartphone app preferences for medication adherence in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2016 May;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]

52. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Feb 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
53. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
54. Kramer J, Künzler F, Mishra V, Presset B, Kotz D, Smith S, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540 [FREE Full text] [doi: 10.2196/11540] [Medline: 30702430]

Abbreviations

BC: British Columbia
mHealth: mobile health

Edited by G Eysenbach; submitted 25.10.19; peer-reviewed by S Smith , S Badawy; comments to author 23.12.19; revised version received 05.05.20; accepted 03.06.20; published 10.08.20.

Please cite as:
Brower J, LaBarge MC, White L, Mitchell MS
Examining Responsiveness to an Incentive-Based Mobile Health App: Longitudinal Observational Study
J Med Internet Res 2020;22(8):e16797
URL: https://www.jmir.org/2020/8/e16797
doi:10.2196/16797
PMID:32773371

Original Paper

Engagement in an Interactive App for Symptom Self-Management during Treatment in Patients With Breast or Prostate Cancer: Mixed Methods Study

Marie-Therése Crafoord[1*], RN, BSc, MSc; Maria Fjell[1*], RN, BSc, OCN; Kay Sundberg[1], RN, PhD; Marie Nilsson[1], MSW, PhD; Ann Langius-Eklöf[1], RN, PhD

Division of Nursing, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

[*]these authors contributed equally

Corresponding Author:
Marie-Therése Crafoord, RN, BSc, MSc
Division of Nursing
Department of Neurobiology, Care Sciences and Society
Karolinska Institutet
Alfred Nobels Allé 23
Stockholm, 14183
Sweden
Phone: 46 8524 837 49
Email: marie-therese.crafoord@ki.se

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/10/e33140

Abstract

Background: Using mobile technology for symptom management and self-care can improve patient-clinician communication and clinical outcomes in patients with cancer. The interactive app Interaktor has been shown to reduce symptom burden during cancer treatment. It includes symptom assessment, an alert system for contact with health care professionals, access to self-care advice, and visualization of symptom history. It is essential to understand how digital interventions operate; one approach is to examine engagement by assessing usage and exploring user experiences. Actual usage in relation to the intended use—adherence—is an essential factor of engagement.

Objective: This study aimed to describe engagement with the Interaktor app among patients with breast or prostate cancer during treatment.

Methods: Patients from the intervention groups of two separate randomized controlled trials were included: patients with breast cancer receiving neoadjuvant chemotherapy (n=74) and patients with locally advanced prostate cancer receiving treatment with radiotherapy (n=75). The patients reported their symptoms daily. Sociodemographic and clinical data were obtained from baseline questionnaires and medical records. Logged data usage was retrieved from the server and analyzed descriptively and with multiple regression analysis. Telephone interviews were conducted with patients about their perceptions of using the app and analyzed using content analysis.

Results: The median adherence percentage to daily symptom reporting was 83%. Most patients used the self-care advice and free text message component. Among the patients treated for breast cancer, higher age predicted a lower total number of free text messages sent (P=.04). Among the patients treated for prostate cancer, higher age (P=.01) and higher education level (P=.04), predicted an increase in total views on self-care advice, while higher comorbidity (P=.004) predicted a decrease in total views on self-care advice. Being married or living with a partner predicted a higher adherence to daily symptom reporting (P=.02). Daily symptom reporting created feelings of having continuous contact with health care professionals, being acknowledged, and safe. Being contacted by a nurse after a symptom alert was considered convenient and highly valued. Treatment and time-related aspects influenced engagement. Daily symptom reporting was perceived as particularly meaningful at the beginning of treatment. Requests were made for advice on diet and psychological symptoms, as well as for more comprehensive and detailed information as the patient progressed through treatment.

Conclusions: Patient engagement in the interactive app Interaktor was high. The app promoted patient participation in their care through continuous and convenient contact with health care professionals. The predictive ability of demographic variables differed between patient groups, but higher age and a higher educational level predicted usage of specific app functions for both patient groups. Patients' experience of relevance and interactivity influenced their engagement positively.

(J Med Internet Res 2020;22(8):e17058) doi:10.2196/17058

KEYWORDS

engagement; adherence; mHealth; mobile app; cancer supportive care; symptom management; usage metrics; breast cancer; prostate cancer

Introduction

Treatments for cancer can lead to challenging symptoms, but most patients are managed as outpatients [1,2]. Patients' own assessments of the occurrence and severity of symptoms and their concerns can inform and support health care professionals in identifying and assessing the potential risks associated with cancer treatment, leading to improved patient outcomes [3,4]. Interventions using mobile technology to support symptom monitoring and self-care among patients being treated for cancer have been shown to improve patient-clinician communication, improve symptom management and self-care ability, reduce symptom burden, and increase survival [4-6].

Even though apps to support symptom management for patients with cancer have increased, few feature evidence-based content or have been tested in rigorous trials [7,8]. Moreover, only some include interactive components, such as support for self-care and communication with health care professionals and peers for immediate clinical management [7,9-12]. A key aspect of digital interventions is to understand how they operate and how they can be enhanced by assessment of usage and user experiences [13-17].

We designed Interaktor, an interactive smartphone and tablet app, to support patient symptom management. The concepts of person-centered and participatory care inspired the development of the intervention [18,19]. The app is available in different versions tailored for patients during treatment for breast cancer, prostate cancer [20], and pancreatic cancer [21] and for older persons receiving home care [22]. The content of the different versions of the app was developed in an iterative process, which included reviews of literature focusing on symptoms and their management, clinical guidelines, interviews with patients, and interviews with health care professionals [21,23-25]. Detailed descriptions about the intervention outline and screenshots have been previously presented [20,26]. Interaktor includes four generic components: (1) Self-assessment through questions about symptoms and concerns, where patients report symptom occurrence, symptom frequency, and their distress level inspired by the Memorial Symptom Assessment Scale is a main component [27]. Patients also have the opportunity to report other symptoms with free text messages. The reports are immediately transferred via a secure server to health care professionals who can monitor patient reports in real time via a web interface. (2) A risk assessment model is included for symptoms that notifies nurses at the clinic by SMS text message when a high level of frequency or distress is reported for a symptom. There are two kinds of alerts: yellow and red. Yellow alerts require that a nurse contact the patient during the daytime, and red alerts require that a nurse contact the patient within 1 hour. (3) Evidence-based self-care advice and links to relevant webpages related to the assessed symptoms and other areas of concern are included. (4) Graphs showing reported symptom history for patients and health care professionals are also included. Interaktor can be used with Android and iOS, but the app is only available for research purposes. Logged data were stored on a separate secure server hosted by the health care company that developed the app.

Studies suggest that patients undergoing treatment for cancer make use of and appreciate opportunities to report symptoms to health care professionals when they are at home [26,28,29]. Mobile technology for health (mHealth) to support self-management has been linked to positive outcomes regarding physical as well as psychological symptoms in the context of cancer care [30,31]. An early system for remote symptom management during cancer treatment included symptom assessments twice a day, tailored advice, and access to informational webpages [5]; the study [5] found that patients differed markedly in the number of reports made and all patients viewed the webpages. Furthermore, modest problems with the technology were described, and patients rated improvements in the communication process with hospital staff and their satisfaction with care [32]. Systems for remote symptom management during cancer treatment have since demonstrated high acceptance [33] as well as long-term feasibility [34]. Most have been web-based and have involved symptom assessments from a home computer or clinic tablet [4,6,33,35].

Usage and user experiences of a web- or mobile-based intervention can be described by the concept of engagement [15]. Engagement is influenced by interconnected factors—some individual, such as demographics, skills, and understanding; some contextual, which include internet access and online environment; and some interventional, such as technical and design features [36]. Patients' prior health behaviors and smartphone experience will affect how relevant and usable an intervention is perceived to be, which affects engagement, and persistent patient engagement is achieved if the intervention is perceived to be usable, relevant, helpful, and interactive [36]. Usage level compared to the intended usage is referred to as *adherence* [37]. The significance of adherence has gained increased recognition, since evidence has emerged that high levels of adherence positively correlate with improved outcomes [38]. However, within eHealth in general, levels of nonusage and dropout have been substantial, and achieving the desired level of patient adherence may be challenging [17]. Relating

usage to intended use requires operationalization and rationalization [37]. Approaches to measuring adherence vary; reported methods include the number of log-ins, website exposure, and modules completed; these stem partly from diversity in the purpose and design of digital interventions but also demonstrate different views of the concept of adherence [38].

High levels of adherence to symptom reporting during cancer treatment have been observed in relation to clinic visits, but there is a lack of large-scale studies examining how patients undergoing treatment for cancer adhere to and perceive symptom monitoring and reporting via a mobile app [7,9].

Treatment for breast cancer consists of different approaches such as chemotherapy, surgery, and radiotherapy [39-41]. Chemotherapy can be administered as adjuvant after surgery, but in recent years, neoadjuvant chemotherapy, which is administered before surgery, has become more common [42]. Neoadjuvant chemotherapy is administered at the oncology clinic at different treatment intervals, depending on the cytotoxic drugs that are given [43]. Treatment for prostate cancer includes three main approaches: active surveillance, surgical treatment, and radiotherapy. These may be combined with antihormonal treatment. Radiotherapy for prostate cancer is administered every weekday at clinics [44]. In Sweden, patients remain at home between treatments. Patients with breast cancer undergoing neoadjuvant chemotherapy meet the physician before each treatment cycle, approximately every second or third week, depending on the chemotherapy regimen [43]. Patients with prostate cancer undergoing radiotherapy meet the physician before the start of radiotherapy and 6 months after completing radiotherapy [44]. Patients are assigned a contact nurse who is responsible for the patient's care throughout the care chain and who the patient can contact during office hours in case of concerns related to the treatment [43,44]. Patients are also provided with telephone numbers for the oncology clinics, for making contact during office hours or during evenings, nights, and weekends.

As most patients undergoing neoadjuvant chemotherapy for breast cancer or radiotherapy for prostate cancer are treated on an outpatient basis, partially self-reliant management of symptoms is necessary. Several studies show decreased symptom burden [24,45], that patients appreciate using Interaktor, and that they feel secure [21,26]. The app is currently being evaluated in two randomized controlled trials that include patients with breast cancer and prostate cancer, with the hypothesis that using the app will improve symptom management, reduce symptom burden, and increase cost-effectiveness through a reduced consumption of health care services in comparison to standard care alone [20].

An understanding of how Interaktor is used and perceived, including the impacts of individual factors, is warranted in order to support the interpretation of the clinical effects of using the app during treatment [45]. Therefore, this study aimed to describe engagement with the Interaktor app among patients with breast and prostate cancer during their treatment.

Methods

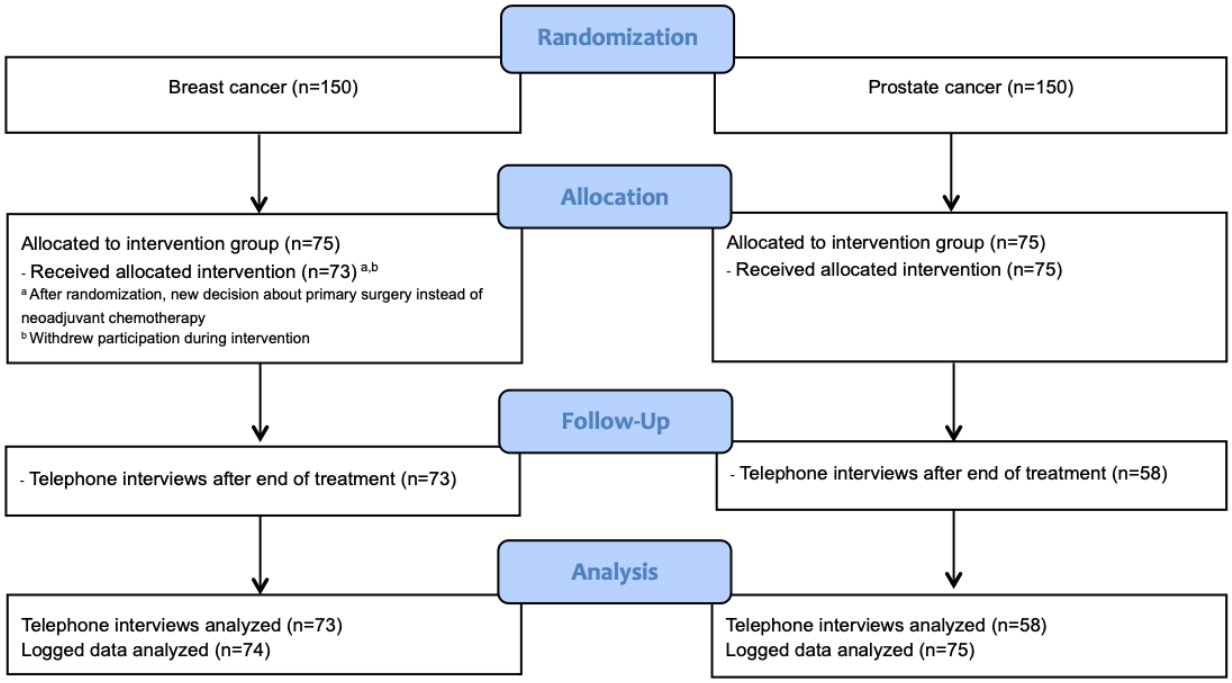
Overview

This study included patients from the intervention groups of two separate randomized controlled trials: patients with breast cancer during neoadjuvant chemotherapy (ClinicalTrials.gov; NCT02479607) and patients with locally advanced prostate cancer during treatment with radiotherapy (ClinicalTrials.gov; NCT02477137). The study comprised logged data from patient reports and interviews with patients. Ethical approval was obtained from the Regional Ethical Review Board of Stockholm (registration no. 2013/1652-31/2 and 201712519-32).

Setting and Sample

The patients were consecutively recruited at two university hospitals in Stockholm, Sweden. Patients with breast cancer or prostate cancer undergoing neoadjuvant chemotherapy and radiotherapy, respectively, who were able to speak and understand Swedish, who were presumed cognitively able to use a mobile app for symptom reporting, who agreed to participate, and who signed a written informed consent form were eligible. Of the 75 patients in the breast cancer group, 1 patient had a change of treatment to surgery instead of neoadjuvant chemotherapy, and 1 patient withdrew their consent to participate in the interview. In the prostate cancer group, 58 patients of the 75 patients participated in the interviews, due to organizational circumstances and difficulties reaching the patients after their treatment had ended (Figure 1).

Figure 1. CONSORT diagram.



Study Procedure

Patients downloaded Interaktor onto their smartphone or tablet. Patients who did not own one were lent a smartphone by the research group (2 in the breast cancer group, 2 in the prostate cancer group). The patients also received an individual log-in for access to the content of the app. The patients received verbal and written instructions on how to use the app and were asked to report their symptoms daily on weekdays during treatment. Patients with breast cancer were asked to start using the app on their first day of neoadjuvant chemotherapy and continue until 2 weeks after treatment had ended—a total of approximately 18 weeks. Patients with prostate cancer started using the app on their first day of radiotherapy and continued until 3 weeks after treatment had ended—a total of approximately 9 weeks. All patients were informed that nurses at the clinic would survey and respond to alerts triggered on weekdays (8 AM to 4 PM). If patients needed support at other times of the day, they were instructed to contact the clinic according to the standard procedure. If a report had not been submitted before 2 PM, a notification was sent out to remind the patient to report. The breast cancer version of the app included a notification that was sent to the patient, suggesting the patient read self-care advice related to alerted symptoms. This function was not included in the prostate cancer version of the app.

Data Collection

Participant Data

Sociodemographic data were obtained from baseline questionnaires. Clinical data and prevailing health status at the time of treatment start were collected from medical records. Health status was used to calculate the comorbidity score using the Charlson Comorbidity Index, which encompasses 19 medical conditions. Each condition has a score based on a relative risk of death within a year. The scores are totaled to yield the comorbidity score, with a range between 0 and 37. A higher value corresponds to greater comorbidity [46].

Logged Data

Data on app usage including symptom reports, triggered alerts, views of self-care advice, and free text messages sent, were made accessible to the researchers through encrypted Excel (2013; Microsoft Inc) files. At the time of the study, it was not possible for data on clicked links or viewed graphs to be logged for later retrieval.

Interviews

Telephone interviews were conducted with the patients shortly after the end of the use of the app, following a semistructured interview guide (Table 1) that we had developed, focusing primarily on usability, utility, and capturing participant feedback on the app (including all components). The interviews lasted between 10 and 20 minutes, during which the authors made notes of the answers in the template of the interview guide. Five of the patients with prostate cancer were individually interviewed in connection to more comprehensive interviews about participation in care. These interviews were recorded and transcribed verbatim, but only the text with respect to using the app was analyzed in this study.

Table 1. Interview guide.

	Question	Follow-up question
1	What was it like to report in the app? (in general)	Easy/difficult? Advantages/disadvantages?
2	What was it like to report your symptoms?	Easy/difficult? Absence of symptom to report?
		Relevant symptom questions and reporting frequency?
3	Have you used the self-care advice? (yes/no)	If yes, your experience? Relevant self-care advice? Have you used the links?
		If no, have you searched for information elsewhere?
4	How have you experienced that the technology has worked?	The log-in procedure? Mobile coverage?
5	Have you used the graphs to follow your symptom history? (yes/no)	If yes, in which way? How did you experience the graphs?
		If no, why?
6	Were you called sometime by a nurse? (yes/no)	If yes, was it after an alert?
7	How did you experience being called by a nurse after an alert?	—
8	Is there anything else you want to add?	Is there anything in the app that you have lacked?

Statistical Analyses

Data management was performed using Excel, and statistical analyses were carried out using SPSS statistical software (version 24.0; IBM Corp). Differences in demographic and clinical characteristics between the two groups were analyzed using two-tailed independent *t* test, Fischer exact test, and the chi-square test. Usage was analyzed with descriptive and inferential statistics on the variables reports sent, alerts triggered, self-care advice section views, and free text messages sent; group median values were calculated and are reported. The intended use was measured by daily symptom reporting during weekdays. Adherence to daily reporting was calculated as the number of weekday reports (excluding multiple daily reports) divided by the total number of reportable weekdays for each patient.

Multiple regression analysis was conducted (using the enter method) to see if the independent variables (predictors) age, comorbidity, marital status, and education level predicted usage of the app. The usage variables (dependent variables) were adherence to daily reporting as intended, total number of alerts triggered, total views on self-care advice, and total number of free text messages sent. Since all of the usage variables were positively skewed, these were normalized using a natural logarithm transformation [47]. Level of significance was determined as $P<.005$.

Qualitative Analysis

The interview notes were analyzed by conventional content analysis [48]. First, two authors read the complete interview notes from each group (breast cancer and prostate cancer) individually, to grasp the entire data. Both authors compiled a data sheet for each group with the patients' answers. The data sheets were reviewed repeatedly, and initial codes were derived. Thereafter, all authors reviewed both sets of data sheets and codes. Due to substantial similarities of codes in the two groups (breast cancer and prostate cancer), one data set containing all coded responses (with credentials) was assembled. Subsequently, all text material were analyzed. In an iterative process, the codes were sorted based on similarity into subcategories and combined based on content into overarching categories. Some exemplifying quotes from the patients' statements are presented in the Results section. During the analytical process, all authors met continually to discuss and come to a consensus.

Results

Sample Characteristics

The patients' sociodemographic (Table 2) and clinical data (Table 3) are presented below. Patients in the breast cancer group were significantly younger than the patients in the prostate cancer group ($P<.001$). Moreover, the patients in the breast cancer group had a statistically significant lower comorbidity score ($P<.001$) and a higher self-reported education level ($P=.005$).

Table 2. Sociodemographic characteristics at baseline of patients with breast cancer and prostate cancer.

Characteristics	Breast cancer (n=74)	Prostate cancer (n=75)	Test statistic	*P* value
Age (in years), median (range)	47 (27-73)	72 (44-81)	–15.127[a]	<.001[a]
Marital status, n (%)	74 (100)	71 (100)	1.204 (2)[b]	.57[b]
Married/cohabitant	58 (78)	53 (74)		
Living apart	3 (4)	6 (9)		
Single	13 (18)	12 (17)		
Highest education level, n (%)	74 (100)	71 (100)	10.627 (2)[c]	.005[c]
University	50 (68)	30 (42)		
Secondary school	18 (24)	25 (35)		
Primary school	6 (8)	16 (23)		
Occupation, n (%)	74(100)	69 (100)	61.330 (2)[b]	<.001[b]
Working	57 (77)	22 (32)		
Sick leave	12 (16)	0 (0)		
Retired/unemployed	5 (7)	47 (68)		

[a]Mann-Whitney *U* test.

[b]Fischer exact test (*df*).

[c]Chi-square (*df*).

Table 3. Clinical characteristics at baseline of patients with breast cancer and prostate cancer.

Characteristics	Breast cancer (n=74)	Prostate cancer (n=75)	Test statistic	*P* value
Charlson Comorbidity Scale score[a], median (IQR)	1.0 (1)	3.0 (1)	5082.5[b]	<.001[b]
PSA[c] (at start of treatment), median (range)	N/A[d]	5.3 (0-53)		
Disease stage (TNM[e]), n (%)				
T1	N/A	1 (1)		
T1C	N/A	17 (23)		
T2	N/A	11 (15)		
T2A	N/A	1 (1)		
T2B	N/A	7 (9)		
T2C	N/A	7 (9)		
T3	N/A	20 (27)		
T3A	N/A	1 (1)		
T3B	N/A	8 (11)		
Missing	N/A	2 (3)		
Histologic grade (Elston-Ellis), n (%)				
Intermediate grade 2	23 (31.1)	N/A		
High grade 3	41 (55.4)	N/A		
Unknown	10 (13.5)	N/A		
Tumor characteristics[f], n (%)				
HER2+ ER+ PR+	9 (12.2)	N/A		
HER2+ ER+ PR–	7 (9.5)	N/A		
HER2+ ER– PR–	13 (17.6)	N/A		
HER2– ER+ PR+	16 (21.6)	N/A		
HER2– ER+ PR–	7 (9.5)	N/A		
HER2– ER– PR+	1 (1.4)	N/A		
Triple negative	21 (28.4)	N/A		
Proliferation rate (Ki-67), n (%)				
≥ 20 %	72 (97.3)	N/A		
< 20 %	2 (2.7)	N/A		
Type of chemotherapy, n (%)				
Anthracyclines, alkylators	3 (4.1)	N/A		
Anthracyclines, alkylators, taxanes	37 (50.0)	N/A		
Anthracyclines, alkylators, antimetabolites, taxanes	9 (12.2)	N/A		
Antimetabolites, alkylators	2 (2.7)	N/A		
Taxanes	11 (14.9)	N/A		
Taxanes, alkylators	1 (1.4)	N/A		
Taxanes, trastuzumab emtansine	2 (2.7)	N/A		
Trastuzumab emtansine	9 (12.2)	N/A		
Gleason score, median (range)	N/A	7 (6-9)		
Number of radiotherapy treatments, median (range)	N/A	25 (25-29)		
Number of neoadjuvant chemotherapy treatments, median (range)	6.0 (2-15)	N/A		

Characteristics	Breast cancer (n=74)	Prostate cancer (n=75)	Test statistic	*P* value
Antihormonal treatment, n (%)				
Yes	N/A	55 (73)		
No	N/A	20 (27)		
Treatment duration (in weeks), median (range)	15 (3-26)	5 (5-6)		

[a]A higher value corresponds to greater comorbidity; the score ranges between 0-37.

[b]Mann-Whitney *U* test.

[c]PSA: Prostate-specific antigen.

[d]N/A: not applicable.

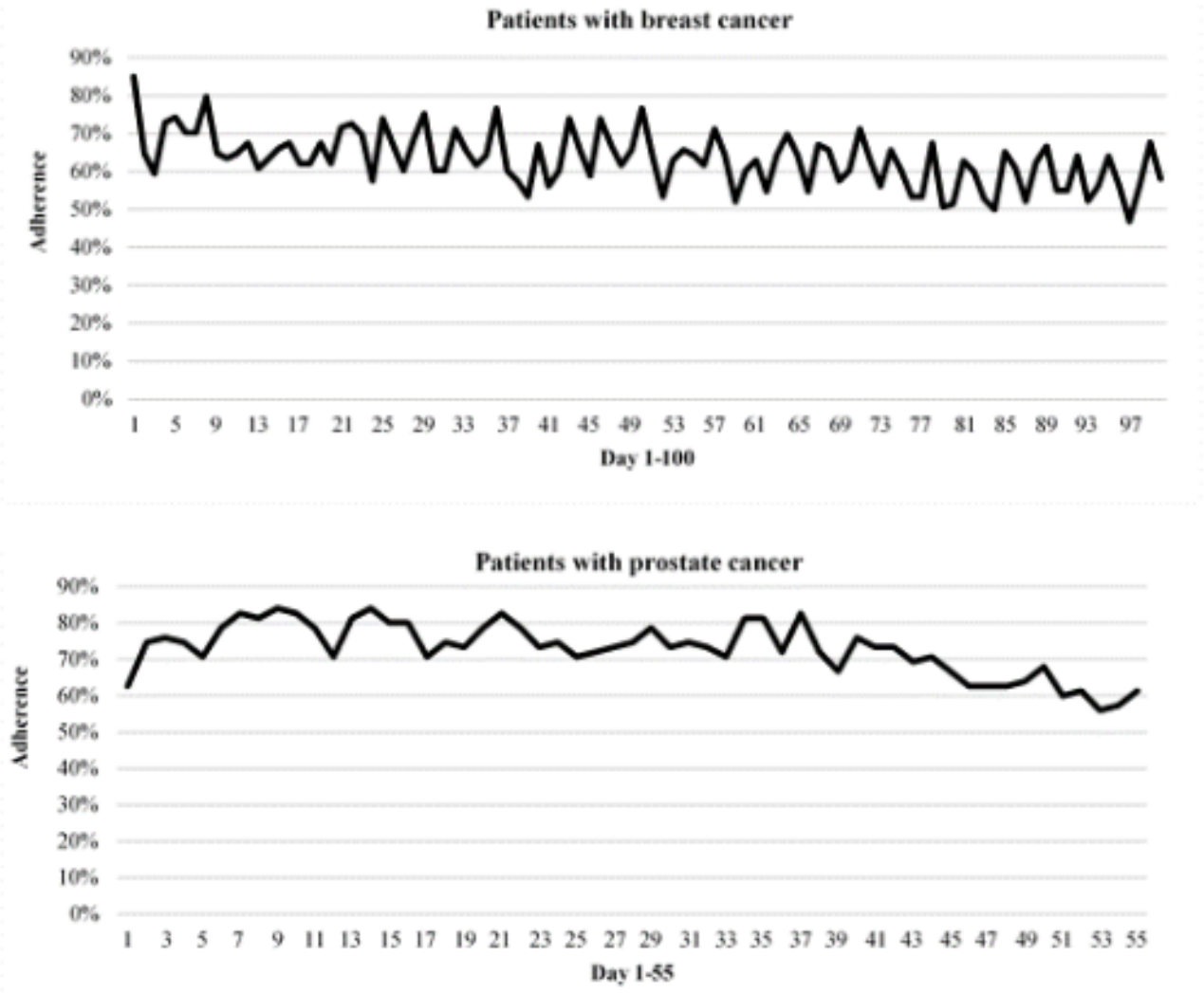
[e]TNM: Classification of malignant tumors.

[f]Breast cancer cell human epidermal growth factor receptor (HER2), estrogen receptor (ER), and progesterone receptor (PR) status.

Logged Data

All patients reported with the app at least once during the study period. Due to differences in treatment schedules, the reporting period in the breast cancer group ranged from 22 to 183 days (median 106, IQR 7). The reporting period in the prostate cancer group ranged from 54 to 89 days (median 63, IQR 11). Median adherence in the breast cancer group was 83% (IQR 36%). In the prostate cancer group, the median adherence percentage was also 83% (IQR 34%). Graphs of adherence patterns over time show that the level of adherence remained stable over time, although it dropped somewhat after day 49 in the breast cancer group and day 39 in the prostate cancer group (Figure 2).

Figure 2. Adherence to symptom reporting over time.



In the breast cancer group, 96% (71/74) of patients triggered at least one alert during the study period; in the prostate cancer group, 72% (54/75). Patients in the breast cancer group triggered a median of 7 alerts (IQR 21, range 1-210). In the prostate cancer group, the median number of alerts triggered was 2 (IQR 9, range 1-60). The distribution of yellow and red alerts was 86% yellow to 14% red in the breast cancer group and 90% yellow to 10% red in the prostate cancer group.

Among patients in the breast cancer group, 100% (74/74) viewed self-care advice at least once. Patients in the breast cancer group viewed a median of 11 (IQR 15) self-care advice topics at least once during the study period, out of 17 self-care advice available. The total number of views was 1075, and 34% (362/1075) were views made after the patient was notified with the suggestion to read related self-care advice after an alert had been triggered. In the prostate cancer group, 87% (65/75) of patients viewed self-care advice at least once and the median number of various self-care advice topics viewed at least once was 5 (IQR 11) out of 16 self-care advice available. The total number of self-care advice views by patients in the prostate cancer group was 697.

Most of the patients used the free text function at least once: 93% (69/74) in the breast cancer group and 75% (56/75) in the prostate cancer group. The free text messages were mainly about symptoms, requesting or declining contact, care-related information, and issues linked to the app or to reporting. There was a variation in how the function was used, some patients wrote short, condensed messages while some wrote longer, richer descriptions.

Patient Characteristics as Predictors of App Usage

In the breast cancer group none of the multiple regression models were statistically significant (Table 4). Higher age predicted a lower total number of free text messages sent (P=.04). In the prostate cancer group the multiple regression model showed that the total number of views on self-care advice was statistically significant ($F_{4,70}$=3.811, P=.007, adjusted R^2=.132) (Table 5). Higher age (P=.01) and higher education level (P=.04) predicted a higher total number of views of self-care advice. Higher comorbidity score predicted fewer self-care advice views (P=.004). Furthermore, being married or cohabitating predicted a higher adherence to daily symptom reporting (P=.02) (Table 5).

Table 4. Multiple regression in patients with breast cancer (n=74).

Dependent and independent variables	B[a]	SE[b]	β[c]	95% CI		*P* value[d]	Adj[e] R^2	*P* value[f]
				Lower	Upper			
Adherence to daily reporting as intended							<.001	.72
Age	–0.006	0.010	–.080	–.027	.015	.56		
Comorbidity	0.094	0.099	.129	–.103	.291	.35		
Marital status	0.039	0.230	.020	–.420	.497	.87		
Educational level	0.224	0.208	.132	–.191	.639	.29		
Total number of alerts							.038	.16
Age	–0.027	0.015	–.243	–.056	.003	.07		
Comorbidity	–0.087	0.140	–.082	–.367	.193	.54		
Marital status	–0.387	0.326	–.137	–1.04	.264	.24		
Educational level	–0.116	0.296	–.046	–.705	.474	.70		
Total views on self-care advice							<.001	.46
Age	–0.007	0.011	–.084	–.028	.015	.54		
Comorbidity	–0.101	0.102	–.134	–.303	.102	.33		
Marital status	0.131	0.236	.066	–.340	.603	.58		
Educational level	0.093	0.214	.053	–.334	.520	.66		
Total number of free text messages							.070	.07
Age	–0.027	0.013	–.295	–.052	–.001	.04		
Comorbidity	0.036	0.114	.044	–.191	.264	.75		
Marital status	–0.030	0.271	–.013	–.571	.511	.91		
Educational level	0.326	0.238	.165	–.149	.800	.18		

[a]B: unstandardized coefficient.

[b]SE: standard error of the unstandardized coefficient.

[c]β: standardized coefficient.

[d]*P* value for the independent variable.

[e]Adj: adjusted for the multiple regression model.

[f]*P* value for the multiple regression model.

Table 5. Multiple Regression in patients with prostate cancer (n=75)

Dependent and independent variables	B[a]	SE[b]	β[c]	95% CI		*P* value[d]	Adj[e] R^2	*P* value[f]
				Lower	Upper			
Adherence to daily reporting as intended							.027	.21
Age	–0.001	0.015	–.008	–.030	.028	.96		
Comorbidity	0.020	0.077	.038	–.134	.173	.80		
Marital status	0.431	0.181	.277	.071	.792	.02		
Educational level	–0.026	0.165	–.018	–.355	.304	.88		
Total number of alerts							< .001	.66
Age	–0.011	0.026	–.067	–.062	.040	.66		
Comorbidity	0.032	0.134	.036	–.236	.300	.81		
Marital status	0.482	0.316	.182	–.148	1.11	.13		
Educational level	0.082	0.289	.033	–.494	.657	.78		
Total views on self–care advice							.132	.007
Age	0.057	0.022	.368	.013	.101	.01		
Comorbidity	–0.347	0.116	–.419	–.579	–.115	.004		
Marital status	0.019	0.273	.008	–.526	.565	.94		
Educational level	0.531	0.250	.232	.032	1.03	.04		
Total number of free text messages							< .001	.84
Age	–0.008	0.024	–.052	–.055	.039	.74		
Comorbidity	–0.029	0.125	–.035	–.278	.220	.82		
Marital status	0.229	0.293	.094	–.356	.814	.44		
Educational level	–0.161	0.268	–.071	–.696	.374	.55		

[a]B: unstandardized coefficient.

[b]SE: standard error of the unstandardized coefficient.

[c]β: standardized coefficient.

[d]*P* value for the independent variable.

[e]Adj: adjusted for the multiple regression model.

[f]*P* value for the multiple regression model.

Perceptions of Using the Interaktor App

Overall

Although all patients were enthusiastic about contributing to the study and to the evaluation and development of the app, some patients noted that they did not recall using the app in detail. Analysis of the interviews resulted in three overarching categories: user friendliness, interaction with the health care professionals, and support for self-care.

User Friendliness

Nearly all patients stated that the app was easy to use and that it took little time to learn how to use it. Few experienced technical problems.

Reporting went quickly, and the app was described as a fast and comfortable way to get support and help. Most patients agreed that the app content was relevant and that the symptoms included in the app covered most symptoms they experienced during their treatment. Only some patients wanted to report symptoms that were not included in the app, such as headache and weight gain.

Symptom reporting was often described as more meaningful and interesting in the beginning when symptoms were new and there was a feeling of apprehension and insecurity of what the treatment would entail. When symptoms had become a fact of everyday life, reporting them via the app daily did not always feel necessary. A few patients described daily symptom reporting as something they perceived as a negative reminder of illness, especially during times when they felt alright.

> *It has been very easy and convenient.…The app is easy to use.…When you feel ill it is a security, but if you feel good it is a negative reminder that you are sick.* [BI-10]

Remembering to report symptoms was sometimes perceived as difficult, particularly when patients felt well. Incorporating or establishing a routine around reporting facilitated recollection, as did the automatic reminder notifications, which were greatly

appreciated. Some patients requested the ability to adjust the reporting time according to personal preferences.

In the breast cancer group, daily reporting could sometimes be stressful and difficult, especially when they felt ill or lacked energy; nevertheless, most patients considered reporting to be more necessary at those times. Furthermore, memory impairment was often described as linked to the cognitive side effects of treatment and some mentioned that graphs were difficult to interpret because the text was too small. In the prostate cancer group, some patients did not notice or use the self-care advice, links, or graphs due to lack of experience with mobile apps or due to forgetfulness. Patients suffering from comorbidities that had symptoms similar to the questions included in the app perceived reporting as difficult to answer, given the available responses.

Interaction With Health Care Professionals

Reporting symptoms generated a feeling of having continuous contact with health care professionals. It also created feelings of being safe, monitored, acknowledged, involved, and cared for. Most patients described being called by a nurse after an alert in positive terms and said it decreased the need for contacting the oncological clinic through other channels. They said that it was a great benefit to not have to call, leave a message on the answering machine, be put on hold, or have to search for the right phone number. On a few occasions, patients were not contacted after alerts.

> *I felt safe reporting every day....It was excellent.…I noticed, before logging off, that I would be called...A huge security.…Having the app and a continuous access to help has made me feel better.* [BI-44]

In the breast cancer group, some patients expressed that alerts should be monitored and responded to around the clock, and several wanted the possibility to choose themselves whether a nurse should call them. Some preferred calling health care professionals, especially when they felt very ill. A small number of patients in the breast cancer group perceived themselves as nagging or bothering the nurses when alerts were triggered.

Some patients indicated a wish to continue reporting after the trial period had ended. They described feelings of being alone when the treatment ended, and they no longer met their health care professionals regularly. At that time, the value of symptom reporting was perceived to increase.

The free text function was generally appreciated and was perceived as useful for reporting additional symptoms and information to the nurse.

Support for Self-Care

The patients perceived the app as supportive during their treatments and described symptom reporting in terms of diary keeping. Reporting symptoms encouraged and supported reflection on their well-being and made patients more aware of symptoms and what they should observe for.

A majority of the patients perceived the self-care advice as valuable, applicable, and informative, especially when a symptom first occurred. The self-care advice gave answers on how to perform self-care to relieve or manage symptoms for themselves. Reading the advice also gave them an idea of what was normal and what they should expect during treatment. There were requests to add more comprehensive information on psychological symptoms and dietary advice. The patients in the breast cancer group described the graphs as useful for comparing symptoms over time and detecting patterns in the symptoms related to the cytotoxic treatment intervals. This could facilitate the planning of activities and enable them to do things during days they felt well. The patients in the prostate cancer group often commented that the graphs enabled them to monitor their well-being by displaying when symptoms increased or decreased and also helped them discern that many days were trouble free.

It was said about the symptom history graphs—

> *It was fun to see that it was getting better… and it trailed with how you perceived to be feeling.* [PI-187]
>
> *I did not follow them.…I want to move on.…Now it is just forward ahead.* [PI-143]

The patients in the breast cancer group perceived the links as useful for gaining further or in-depth information and support. Patients in the prostate cancer group were more likely to describe the self-care advice, links, and graphs as superfluous when they experienced mild or less persistent symptoms.

Discussion

Principal Results

The findings of this study show that the app was largely used as intended and appreciated by patients undergoing treatment for breast cancer and prostate cancer. The app gave the patients a feeling of assurance by offering a convenient method to contact their health care professionals and the security of being monitored via the symptom reports. Furthermore, the app promoted self-care by facilitating self-monitoring and learning about symptoms.

In this study, adherence to daily symptom reporting in the app was 83%. This is high in comparison to a review [49] in which it was concluded that around half of participants in web-based health interventions for chronic conditions and lifestyle and mental health management adhered to the interventions. Plausible explanations for the high adherence to the Interaktor app are that the app was interactive and easily available on a smartphone or tablet when compared to a computer-based system. Previous research [49,50] shows that intervention characteristics such as intended usage frequency, updates, and persuasive design increase patients' adherence to web-based interventions. Moreover, the app may reflect patients' need for frequent, continuous contact with health care professionals during treatment for cancer [51].

Most patients perceived the content of the Interaktor app as relevant, and the fact that Interaktor was developed in collaboration with patients and health care professionals is likely to have contributed to this result. Two previous studies [21,26] of the Interaktor app among patients treated for cancer yielded results in line with this study, both in terms of adherence levels above 80% and interviews revealing that patients appreciate

and perceive the app as supportive in their symptom management.

There was a temporal aspect to how patients perceived using the app. At the beginning of treatment, when the situation was new or they experienced a new symptom, daily symptom reporting was considered especially meaningful. Later, as patients became more experienced and familiar with their symptoms, some noted a lower inclination to report each day. Also, patients commented that after some time, as they acquired general knowledge about the disease and treatment, they felt a desire for more individualized and in-depth information than that contained in the app.

The observation that memory impairment and feeling ill influenced engagement adversely contrasts with a study [52] that showed that increased use of a web-based symptom management system was predicted by higher levels of symptom distress among men with prostate cancer, and it may be of value to investigate the effect of symptom burden on app usage in future studies.

The interviews indicated that some patients with comorbidities felt the need to add clarifications to responses available in the self-assessment form by free text. This finding warrants further study, considering the expanding community with multiple conditions, and it contrasts somewhat from the findings of a study [52] that showed that increased use of the system was associated with the absence of comorbidities.

The findings described above are in line with a conceptual model where perceived credibility and personal relevance influence engagement [36]. This study accumulates existing evidence, which imply the need to develop cancer-supportive digital interventions that are interactive and tailored [53]. Tailoring can be performed by the individual patient before the intervention (pretailoring), by preference to promote autonomy (self-tailoring), and within-person as health status or needs evolve [54]. In future technological development of the app, it might be useful to integrate an option for the individual patient to add or exclude symptoms or concerns in the self-assessment component to make it more person-centered. Expanding the interactive features in Interaktor by adding or updating information as patients progress through treatment may also be a way to maintain and enhance patients' experience of personal relevance [36].

The adjusted R^2 values in the regression models in this study were low, indicating that none or only one predictor was correlated to each dependent variable. The only variable tested that predicted usage for both groups was age; higher age predicted a decrease of the total number of free text messages sent in the breast cancer group, while a higher age predicted more self-care advice views in the prostate cancer group. It has previously been suggested that demographic variables may be too broad to indicate usage motives and preferences for mHealth [55]. Higher education level predicted usage in the prostate cancer group, specifically, predicting a higher number of views of self-care advice. Furthermore, in the prostate cancer group, being married or cohabiting predicted higher adherence to daily reporting as intended than that of patients who were single. These findings are in line with theory as well as research suggesting that social support and education level influence the adoption and usage of web-based interventions [17,56]. Both social support and higher education level have been associated with a significant increase in engagement [15].

In the breast cancer group, it was suggested by patients that they should be responsible for contacting health care professionals. This finding is noteworthy, as it pertains to patients' sense of control and self-reliance. From the perspective of a review on person-centered participation [57], this may signify a patient with confidence in their own experience-based knowledge and expertise, voicing a willingness for increased responsibility and management of their care. This is an essential precondition if we are to achieve equality and partnership among health care professionals and patients, but it is always preceded by phases of dialogue, knowledge building, and information sharing [57]. Previous studies have demonstrated that being female, having a higher education level, and being younger are predictive for preferring a more active role in care [58,59].

Strengths and Limitations

Analysis of usage metrics enables a systematic assessment of, and insight into, patients' exposure to and behavior throughout an intervention relating to frequency, depth, and breadth of use [17] but cannot disclose how use was perceived [60]. On the other hand, interviews or self-reports are associated with a risk of social desirability and challenges in communication, such as discrepancies in researcher and respondent terminology [60].

To counteract the limitations of each method, a combination was used in this study. Below are two examples of findings that reinforce how relying on only one source of data to study engagement is not sufficient. In the interviews, patients described aspects that were perceived to influence their engagement, which were not visible in the logged data. Patients stated that remembering to report daily was sometimes difficult, and moreover, patients noted that daily symptom reporting was perceived as more meaningful in the beginning. But the logged data show that adherence to daily symptom reporting is high and stable throughout the time of the intervention as adherence patterns do not decrease markedly over time. Relying on either one of the methods would have resulted in a less comprehensive account of patient engagement.

It may be a limitation that all the interviews took place after the intervention was completed and collecting data during the intervention as well as afterward may have added additional insights relating to usability and learning [61]. Symptom graphs and self-care advice features went undiscovered by a small number of patients in the significantly older prostate cancer group, indicating that enhanced patient training and further reviews of usability may be profitable in future studies. Research contributing to an understanding of how adherence can be promoted is starting to emerge [50,61,62]. However, a comparison of adherence levels between studies is debatable due to limited consistency in how adherence is reported [37,38]. The results of this study provide valuable knowledge, as it relates both to the patients' actual usage versus intended use of the intervention and to how the patients thought about using it. The strength of this study consolidating investigations of two

distinct patient groups is somewhat limited as the two versions of Interaktor slightly differed. The prostate cancer version of the app was developed and clinically trialed before the planning of this study and did not have the function that automatically prompted reading of specific self-care advice based on symptom reports. It is also a limitation that the use of links and graphs were not logged. Both are, in a sense, reflective of a learning curve in innovation research and will be taken into consideration in forthcoming studies.

Conclusions

High patient engagement in the Interaktor app was achieved. Using the app promotes self-care by facilitating self-monitoring and timely advice. Furthermore, it provides assurance through continuous and convenient contact with health care professionals.

The study supports the notion that interactivity enhances patients' feelings of personal relevance and thus increases engagement. The predictive ability of demographic variables differed between patient groups, but higher age and a higher educational level predicted the usage of specific app functions for both patient groups.

Taken together, the findings suggest a role for the use of an interactive app, such as Interaktor, to promote patients' participation in their care. The findings may have relevance for outpatients undergoing other cancer treatments associated with a risk of toxicities.

Acknowledgments

The study was supported by grants from *Karolinska Institutet*; the Swedish Research Council; the Swedish Research Council for Health, Working Life and Welfare; the Swedish Cancer Foundation; the Kamprad Family Foundation for Entrepreneurship, Research and Charity; and the Cancer Research Funds of Radiumhemmet. The authors would like to thank the patients who participated in this study and the health care professionals at the oncology clinics who assisted us. We would also like to thank Health Navigator for technical support and collaboration, and Jan Kowalski, statistician, who provided expertise and support during the statistical analysis process.

Conflicts of Interest

None declared.

References

1. McKenzie H, Hayes L, White K, Cox K, Fethney J, Boughton M, et al. Chemotherapy outpatients' unplanned presentations to hospital: a retrospective study. Support Care Cancer 2011 Jul;19(7):963-969. [doi: 10.1007/s00520-010-0913-y] [Medline: 20499108]
2. Lai XB, Ching SSY, Wong FKY. A qualitative exploration of the experiences of patients with breast cancer receiving outpatient-based chemotherapy. J Adv Nurs 2017 Oct;73(10):2339-2350. [doi: 10.1111/jan.13309] [Medline: 28370172]
3. Hinami K, Alkhalil A, Chouksey S, Chua J, Trick WE. Clinical significance of physical symptom severity in standardized assessments of patient reported outcomes. Qual Life Res 2016 Sep;25(9):2239-2243. [doi: 10.1007/s11136-016-1261-2] [Medline: 26980417]
4. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
5. Kearney N, McCann L, Norrie J, Taylor L, Gray P, McGee-Lennon M, et al. Evaluation of a mobile phone-based, advanced symptom management system (ASyMS) in the management of chemotherapy-related toxicity. Support Care Cancer 2009 Apr;17(4):437-444. [doi: 10.1007/s00520-008-0515-0] [Medline: 18953579]
6. Ruland CM, Andersen T, Jeneson A, Moore S, Grimsbø GH, Børøsund E, et al. Effects of an internet support system to assist cancer patients in reducing symptom distress: a randomized controlled trial. Cancer Nurs 2013;36(1):6-17. [doi: 10.1097/NCC.0b013e31824d90d4] [Medline: 22495503]
7. Rincon E, Monteiro-Guerra F, Rivera-Romero O, Dorronzoro-Zubiete E, Sanchez-Bocanegra CL, Gabarron E. Mobile phone apps for quality of life and well-being assessment in breast and prostate cancer patients: systematic review. JMIR Mhealth Uhealth 2017 Dec 04;5(12):e187 [FREE Full text] [doi: 10.2196/mhealth.8741] [Medline: 29203459]
8. Giunti G, Giunta DH, Guisado-Fernandez E, Bender JL, Fernandez-Luque L. A biopsy of Breast Cancer mobile applications: state of the practice review. Int J Med Inform 2018 Feb;110:1-9 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.022] [Medline: 29331247]
9. Warrington L, Absolom K, Conner M, Kellar I, Clayton B, Ayres M, et al. Electronic systems for patients to report and manage side effects of cancer treatment: systematic review. J Med Internet Res 2019 Jan 24;21(1):e10875 [FREE Full text] [doi: 10.2196/10875] [Medline: 30679145]
10. Osborn J, Ajakaiye A, Cooksley T, Subbe CP. Do mHealth applications improve clinical outcomes of patients with cancer? a critical appraisal of the peer-reviewed literature. Support Care Cancer 2020 Mar;28(3):1469-1479 [FREE Full text] [doi: 10.1007/s00520-019-04945-4] [Medline: 31273501]

11. Richards R, Kinnersley P, Brain K, McCutchan G, Staffurth J, Wood F. Use of mobile devices to help cancer patients meet their information needs in non-inpatient settings: systematic review. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e10026 [FREE Full text] [doi: 10.2196/10026] [Medline: 30552082]
12. Boulos MNK, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
13. Gomolin A, Lebouché B, Engler K, Vedel I. Optimizing smartphone intervention features to improve chronic disease management: A rapid review. Health Informatics J 2019 Dec 12:1460458219891377. [doi: 10.1177/1460458219891377] [Medline: 31826703]
14. Doherty K, Doherty G. Engagement in HCI. ACM Comput. Surv 2018 Jan 23;51(5):1-39. [doi: 10.1145/3234149]
15. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
16. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J. Am. Soc. Inf. Sci 2008 Apr 28;59(6):938-955. [doi: 10.1002/asi.20801]
17. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
18. Olsson L, Jakobsson UE, Swedberg K, Ekman I. Efficacy of person-centred care as an intervention in controlled trials - a systematic review. J Clin Nurs 2013 Feb;22(3-4):456-465. [doi: 10.1111/jocn.12039] [Medline: 23231540]
19. Frank C, Asp M, Fridlund B, Baigi A. Questionnaire for patient participation in emergency departments: development and psychometric testing. J Adv Nurs 2011 Mar;67(3):643-651. [doi: 10.1111/j.1365-2648.2010.05472.x] [Medline: 21166838]
20. Langius-Eklöf A, Crafoord M, Christiansen M, Fjell M, Sundberg K. Effects of an interactive mHealth innovation for early detection of patient-reported symptom distress with focus on participatory care: protocol for a study based on prospective, randomised, controlled trials in patients with prostate and breast cancer. BMC Cancer 2017 Jul 04;17(1):466 [FREE Full text] [doi: 10.1186/s12885-017-3450-y] [Medline: 28676102]
21. Gustavell T, Langius-Eklöf A, Wengström Y, Segersvärd R, Sundberg K. Development and feasibility of an interactive smartphone app for early assessment and management of symptoms following pancreaticoduodenectomy. Cancer Nurs 2018 Mar 27:e1-e10. [doi: 10.1097/NCC.0000000000000584] [Medline: 29596113]
22. Göransson C. Developing and evaluating an interactive app to support self-care among older persons receiving home care. Doctoral dissertation. In: Developing and evaluating an interactive app to support self-care among older persons receiving home care. Örebro: Örebro University; 2019.
23. Blomberg K, Wengström Y, Sundberg K, Browall M, Isaksson A, Nyman MH, et al. Symptoms and self-care strategies during and six months after radiotherapy for prostate cancer - scoping the perspectives of patients, professionals and literature. Eur J Oncol Nurs 2016 Apr;21:139-145. [doi: 10.1016/j.ejon.2015.09.004] [Medline: 26482003]
24. Sundberg K, Wengström Y, Blomberg K, Hälleberg-Nyman M, Frank C, Langius-Eklöf A. Early detection and management of symptoms using an interactive smartphone application (Interaktor) during radiotherapy for prostate cancer. Support Care Cancer 2017 Feb 24:2195-2204. [doi: 10.1007/s00520-017-3625-8] [Medline: 28236145]
25. Gustavell T, Sundberg K, Frank C, Wengström Y, Browall M, Segersvärd R, et al. Symptoms and self-care following pancreaticoduodenectomy: perspectives from patients and healthcare professionals - Foundation for an interactive ICT application. Eur J Oncol Nurs 2017 Feb;26:36-41. [doi: 10.1016/j.ejon.2016.12.002] [Medline: 28069150]
26. Langius-Eklöf A, Christiansen M, Lindström V, Blomberg K, Hälleberg Nyman M, Wengström Y, et al. Adherence to report and patient perception of an interactive app for managing symptoms during radiotherapy for prostate cancer: descriptive study of logged and interview data. JMIR Cancer 2017 Oct 31;3(2):e18 [FREE Full text] [doi: 10.2196/cancer.7599] [Medline: 29089290]
27. Portenoy RK, Thaler HT, Kornblith AB, Lepore JM, Friedlander-Klar H, Kiyasu E, et al. The Memorial Symptom Assessment Scale: an instrument for the evaluation of symptom prevalence, characteristics and distress. Eur J Cancer 1994;30A(9):1326-1336. [Medline: 7999421]
28. Egbring M, Far E, Roos M, Dietrich M, Brauchbar M, Kullak-Ublick GA, et al. A mobile app to stabilize daily functional activity of breast cancer patients in collaboration with the physician: a randomized controlled clinical trial. J Med Internet Res 2016 Sep 06;18(9):e238 [FREE Full text] [doi: 10.2196/jmir.6414] [Medline: 27601354]
29. Beck SL, Eaton LH, Echeverria C, Mooney KH. SymptomCare@Home: developing an integrated symptom monitoring and management system for outpatients receiving chemotherapy. Comput Inform Nurs 2017 Oct;35(10):520-529 [FREE Full text] [doi: 10.1097/CIN.0000000000000364] [Medline: 28570285]
30. Escriva Boulley G, Leroy T, Bernetière C, Paquienseguy F, Desfriches-Doria O, Préau M. Digital health interventions to help living with cancer: a systematic review of participants' engagement and psychosocial effects. Psychooncology 2018 Dec;27(12):2677-2686. [doi: 10.1002/pon.4867] [Medline: 30152074]
31. Hernandez Silva E, Lawler S, Langbecker D. The effectiveness of mHealth for self-management in improving pain, psychological distress, fatigue, and sleep in cancer survivors: a systematic review. J Cancer Surviv 2019 Feb;13(1):97-107. [doi: 10.1007/s11764-018-0730-8] [Medline: 30635865]

32. McCann L, Maguire R, Miller M, Kearney N. Patients' perceptions and experiences of using a mobile phone-based advanced symptom management system (ASyMS) to monitor and manage chemotherapy related toxicity. Eur J Cancer Care (Engl) 2009 Mar;18(2):156-164. [doi: 10.1111/j.1365-2354.2008.00938.x] [Medline: 19267731]
33. Bock M, Moore D, Hwang J, Shumay D, Lawson L, Hamolsky D, et al. The impact of an electronic health questionnaire on symptom management and behavior reporting for breast cancer survivors. Breast Cancer Res Treat 2012 Aug;134(3):1327-1335. [doi: 10.1007/s10549-012-2150-1] [Medline: 22798157]
34. Judson TJ, Bennett AV, Rogak LJ, Sit L, Barz A, Kris MG, et al. Feasibility of long-term patient self-reporting of toxicities from home via the internet during routine chemotherapy. J Clin Oncol 2013 Jul 10;31(20):2580-2585 [FREE Full text] [doi: 10.1200/JCO.2012.47.6804] [Medline: 23733753]
35. Berry DL, Hong F, Halpenny B, Partridge A, Fox E, Fann JR, et al. The electronic self report assessment and intervention for cancer: promoting patient verbal reporting of symptom and quality of life issues in a randomized controlled trial. BMC Cancer 2014;14:513 [FREE Full text] [doi: 10.1186/1471-2407-14-513] [Medline: 25014995]
36. Short C, Rebar A, Plotnikoff R, Vandelanotte C. Designing engaging online behaviour change interventions: a proposed model of user engagement. Health Psychol Rev.: European Health Psychology Society; 2015. URL: https://www.ehps.net/ehp/index.php/contents/issue/view/ehp.v17.i1/pdf_43 [accessed 2020-07-22]
37. Sieverink F, Kelders SM, van GJE. Clarifying the concept of adherence to ehealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
38. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
39. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Long-term outcomes for neoadjuvant versus adjuvant chemotherapy in early breast cancer: meta-analysis of individual patient data from ten randomised trials. Lancet Oncol 2018 Jan;19(1):27-39 [FREE Full text] [doi: 10.1016/S1470-2045(17)30777-5] [Medline: 29242041]
40. Van de Wiel M, Dockx Y, Van den Wyngaert T, Stroobants S, Tjalma WAA, Huizing MT. Neoadjuvant systemic therapy in breast cancer: challenges and uncertainties. Eur J Obstet Gynecol Reprod Biol 2017 Mar;210:144-156. [doi: 10.1016/j.ejogrb.2016.12.014] [Medline: 28039758]
41. Swain SM. Chemotherapy: updates and new perspectives. Oncologist 2011;16 Suppl 1:30-39 [FREE Full text] [doi: 10.1634/theoncologist.2011-S1-30] [Medline: 21278439]
42. Steenbruggen TG, van Ramshorst MS, Kok M, Linn SC, Smorenburg CH, Sonke GS. Neoadjuvant therapy for breast cancer: established concepts and emerging strategies. Drugs 2017 Aug;77(12):1313-1336. [doi: 10.1007/s40265-017-0774-5] [Medline: 28616845]
43. Nationellt vårdprogram bröstcancer. cancercentrum.se. Stockholm, Sweden: Regional cancercentrum i samverkan; 2020 May 21. URL: https://www.cancercentrum.se/samverkan/cancerdiagnoser/brost/vardprogram/ [accessed 2019-01-30]
44. Prostatacancer Nationellt vårdprogram (National Care Program Prostate Cancer). Regionala Cancercentrum i Samverkan. Stockholm, Sweden; 2020 Mar 03. URL: https://www.cancercentrum.se/samverkan/cancerdiagnoser/prostata/vardprogram/ [accessed 2020-04-01]
45. Fjell M, Langius-Eklöf A, Nilsson M, Wengström Y, Sundberg K. Reduced symptom burden with the support of an interactive app during neoadjuvant chemotherapy for breast cancer - A randomized controlled trial. Breast 2020 Jun;51:85-93 [FREE Full text] [doi: 10.1016/j.breast.2020.03.004] [Medline: 32247251]
46. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [Medline: 3558716]
47. Altman D. Practical statistics for medical research. London: Chapman and Hall; 1990:9780412276309.
48. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
49. Kelders SM, Kok RN, Ossebaard HC, Van GJEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
50. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
51. Waibel S, Henao D, Aller M, Vargas I, Vázquez M. What do we know about patients' perceptions of continuity of care? a meta-synthesis of qualitative studies. Int J Qual Health Care 2012 Feb;24(1):39-48. [doi: 10.1093/intqhc/mzr068] [Medline: 22146566]
52. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013;15(3):e34 [FREE Full text] [doi: 10.2196/jmir.2285] [Medline: 23454601]
53. Moradian S, Voelker N, Brown C, Liu G, Howell D. Effectiveness of internet-based interventions in managing chemotherapy-related symptoms in patients with cancer: a systematic literature review. Support Care Cancer 2018 Feb;26(2):361-374. [doi: 10.1007/s00520-017-3900-8] [Medline: 28948360]

54. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
55. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 02;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
56. Ryan C, Bergin M, Wells JS. Theoretical perspectives of adherence to web-based interventions: a scoping review. Int J Behav Med 2017 Jul 20:17-29. [doi: 10.1007/s12529-017-9678-8] [Medline: 28730402]
57. Thórarinsdóttir K, Kristjánsson K. Patients' perspectives on person-centred participation in healthcare: a framework analysis. Nurs Ethics 2014 Mar;21(2):129-147. [doi: 10.1177/0969733013490593] [Medline: 23812560]
58. Florin J, Ehrenberg A, Ehnfors M. Clinical decision-making: predictors of patient participation in nursing care. J Clin Nurs 2008 Nov;17(21):2935-2944. [doi: 10.1111/j.1365-2702.2008.02328.x] [Medline: 19034992]
59. Kolovos P, Kaitelidou D, Lemonidou C, Sachlas A, Sourtzi P. Patients' perceptions and preferences of participation in nursing care. Journal of Research in Nursing 2016 Mar 08;21(4):290-303. [doi: 10.1177/1744987116633498]
60. Lalmas M, O'Brien H, Yom-Tov E. Measuring User Engagement. Synthesis Lectures on Information Concepts, Retrieval, and Services 2014 Nov;6(4):1-132. [doi: 10.2200/s00605ed1v01y201410icr038]
61. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of technology-based strategies to promote engagement with digital interventions: a systematic review protocol. JMIR Res Protoc 2015;4(2):e47 [FREE Full text] [doi: 10.2196/resprot.3990] [Medline: 25921274]
62. Beatty L, Kemp E, Binnion C, Turner J, Milne D, Butow P, et al. Uptake and adherence to an online intervention for cancer-related distress: older age is not a barrier to adherence but may be a barrier to uptake. Support Care Cancer 2017 Jun;25(6):1905-1914. [doi: 10.1007/s00520-017-3591-1] [Medline: 28155018]

Abbreviations

IQR: interquartile range
mHealth: mobile technology for health

Edited by C Gibbons, G Eysenbach; submitted 20.11.19; peer-reviewed by P Innominato, L McCann; comments to author 10.03.20; revised version received 05.05.20; accepted 20.05.20; published 10.08.20.

Please cite as:
Crafoord MT, Fjell M, Sundberg K, Nilsson M, Langius-Eklöf A
Engagement in an Interactive App for Symptom Self-Management during Treatment in Patients With Breast or Prostate Cancer: Mixed Methods Study
J Med Internet Res 2020;22(8):e17058
URL: https://www.jmir.org/2020/8/e17058
doi:10.2196/17058
PMID:32663140

Original Paper

Mobile App for Mental Health Monitoring and Clinical Outreach in Veterans: Mixed Methods Feasibility and Acceptability Study

Lisa M Betthauser[1,2], MBA, PhD; Kelly A Stearns-Yoder[1,2], MA; Suzanne McGarity[1,2], PhD; Victoria Smith[3], BSc; Skyler Place[3], PhD; Lisa A Brenner[1,2,4], PhD

[1]Veterans Affairs, Rocky Mountain Mental Illness Research Education and Clinical Center, Aurora, CO, United States

[2]Department of Physical Medicine and Rehabilitation, School of Medicine, University of Colorado, Aurora, CO, United States

[3]Cogito Corporation, Boston, MA, United States

[4]Departments of Psychiatry and Neurology, School of Medicine, University of Colorado, Aurora, CO, United States

Corresponding Author:
Lisa M Betthauser, MBA, PhD
Veterans Affairs
Rocky Mountain Mental Illness Research Education and Clinical Center
1700 N Wheeling St
Aurora, CO, 80045
United States
Phone: 1 720 723 4867
Email: lisa.betthauser@va.gov

Abstract

Background: Advances in mobile health (mHealth) technology have made it possible for patients and health care providers to monitor and track behavioral health symptoms in real time. Ideally, mHealth apps include both passive and interactive monitoring and demonstrate high levels of patient engagement. Digital phenotyping, the measurement of individual technology usage, provides insight into individual behaviors associated with mental health.

Objective: Researchers at a Veterans Affairs Medical Center and Cogito Corporation sought to explore the feasibility and acceptability of an mHealth app, the Cogito Companion.

Methods: A mixed methodological approach was used to investigate the feasibility and acceptability of the app. Veterans completed clinical interviews and self-report measures, at baseline and at a 3-month follow-up. During the data collection period, participants were provided access to the Cogito Companion smartphone app. The mobile app gathered passive and active behavioral health indicators. Data collected (eg, vocal features and digital phenotyping of everyday social signals) are analyzed in real time. Passive data collected include location via global positioning system (GPS), phone calls, and SMS text message metadata. Four primary model scores were identified as being predictive of the presence or absence of depression or posttraumatic stress disorder (PTSD). Veterans Affairs clinicians monitored a provider dashboard and conducted clinical outreach when indicated.

Results: Findings suggest that use of the Cogito Companion app was feasible and acceptable. Veterans (n=83) were interested in and used the app; however, active use declined over time. Nonetheless, data were passively collected, and outreach occurred throughout the study period. On the Client Satisfaction Questionnaire–8, 79% (53/67) of the sample reported scores demonstrating acceptability of the app (mean 26.2, SD 4.3). Many veterans reported liking specific app features (day-to-day monitoring) and the sense of connection they felt with the study clinicians who conducted outreach. Only a small percentage (4/67, 6%) reported concerns regarding personal privacy.

Conclusions: Feasibility and acceptability of the Cogito Corporation platform to monitor mental health symptoms, behaviors, and facilitate follow-up in a sample of veterans were supported. Clinically, platforms such as the Cogito Companion system may serve as useful methods to promote monitoring, thereby facilitating early identification of risk and mitigating negative psychiatric outcomes, such as suicide.

(J Med Internet Res 2020;22(8):e15506) doi:10.2196/15506

KEYWORDS

veterans; mobile app; smartphone; mental health; acceptability; feasibility

Introduction

Mental health disorders are common, with 46.6 million United States (US) adults using mental health services in 2017 [1]. Approximately 2.1 million US individuals received Veterans Affairs mental health care between 2006 and 2010 [2]. Although not all veterans seek care at Veterans Health Administration facilities [3], one recent study [4], using data from the 2012-2014 National Survey on Drug Use and Health, found that 34% of veterans with any mental illness had received at least one form of mental health treatment in the preceding 12 months. Veterans commonly experience symptoms, such as those associated with posttraumatic stress, mood, and anxiety disorders. In the past, these symptoms have been monitored using self-report measures, which can yield unreliable and incomplete information [5].

Advances in smartphone technology, such as mobile health (mHealth) apps, provide the opportunity for continuous monitoring and tracking of mental health symptoms and behaviors [6-8]. The majority of individuals living in the US own and use smartphones [9], which increases the feasibility of using this technology to detect behavior change; mHealth apps which employ methods such as digital phenotyping can assist patients in adhering to treatment via self-monitoring, as well as assist health care providers in monitoring symptom progression outside of clinical visits. Ideally, such monitoring would occur in real time [7], and include subjective self-report of mental health symptoms, as well as objective behavioral data drawn from passive monitoring of an individual's smartphone. Such behavioral data may be used to observe and track changes in functioning (eg, changes in social connectivity as indicated by text messaging patterns). Clinical outreach can then be initiated. Although the Department of Defense and Veterans Affairs have created and distributed mHealth apps for service members and veterans [10], these apps predominantly provide psychoeducation or facilitate coping strategies, but do not allow patients and health care providers to interactively track symptoms and behaviors in real time.

Cogito Corporation, in conjunction with the Defense Advance Research Project Agency of the Department of Defense, developed a secure, federally compliant mHealth smartphone app, the Cogito Companion, to collect and analyze passive data from the personal smartphones of participants. Data collected (eg, vocal features and digital phenotyping of everyday social signals) are analyzed in real time. Passive data collected include location via global positioning system (GPS) information, phone calls, and SMS text message metadata (including counts and frequency [11]). These data were transformed in Cogito's proprietary algorithms to provide scores of behavior health indicators that can be shared directly with patients in the mobile app and with providers in real time via a clinical dashboard [11].

Initial data sets (civilians and veterans experiencing at least one symptom of depression or posttraumatic stress disorder) were used to create Cogito's proprietary algorithms. Four primary model scores were identified as being predictive of the presence or absence of depression or posttraumatic stress disorder based on DSM-IV symptom criteria and included *mood* (indicative of a depressed mood almost every day), *out-and-about* (indicative of diminished interest or pleasure in activities), *socially connected* (indicative of avoidance of activities, places, and people), and *energized* (indicative of fatigue or loss of energy). Model scores ranged from 0 to 100, where a score of 100 indicated a low probability of the presence of the symptom, and 0 indicated a high probability of the presence of the symptom. While initial validity of the model scores and of the feasibility of implementing of the Cogito Companion system were demonstrated, further investigation of acceptability of this app to facilitate symptom monitoring in veterans was suggested [11].

The objective of this study was to evaluate whether veterans would be willing to download and use the app thereby allowing clinicians to monitor output from the app and provide outreach, if indicated. Such acceptability and feasibility data would yield important information for future implementation or efficacy studies in this population.

Methods

Study Design

A mixed methodological approach was used to test the feasibility and acceptability of the Cogito Companion system. Acceptability was defined as the perceived suitability of the intervention [12] and was measured by completion of a satisfaction questionnaire and with data from structured qualitative interviews. Acceptability was determined using a criterion defined a priori—more than 70% of the sample providing a score of 24 or higher on the Client Satisfaction Questionnaire–8 and responses to structure qualitative interviews. Feasibility [12] was defined as the ease of implementation of the app among veterans. Feasibility was descriptively measured based on ease of recruitment, rate of enrollment, and rate of attrition; and frequency of clinical outreach calls based on Cogito Companion model scores and responses to a self-reported suicide risk item on biweekly mental health surveys. Veterans' use of the app was measured by in-app usage metrics and completion of a post-app user experience questionnaire.

Participants

Participants were US military veterans receiving or eligible to receive care at a mountain state Veterans Affairs Medical Center. Veterans were included if they were between the ages of 18 and 89, had an Android-based smartphone, were willing to use their personal data plan during the study, and were able to provide informed consent. Participants were excluded if they were participating in a conflicting interventional study at the local Veterans Affairs Medical Center. Participants were compensated for the baseline and follow-up visits and were provided monthly payments to offset smartphone data costs.

Procedures

Colorado Multiple Institutional Review Board approval was obtained. Data collection occurred between October 2016 and September 2018. During the baseline assessment, veterans completed clinical interviews and self-report questionnaires.

Study personnel also provided instructions on how to download the Cogito Companion app and answered any questions that arose during the download process. Participants were provided with an app user guide and were informed that the app would collect behavioral information from their smartphone to evaluate their mood, level of stress, and general well-being. They were told they could interact with the app by monitoring their own scores and by leaving audio recordings. App interaction also occurred via biweekly mental health surveys. A notification was displayed by the app when surveys became available. Resources regarding mental health, well-being, and safety were also embedded within the app. After 3 months, participants were contacted to complete an in-person or over-the-phone follow-up assessment and were instructed to uninstall the app. Follow-up assessments included clinical interviews and self-report questionnaires. Data from full measures are available upon request, as not all data is presented in this paper.

Measures and Materials

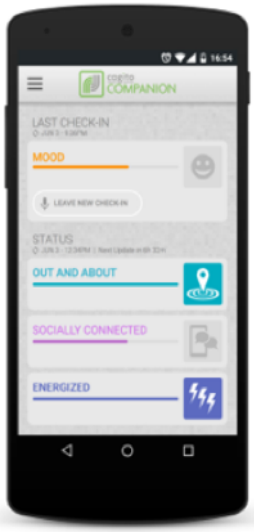
Clinical Interviews and Assessments

The Structured Clinical Interview for DSM-5 Disorders-Research Version [13] was used to assess current and lifetime psychiatric diagnoses. The Ohio State University Traumatic Brain Injury Identification Method [14], a valid and reliable structured clinical interview procedure, was used to obtain lifetime history of traumatic brain injury [14,15]. The Narrative Evaluation of Intervention Interview [16], a structured qualitative interview, was used to obtain feedback regarding the intervention. The Patient Health Questionnaire–9 [17], a 9-item self-report module from the Primary Care Evaluation of Mental Disorders diagnostic instrument, was used every 2 weeks to assess and monitor symptoms of depression: item 9 (ie, thoughts of being better off dead) was used to monitor ongoing suicide risk. The Client Satisfaction Questionnaire–8 [18], an 8-item questionnaire, was used to assess patient satisfaction with the intervention. Cogito's Technology Assessment Survey was used to evaluate how likely participants would be to share data from this app with others, overall privacy perceptions, and the degree to which behaviors may have changed as a result of app use.

Cogito Companion App

Once downloaded, the Cogito Companion app (see Figure 1) was able to continuously collect passive data [11]. Model scores generated by the passive data were displayed daily after the first 7 days postdownload. Model scores for audio check-ins were available immediately after the recording was completed. These scores were also presented in a clinician dashboard for the study team to monitor. Clinical outreach occurred based on 3 primary model scores—socially connected, out-and-about, and energized—or if a response to biweekly Patient Health Questionnaire–9 item 9 (ie, thoughts of being better off dead) was a score of 2 or greater. Clinical outreach facilitated treatment with mental health providers within the Veterans Affairs Medical Center, as needed, if the veteran appeared to be at increased risk for suicide or struggling with risk factors associated with suicide risk. The audio check-in score was not used for clinical monitoring purposes.

Figure 1. Cogito Companion app.



Data Analysis

Quantitative Analysis

Sample characteristics and descriptive analyses were performed using SAS software (version 9.4; SAS Institute). Data are presented as counts, percentages, median, and range.

Qualitative Analysis

Qualitative interviews were analyzed for all participants who completed the 3-month follow-up visit. Investigators used a thematic analysis method [19] rooted in the qualitative descriptive approach [20]. Thematic analysis included six steps: (1) familiarization with the data by reading transcripts, (2) generation of initial codes, (3) identification of themes, (4) review of themes, (5) refinement of the themes, and (6) selection

of quotes and examples. We ensured that data saturation [21] occurred, that is, that no new themes or new experiences with the app were noted. Furthermore, data adequacy [22] was achieved; themes that emerged were comprehensive and little to no variation was observed across interviews. All salient themes are presented below.

Scientific rigor was facilitated through a step-by-step approach to examining the interviews. Three investigators (LMB, KSY, SM), who were not affiliated with Cogito Companion, noted and shared their biases at the start of their coding process. At least two investigators reviewed each interview, noting themes, impressions and descriptive quotes on previously constructed coding sheets based on the structure of the qualitative interview (ie, description and evaluation of the app) [16]. Investigators met to review, document, and achieve consensus regarding common themes and identify illustrative quotes. Descriptive summaries of commonly observed themes were generated [20,23]. Themes of benefits, motivation to use the app, barriers experienced using the app, and general use of the app were noted.

Results

Baseline Sample Characteristics

Sample characteristics for the participants (N=83) are displayed in Table 1. Individuals in the sample were predominantly men (72/83, 86.7%), white (52/83, 62.7%), and had served in the Army (46/83, 55.4%).

Table 1. Baseline characteristics of the sample.

Participant characteristics (N=83)	Value
Age (years), median (range[a])	50 (24-76)
Gender, n (%)	
Male	72 (87)
Female	11 (13)
Race, n (%)	
White	52 (63)
Black or African American	16 (19)
Multiracial	5 (6)
Other	10 (12)
Military branch, n (%)	
Army	46 (55)
Air Force	12 (15)
Navy	8 (10)
Marines	16 (19)
Multiple branches	1 (1)
Rank at separation, n (%)	
Enlisted	57 (69)
Noncommissioned officer	20 (24)
Officer	6 (7)
Deployed, n (%)	
No	30 (36)
Yes	53 (64)
Combat experience, n (%)	
No	42 (51)
Yes	41 (49)
Marital status, n (%)	
Married	29 (35)
Single	20 (24)
Cohabitating	4 (5)
Widowed	4 (5)
Divorced or separated	26 (31)
Employment status, n (%)	
Full-time	23 (28)
Part-time	7 (8)
Unemployed	30 (36)
Retired	22 (27)
Missing	1 (1)
Psychological disorders[b], n (%)	
Alcohol abuse	12 (15)
Major depressive disorder	14 (17)
Posttraumatic stress disorder	20 (25)

Participant characteristics (N=83)	Value
Substance abuse	9 (11)
Missing[c]/none	28 (34)
History of TBI, n (%)	
Mild only	48 (58)
Moderate to severe only	4 (5)
Mild and moderate to severe	9 (11)
None	22 (26)
Missing	1 (1)

[a]Minimum to maximum.

[b]N=82.

[c]Data for 1 individual was missing.

Feasibility

A flowchart of study eligibility, enrollment, and completion is presented in Figure 2. Although many veterans expressed an interest in the study, about one-third (102/300, 34%) were ineligible because of smartphone or data plan related issues (eg, iPhone versus Android-based smartphone). Scheduling issues and difficulty contacting veterans had an impact on enrollment. Approximately 18% (36/198) of veterans who were eligible were not interested in participating; reasons for this were not collected. No feasibility issues regarding the app download process were observed. Veterans who participated in the 3-month follow-up (67/83, 81%) were asked to subjectively rate the frequency of their app use. These data are shown in Table 2. App interaction data over the course of the study (approximately 90 days) are shown in Figure 3.

Figure 2. Recruitment flow diagram.

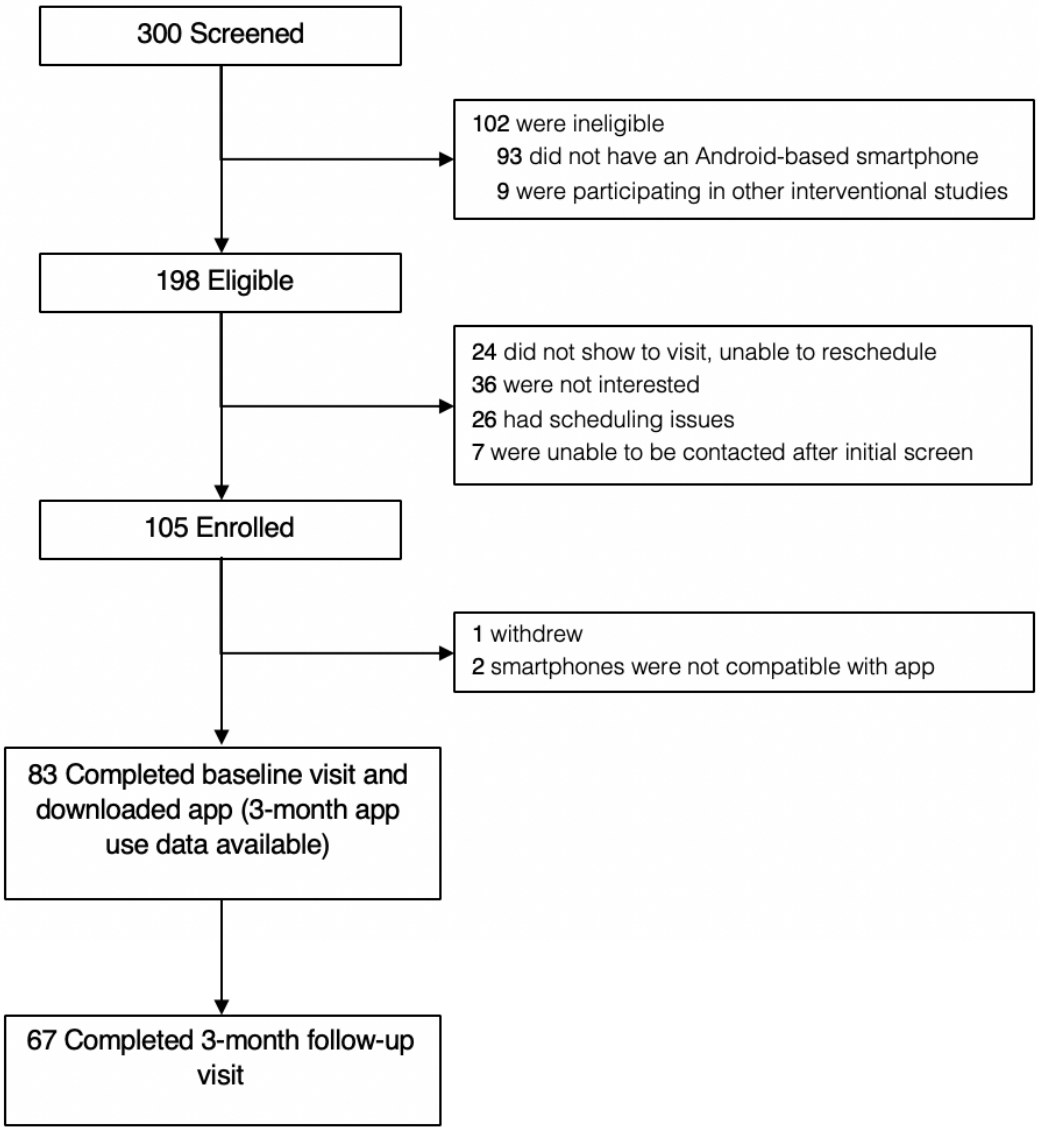


Table 2. Cogito Companion app use over the 3-month study period.

App action	Number of participants (N=83), n (%)	Number of times, median (range[a])
Clicked on the home screen	83 (100)	24 (1-191)
Clicked on the help screen	62 (75)	1 (1-9)
Clicked on a daily model score	81 (98)	9 (1-49)
Clicked on a survey	74 (89)	4 (1-41)
Clicked on an audio recording	74 (89)	6.5 (1-83)
First biweekly PHQ-9[b] survey completed	54 (65)	N/A[c]
Last biweekly PHQ-9 survey completed	39 (47)	N/A
How often did you use the app?[d]		
Never	1 (2)	N/A
Occasionally	17 (25)	N/A
Moderately	24 (36)	N/A
Frequently/often	25 (37)	N/A

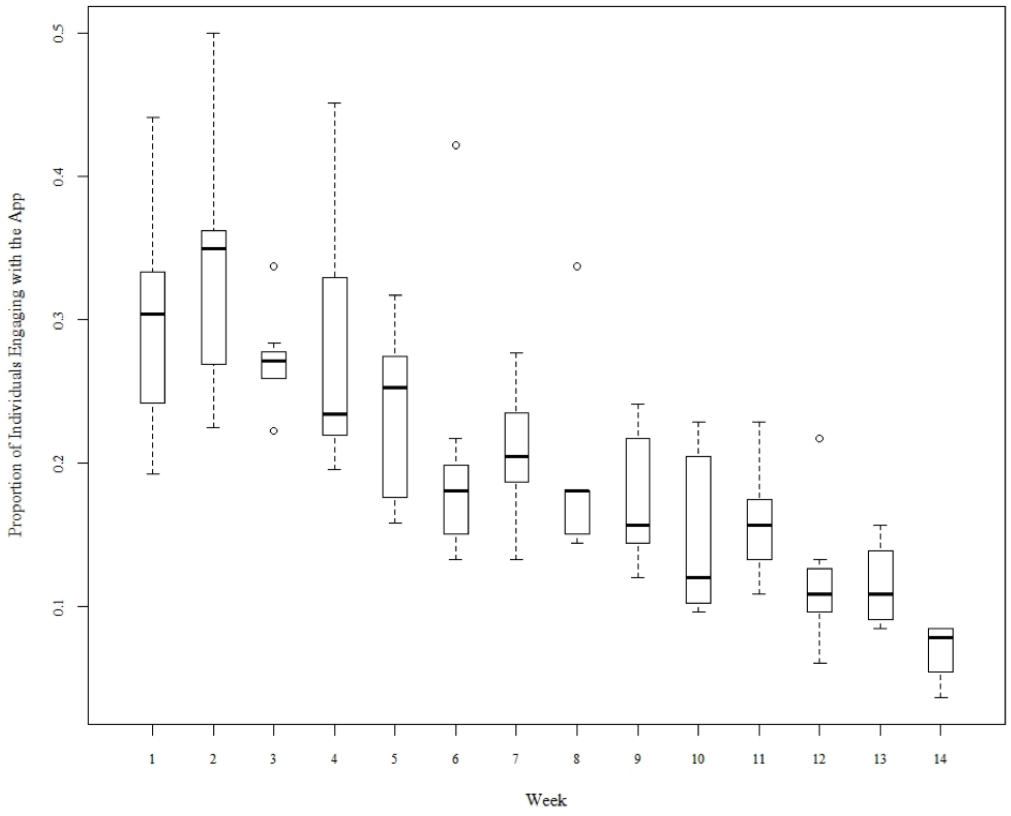
[a]Minimum to maximum.

[b]PHQ-9: Patient Health Questionnaire–9.

[c]N/A: not applicable.

[d]Question was administered at follow-up approximately 3 months postdownload. N=67.

Figure 3. Cogito Companion app engagement. Note: Active app use data is incomplete for the first 7 participants. All available data is displayed.



A potential limitation of app use was participant misunderstanding about which smartphone behaviors contributed to their model scores, though it is unclear whether this had an impact on app use. A veteran commented:

> *...the out-and-about and socially connected, I wasn't sure about those two...I wasn't comfortable with [the scores] only because I didn't know.*

Some participants noted that they did not know how frequently they should be answering surveys, leaving audio check-ins, and interacting with the app. Many participants recommended having additional notifications to complete surveys, complete check-ins, and for in-app feedback. Furthermore, some veterans believed scores could be intentionally influenced. With respect to the mood score derived in part from the audio check-in, one veteran stated:

> *...so if I try to sound like I'm in a good mood, like with a smile on my face...it can be manipulated.*

Technology-related issues emerged as additional barriers to using the app. For example, challenges with wireless internet connectivity disrupted functionality (eg, delayed data transfer resulted in model scores not being calculated or updated). Some participants noted that the app was "buggy," and scores were "erratic" which may have been due to limited connectivity, lack of SMS text messaging, or disabled GPS functions. Nonclinical study team members contacted 40 veterans to address these technological issues, with 40% (16/40) reporting resolution with one outreach. A system upgrade that disabled survey and audio check-in notifications temporarily had an impact on the participation of 15% (6/40) veterans.

Clinical team members contacted 42% (35/83) veterans over the course of the study either because of their item-9 response on the Patient Health Questionnaire–9 or because of an observable change in the model scores on the clinical dashboard. Approximately 26% (9/35) of the veterans who were contacted by clinical team members had endorsed suicidal ideation on the Patient Health Questionnaire–9. Clinical team members conducted risk assessments, reviewed safety plans, and facilitated outreach to the veteran's primary mental health care provider. Approximately 30% (10/35) of the clinical contacts were due to observable changes in the model scores; 17% (6/35) reported experiencing increase in posttraumatic stress disorder symptoms or reported psychosocial stressors such as difficulty finding employment or difficulties in relationships, approximately 25% (9/35) veterans reported feeling "fine" and "great," and the remainder (5/35, 14%) reported physical health issues (eg, colds, postoperative recovery status, or pain) impacting changes in their activities and use of their phone. The remaining clinical contacts (15/35, 43%) were because of missing scores. In such cases, participants required further operational support such as password resets, or instructions to turn on notifications or allowances specific to the app (eg, GPS).

Acceptability

Results from the Client Satisfaction Questionnaire–8 suggest that 79% (53/67) of the participants found the app intervention acceptable, a score of 24 or above (mean 26.2, SD 4.3). Data from the Cogito Technology Assessment Survey suggest that about half of veterans (34/67, 51%) enjoyed using the app a lot, 45% (30/67) enjoyed using the app a little, and 4% (3/67) did not enjoy using the app at all. Many described the convenience and accessibility of using the app, noting:

> *I could just do it at my own leisure.*
>
> *...it was easily accessible...*
>
> *It was on your phone, you carry your phone everywhere, that was the biggest thing.*

More than half (35/67, 52%) of the veterans reported liking the audio check-in feature; 34% (23/67) reported liking the feature of reviewing past scores on the Cogito Technology Assessment Survey. Veterans identified these features as useful "tool[s]" for monitoring daily activity and stress and building self-awareness. A veteran stated:

> *It made me monitor my stress level, my attitude...monitored my stress and my everyday activities and everything that I do.*

Many participants identified the audio check-in feature of the app as the primary way they monitored their day. One participant said:

> *It made me conscientious about how I'm feeling.*

Several veterans referred to the audio check-in as a "diary," with one veteran stating:

> *Describing my day, I guess it gave me maybe a rehash of the day and I was able to kind of work through the bad days and stuff, and even on good days I'd even say "hey, it was a good day!" Cool you can have good days.*

Veterans commented that the audio check-in feature allowed them to "vent" about their day, without concern of others listening to their expressed thoughts.

Approximately 33% (22/67) of the participant liked the out-and-about app feature. One veteran used these scores to help create change:

> *...location, moving around and like how physical you are. Coming into contact with others…I was a person that didn't want to be bothered with nobody...I liked to isolate myself…but you know, now I see that I'm more out-and-about, I'm in contact with lots of other people, I'm in this arena football league.*

Fewer veterans endorsed liking the socially connected feature (18/67, 27%). A small subset discussed how this model score increased awareness of a lack of social connection. One veteran stated:

> *[The app] makes you think if you haven't gone out with your friends or socialized in a week or two just makes you think "hey, it's been awhile, maybe I should go out and do something.*"

Other veterans discussed the sense of connection and caring they felt with study clinicians who monitored the clinician dashboard. One participant who was the recipient of a clinical outreach during the study expressed:

> *In other stuff…anytime I reached out before, it was like I was reaching out to a robot... And then, I am using this app instead and I am answering questions and I started getting phone calls and getting help. After I would answer the survey questions from the Companion app and someone would call me to check in on me. At first, I thought that was odd but then I thought that was kinda cool. I actually liked that.*

Slightly less than five percent (3/67, 4%) of the veterans reported not liking the app. Moreover, qualitative data suggest that a small subset of participants felt neutral about the app. Many of the participants who expressed neutrality noted its potential utility for different populations, such as those struggling with more severe symptoms of posttraumatic stress disorder, depression, or those at increased risk for suicide. Furthermore, some participants noted the belief that individuals who are more "tech savvy" may reap great benefit.

Very few participants felt that the app violated their personal privacy (4/67, 6%). Nonetheless, qualitatively, a small number of participants shared privacy and monitoring concerns. One veteran stated:

> *I feel like I'm being monitored, my emotions or how I feel or what I'm doing...when I'm out and about you know like it follow me. I feel like I'm being spied on, but I know I'm not.*

A small group of veterans believed that the study team was listening to the recordings, as one veteran noted:

> *I felt that I had to be careful about what I said or how I said it.*

This hesitancy may have had an effect on the frequency and types of messages they recorded. This theme tended to emerge for participants who either expressed a lack of understanding of the app or the technology embedded within the app at the 3-month follow-up visit, or who shared larger concerns regarding privacy and the government on whole

Discussion

Principal Findings

The feasibility of implementing the Cogito Companion app and system was evaluated via willingness of veterans to download and use the app; thereby facilitating tracking and outreach by study clinicians. Veterans were willing to enroll in this study, download the app, and keep the app on their phone for the 3-month study period. Participants varied in terms of their interaction with app features. That is, some participants were highly interactive with the app, checking their own model scores, leaving audio check-ins, and completing biweekly surveys. Other participants were not as interactive, leaving audio check-ins or completing surveys only. Another subset had minimal interaction with the app. Overall, we found that the highest active use of the app by veterans occurred in the first 2 weeks, with a decline in interaction over the course of the study. This finding is comparable to other studies that have found that mHealth app use declines over time [24,25]. This pattern has also been seen in mental health interventions, where treatment engagement declines over time [26,27]. Thus, continued investigation into maintaining patient engagement should be considered. Future studies will benefit from examining the integration of gamification techniques or leveraging intrinsic motivation principles to increase the duration of engagement with the app while maintaining successful clinical outcomes.

As noted above, active engagement declined over time; however, clinical tracking of the passive data and biweekly surveys were the unique and important features of this app. Feasibility of daily clinical monitoring was supported. Clinical outreach for approximately 40% (35/83) of the total sample occurred. Of these clinical contacts, one-third (10/35, 29%) were contacted due to changes in the passive data model scores. Frequently, the passive data signaled changes in physical health, psychological functioning or psychosocial stressors (eg, job, relationships, etc). The identification of possible suicide risk via biweekly Patient Health Questionnaire–9 survey responses was another primary reason for clinical contacts.

Acceptability was determined by veteran self-report and qualitative interview at the end of the 3-month study period. Veterans in this study reported overall satisfaction with the Cogito Companion app, demonstrating acceptability. That is, the scores on the Client Satisfaction Questionnaire–8 were above the cutoff indicative of acceptability of the app. At the end of the study about half of the sample (34/67, 51%) reported enjoying the app a lot. Qualitative interviews yielded specific feedback regarding features of the app and level of interaction with the app. Some veterans found completing self-assessments and monitoring their day-to-day symptoms and overall mood to be useful. This is consistent with prior literature that found that patients receiving mental health services were open to monitoring their health via their smartphones [28,29]. Such monitoring may provide an innovative way to improve access to care or to facilitate measurement-based care [30]. Veterans reported finding other app features such as the out-and-about and socially connected scores to be less useful. Research examining which subset of veterans liked or disliked these model scores is needed.

Qualitative feedback provided by the veterans did not indicate a need to significantly modify app features. On the other hand, veterans endorsed wanting a clearer understanding of how the app model scores were obtained. Future studies will benefit from integrating real-time feedback and helpful definitions within the app to properly orient users across the entire duration of app use. Although the study team expected some effect on app use as a result of concerns about privacy, a relatively small subset of the sample expressed these concerns; thereby suggesting that concerns about privacy may not be a limitation for future efficacy trials.

A subset of veterans reported liking a sense of connection to providers conducting outreach calls. Many mHealth apps have primarily passive functions, such as providing psychoeducation and self-management tools and coping strategies. The Cogito Companion app extends these functions by including daily clinical tracking of symptoms and behaviors with the potential for subsequent clinical outreach and intervention, as indicated. Although further evidence is needed, integration of the app and accompanying clinical dashboard into mental health clinics and

organizations may be an acceptable way for health care providers to facilitate consistent and caring clinical contacts with their patients.

Even veterans who reported a limited benefit from participation with the app acknowledged the impact such technology may have for at-risk populations. Whereas this study had an inclusive recruitment approach, future studies may wish to study the acceptability, feasibility, and efficacy of mHealth apps such as the Cogito Companion in specific populations. High-risk, chronically health-challenged, or rural-dwelling individuals may derive benefits from such apps. Future investigations of mHealth apps will benefit from clinical trials to establish rigorous and validated mHealth apps among diverse populations.

Limitations

A few feasibility limitations that were noted were related to technological matters. At the time of this study, the Cogito Companion system was compatible with Android-based smartphones only, which did not permit enrollment of iPhone users. Although it is not expected that differences would be observed between Android-based smartphone and iPhone users, future research may want to ensure that the app is compatible with all smartphone platform types ensuring generalizability. Other limitations included disruption of use due to app upgrades and interruptions in internet connectivity. While technological limitations are expected for any mHealth app, future investigations should account for these issues and take steps to minimize disruptions. Another technological limitation may be the app's reliance on participant use of text messaging as a necessary variable to populate specific model scores (ie, socially connected and energized). It could be that some veterans do not use SMS text messaging to communicate with family and friends. In fact, a survey on mobile messaging and social media found that 36% of smartphone users communicate via messaging apps separate from SMS text messaging [31]. This same survey report showed that these types of messaging apps are most popular among those aged 18 to 29 years. If veterans in this study used other social media modalities to communicate (such as FaceBook Messenger, SnapChat, WhatsApp, etc), data would not be collected or recorded on the Cogito Companion app nor observed in the clinical dashboard. As a result, passive model score tracking for both the veteran and the clinician may have been less useful and less reflective of daily functioning.

Conclusions

Feasibility and acceptability of a smartphone app to monitor mental health symptoms, behaviors, and follow-up were generally supported. Veterans reported that the app was generally convenient and easy to use. Benefits of use included an improved sense of self-awareness, motivation to increase activity and engagement, and enhanced sense of connection with care providers. Technological improvements and a reduction of technological barriers are needed to ensure continued use. The use of mHealth apps such as the Cogito Companion app may provide an accessible and portable way for patients to track symptoms or behaviors and share changes with providers.

Acknowledgments

We would like to thank Channah Rubin, previously employed by Cogito Corporation, and the research assistants at the Rocky Mountain Mental Illness Research Center for their assistance to the success of this project. This project was supported by the Veterans Health Administration Office of Mental Health and Suicide Prevention and the Rocky Mountain Mental Illness Research Center at the Department of Veterans Affairs. The views, opinions, and findings contained in this paper are those of the authors and should not be construed as an official Department of Veterans Affairs position, policy, or decision unless so designated by other documentation.

Conflicts of Interest

LAB, LMB, KAS-Y, SP, and VS are currently involved in the Military Suicide Research Consortium grant W81XWH-16-2-004 Facilitating Assessment of At-Risk Sailors with Technology (FAAST) based on pilot work published in this article. SP and VS declare that they are employees of Cogito Corporation.

References

1. Mental Health Information. National Institute of Mental Health. 2019 Feb. URL: https://www.nimh.nih.gov/health/statistics/mental-illness.shtml [accessed 2020-07-08]
2. VA Mental Health: number of veterans receiving care, barriers faced, and efforts to increase access. Report to the Ranking Member, Committee on Veterans' Affairs, House of Representatives. Washington, D.C: United States Government Accountability Office; 2011 Oct. URL: https://www.gao.gov/new.items/d1212.pdf [accessed 2020-07-08]
3. Institute Of Medicine. Returning home from Iraq and Afghanistan: assessment of readjustment needs of veterans, service members, and their families. In: Returning Home from Iraq and Afghanistan. Washington, D.C: The National Academies Press; 2013:A.
4. Pietrzak RH, Johnson DC, Goldstein MB, Malley JC, Southwick SM. Perceived stigma and barriers to mental health care utilization among OEF-OIF veterans. Psychiatric Services 2009 Aug;60(8):1118-1122. [doi: 10.1176/ps.2009.60.8.1118] [Medline: 19648201]
5. Bush NE, Ouellette G, Kinn J. Utility of the T2 Mood Tracker mobile application among army warrior transition unit service members. Mil Med 2014 Dec;179(12):1453-1457. [doi: 10.7205/MILMED-D-14-00271] [Medline: 25469967]

6. Areàn PA, Hoa Ly K, Andersson G. Mobile technology for mental health assessment. Dialogues Clin Neurosci 2016 Dec;18(2):163-169 [FREE Full text] [Medline: 27489456]
7. Torous J, Powell AC. Current research and trends in the use of smartphone applications for mood disorders. Internet Interventions 2015 May;2(2):169-173. [doi: 10.1016/j.invent.2015.03.002]
8. Erbes CR, Stinson R, Kuhn E, Polusny M, Urban J, Hoffman J, et al. Access, utilization, and interest in mHealth applications among veterans receiving outpatient care for PTSD. Mil Med 2014 Nov;179(11):1218-1222. [doi: 10.7205/MILMED-D-14-00014] [Medline: 25373044]
9. Mobile Fact Sheet. Pew Research Center Internet and Technology. Washington, D.C: Pew Research Center; 2019 Jun 12. URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2020-07-08]
10. VA Mobile Apps. VA.: Veterans Affairs; 2020 Jun. URL: https://www.myhealth.va.gov/mhv-portal-web/mobile-apps [accessed 2020-07-08]
11. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. J Med Internet Res 2017 Mar 16;19(3):e75 [FREE Full text] [doi: 10.2196/jmir.6678] [Medline: 28302595]
12. Feeley N, Cossette S, Côté J, Héon M, Stremler R, Martorella G, et al. The importance of piloting an RCT intervention. Can J Nurs Res 2009 Jun;41(2):85-99. [Medline: 19650515]
13. First MB, Williams JBW, Karg RS, Spitzer RSCIFDSM5V(5FDSM5, Research VSCID5R. Structured Clinical Interview for DSM-5 - Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV). Arlington, VA: American Psychiatric Association; 2015.
14. Corrigan JD, Bogner J. Initial reliability and validity of the Ohio State University TBI Identification Method. J Head Trauma Rehabil 2007;22(6):318-329. [doi: 10.1097/01.HTR.0000300227.67748.77] [Medline: 18025964]
15. Bogner J, Corrigan JD. Reliability and predictive validity of the Ohio State University TBI identification method with prisoners. J Head Trauma Rehabil 2009;24(4):279-291. [doi: 10.1097/HTR.0b013e3181a66356] [Medline: 19625867]
16. Hasson-Ohayon I, Roe D, Kravetz S. A qualitative approach to the evaluation of psychosocial interventions for persons with severe mental illness. Psychological Services 2006 Nov;3(4):262-273. [doi: 10.1037/1541-1559.3.4.262]
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
18. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982;5(3):233-237. [Medline: 10259963]
19. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
20. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [Medline: 10940958]
21. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822x05279903]
22. Morrow SL. Quality and trustworthiness in qualitative research in counseling psychology. Journal of Counseling Psychology 2005 Apr;52(2):250-260. [doi: 10.1037/0022-0167.52.2.250]
23. Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description - the poor cousin of health research? BMC Med Res Methodol 2009 Jul 16;9(1):52 [FREE Full text] [doi: 10.1186/1471-2288-9-52] [Medline: 19607668]
24. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]
25. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
26. Swift JK, Greenberg RP. Premature discontinuation in adult psychotherapy: a meta-analysis. J Consult Clin Psychol 2012 Aug;80(4):547-559. [doi: 10.1037/a0028226] [Medline: 22506792]
27. Cully JA, Tolpin L, Henderson L, Jimenez D, Kunik ME, Petersen LA. Psychotherapy in the veterans health administration: missed opportunities? Psychol Serv 2008 Nov;5(4):320-331 [FREE Full text] [doi: 10.1037/a0013719] [Medline: 25177213]
28. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014;2(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2994] [Medline: 25098314]
29. Proudfoot J, Parker G, Hadzi PD, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
30. FY 2018-2024 Strategic Plan. U.S. Department of Veterans Affairs. Washington, D.C: U.S. Department of Veterans Affairs; 2019 May 31. URL: https://www.va.gov/oei/docs/VA2018-2024strategicPlan.pdf [accessed 2020-07-08]
31. Mobile messaging and social media 2015. Pew Research Center, Internet and Technology. Washington, D.C: Pew Research Center; 2015 Aug 19. URL: https://www.pewresearch.org/internet/2015/08/19/mobile-messaging-and-social-media-2015/ [accessed 2020-07-06]

Abbreviations

DSM: Diagnostic and Statistical Manual of Mental Disorders
GPS: global positioning system
mHealth: mobile health

Edited by G Eysenbach; submitted 15.07.19; peer-reviewed by B Kohrt, A Poudyal, A Pratap, D Leightley; comments to author 21.10.19; revised version received 30.04.20; accepted 13.05.20; published 11.08.20.

Please cite as:
Betthauser LM, Stearns-Yoder KA, McGarity S, Smith V, Place S, Brenner LA
Mobile App for Mental Health Monitoring and Clinical Outreach in Veterans: Mixed Methods Feasibility and Acceptability Study
J Med Internet Res 2020;22(8):e15506
URL: https://www.jmir.org/2020/8/e15506
doi:10.2196/15506
PMID:32779572

Original Paper

Mobile Health App for Japanese Adult Patients With Asthma: Clinical Observational Study

Norihiro Harada[1,2], MD, PhD; Sonoko Harada[1,3], PhD; Jun Ito[1], MD, PhD; Ryo Atsuta[1], MD, PhD; Satoshi Hori[4], MD, PhD; Kazuhisa Takahashi[1], MD, PhD

[1]Department of Respiratory Medicine, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan

[2]Research Institute for Diseases of Old Ages, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan

[3]Atopy (Allergy) Research Center, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan

[4]Department of Electric Medical Intelligence Management, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan

Corresponding Author:
Norihiro Harada, MD, PhD
Department of Respiratory Medicine
Juntendo University Faculty of Medicine and Graduate School of Medicine
3-1-3 Hongo
Bunkyo-ku
Tokyo, 113-8431
Japan
Phone: 81 3 5802 1063
Fax: 81 3 5802 1617
Email: nor@juntendo.ac.jp

Abstract

Background: Inappropriate asthma control reduces quality of life and causes increased exacerbations. Mobile health (mHealth) employs information and communication technology for surveying health-related issues.

Objective: This noninterventional, observational study assessed current real-world asthma control levels among Japanese patients with asthma and cough variant asthma (CVA) using the Zensoku-Log app.

Methods: We developed the app using the ResearchKit platform and conducted a mobile-based, self-reporting, observational survey among patients with asthma and CVA. The app was downloaded 7855 times between February 2016 and February 2018, and enabled collection of data on symptoms, comorbidities, quality of life, medications, asthma control, and adherence.

Results: Of the 1744 eligible participants (median age 33 years; range 20-74 years; male-to-female ratio 38.7:61.3), 50.97% (889/1744) reported unscheduled visits, 62.84% (1096/1744) reported regularly scheduled visits, 23.14% (402/1737) smoked, and 40.75% (705/1730) had pets. In addition, 91.89% (1598/1739) of participants had atopic predisposition, including allergic rhinitis and atopic dermatitis. Daily inhaled corticosteroid and oral corticosteroid treatment had been prescribed for 89.45% (1552/1735) and 22.07% (383/1735) of participants, respectively. Although an asthma control questionnaire demonstrated poor asthma control in 58.48% (1010/1727), a leukotriene receptor antagonist, theophylline, and a long-acting muscarinic antagonist had been prescribed for only 30.66% (532/1735), 15.91% (276/1735), and 4.38% (76/1735), respectively. The Adherence Starts with Knowledge 12 total score was 29. In the 421 participants who repeated the questionnaire, asthma control increased significantly between the initial and last rounds (P=.002).

Conclusions: Users of this mHealth app in Japan had poorly controlled asthma and may need more treatment for asthma and their comorbidities. Repeated app users demonstrated improved asthma control.

Trial Registration: UMIN Clinical Trial Registry UMIN000021043; https://upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000023913.

(J Med Internet Res 2020;22(8):e19006) doi:10.2196/19006

KEYWORDS

asthma; cough variant asthma; mobile health; ResearchKit

Introduction

Asthma

Asthma is one of the most common chronic diseases that poses a serious global health problem; it affects around 334 million people worldwide and its prevalence is increasing every year [1]. In Japanese adults, asthma prevalence has increased from about 1% to 6%-10% since the 1960s [2]. Although regular follow-up is recommended for patients with asthma, adherence to asthma treatment guidelines is generally poor in many countries [3-6]. As a result of poor adherence, appropriate asthma control is not achieved, which contributes to reduced quality of life and depressive symptoms, and causes sleeplessness, daytime fatigue, and school and work absenteeism as well as an increased number of exacerbations [7-9]. Previous reports have shown that adherence to asthma medication regimens tends to be not so good in Japanese patients with asthma, even though they are under specialist care for their asthma [10-13]. These data suggest that for many Japanese patients, asthma remains poorly controlled, despite the availability of nationally promoted standards and improved agents for asthma treatment.

Mobile Health App

Mobile health (mHealth) involves the use of information and communication technology to improve the understanding of health care and is superseding previous forms of surveys, including traditional postal mail, telephone, and web page–based surveys [14,15]. Smartphones, which are used by around 40%-60% people in Japan, are used to collect real-world data directly from patients, which has been shown to be both cost-effective and fast [16-19]. mHealth research apps are often based on the ResearchKit platform (Apple Inc.), an iOS-based open-source framework for mobile medical research that was released in 2015. Moreover, iPhones (Apple Inc.) are used by around 60% of Japanese smartphone users, based on 2016 figures [20]. ResearchKit has been used in observational real-world studies of cardiovascular health, asthma, rheumatoid arthritis, Parkinson disease, type 2 diabetes, and cancer [14,17-19,21-23]. Although ResearchKit-based apps have been used in real-world observational studies of asthma, there are currently no studies using this app for Japanese patients with asthma. This noninterventional and observational study using the Zensoku-Log app assessed the current levels of asthma control as reported by Japanese patients with asthma and cough variant asthma (CVA) in a real-world setting.

Methods

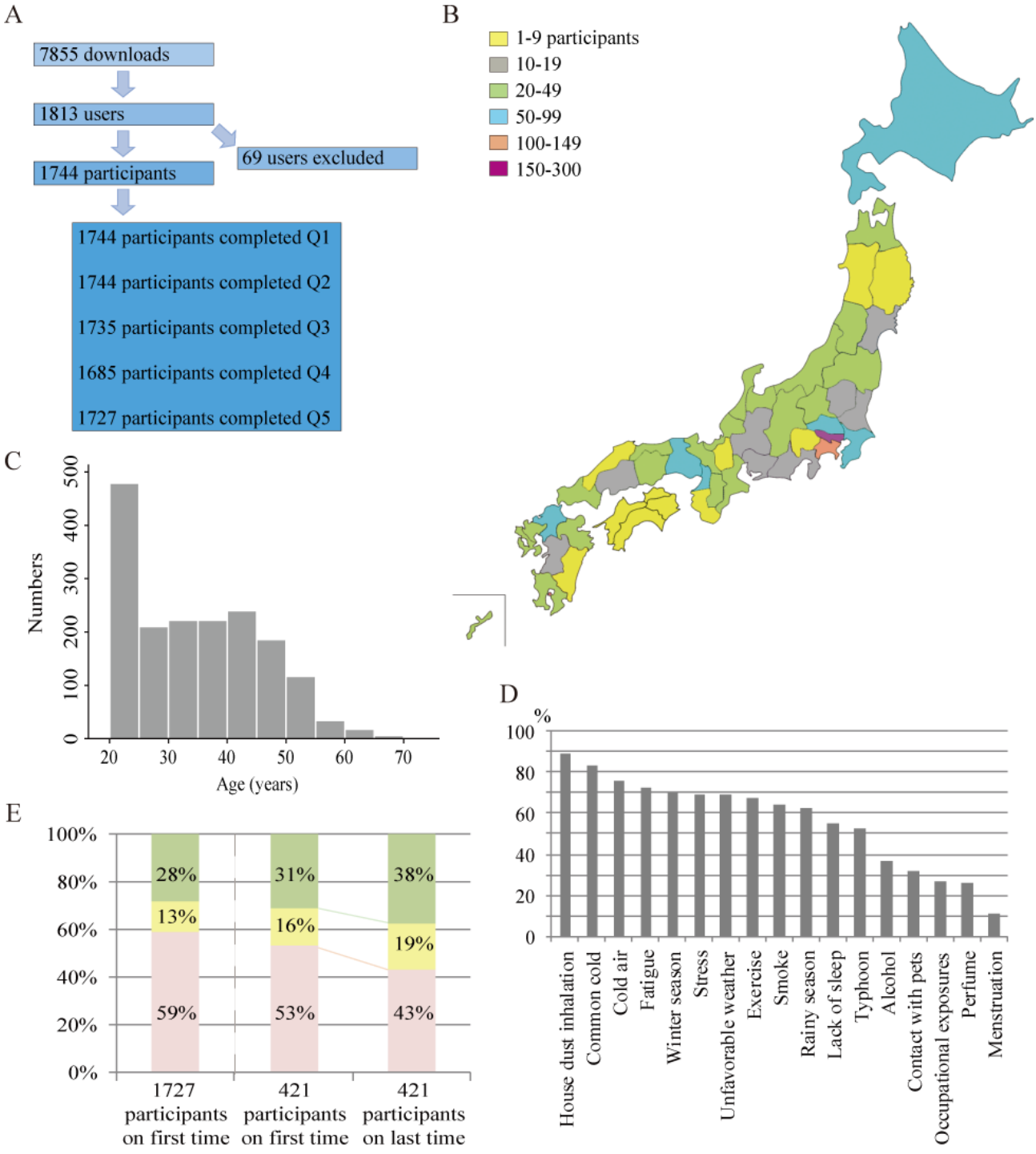
App and Participant Recruitment

We created and developed the Zensoku-Log app using the ResearchKit platform of Medical Logue, Inc. The Zensoku-Log app was released on February 16, 2016, through the Apple App Store in Japan for iPhone 6, iPhone 6 Plus, iPhone 6s, and iPhone 6s Plus (Apple Inc.). Upon downloading and opening the app, participants had to view all informed consent screens and read about the risks and benefits of participating in the study and their right to withdraw from participation. During the electronic informed consent process, participants agreed to share data with the understanding that their personally identifiable information would be collected; they were also provided with an option for *Consent withdrawal* in the app that will allow them to withdraw from the study at any time without having to provide reasons. Once they consented to participate, they answered questions related to eligibility screening (age 20 years or older with physician-diagnosed asthma or CVA) and questions to confirm the certainty of their diagnosis of asthma or CVA (Multimedia Appendix 1).

This study was reviewed and approved by the Juntendo University Research Ethics Committee (Tokyo, Japan), and was registered in the UMIN Clinical Trial Registry (UMIN000021043) on February 16, 2016 (http://www.umin.ac.jp/). The app recorded all data collected for this study through a set of web services for data storage and security developed and operated by C2, Inc.

We conducted this mobile-based, self-reporting, observational survey of patients with asthma in Japan between February 2016 and February 2018. The Zensoku-Log app was downloaded 6491 times within 2 weeks of its launch and was downloaded 7855 times overall during the study period. Of the users who downloaded the app, 1813 consented to participate in the study and agreed to share their data; however, 69 users were excluded (17 had input incorrect information and 52 were not diagnosed as having asthma or CVA). The remaining 1744 participants completed the Question 1 (Q1) and Question 2 (Q2) sections of the survey questionnaire (Figure 1A and Multimedia Appendix 1).

Figure 1. Recruitment process, geographic distribution, asthma triggers, and monthly Mobile Asthma Control Questionnaire (MACQ) surveys. (A) Patient recruitment process. (B) The geographic distribution of baseline users in Japan. (C) The age distribution of baseline users. (D) Percentages of factors that worsen asthma. (E) Proportion of patients with asthma control based on the MACQ surveys. The left column indicates the MACQ score for the 1727 patients who initially answered. Among the 421 patients who repeated the MACQ survey, the middle and the right column represents the MACQ score for the initial round and the last round, respectively. Green, yellow, and pink columns indicate total control (score of 15), well control (score of 13 or 14), and poor control (score of less than 13) of asthma, respectively.



Study Survey Design

We were unable to use certain standard validated questionnaires and quality of life surveys for patients with asthma directly in this study due to licensing constraints. Therefore, our asthma specialists developed a survey questionnaire by incorporating the general content used by validated survey instruments. Our survey questionnaire evaluated characteristics, including medical history, complications, familial atopic predisposition, social history, asthma trigger, and symptoms, in Q1 and Q2, asthma medications in the Question 3 (Q3) section, adherence to asthma medication in the Question 4 (Q4) section, and asthma control in the Question 5 (Q5) section (Multimedia Appendix 1). After the included participants (N=1744) initially answered Q1, 1744, 1735, 1685, and 1727 participants completed Q2, Q3, Q4, and Q5, respectively (Figure 1A). Furthermore, the Zensoku-Log

app could collect user locations (latitude/longitude) and atmospheric pressure optionally at any time using the device's global positioning system. After the first survey, users were able to repeat their answers in Q4 and Q5 optionally.

Monthly Questionnaire for Assessing Asthma Control

The monthly Mobile Asthma Control Questionnaire (MACQ) in Q5 was created based on the Asthma Control Test and *assessment of asthma symptom control* section in the Global Initiative for Asthma guidelines [24-26]. The MACQ survey is a patient-completed questionnaire with 5 items that assess daytime asthma symptoms, nocturnal symptoms, activity limitations due to asthma, use of rescue medications, and asthma attack (Multimedia Appendices 1 and 2). A total of 4 items included 3 response options, corresponding to a 3-point system, and 1 item for asthma attack included 2 response options, as indicated in Multimedia Appendix 2. The score to be validated in this study is called the MACQ score, which was computed by summing all these 5 items to yield a score ranging from 5 to 15. An MACQ score of 15, a score of 13 or 14, and a score of less than 13 were defined as total control, well control, and poor control of asthma, respectively.

Adherence Starts With Knowledge 12 (ASK-12)

The Adherence Starts with Knowledge 12 (ASK-12) questionnaire for Japanese adults with asthma was used for assessing adherence to asthma medication in our app [10,27,28]. For each item of the ASK-12 questionnaire, participants were asked to rate their potential barriers to medication adherence using a 5-point scale (all items are listed in Multimedia Appendix 1). For items 1-3 and 8-12, higher scores indicated stronger barriers to adherence. Items 4-7 were reverse-scored so that their final scores were in the same direction as those of the other 8 items. The ASK-12 total score was computed by summing all 12 items. The total barrier count (TBC) was developed as a simplified scoring approach [27,29]. To compute the TBC, the response to each item was dichotomized, with some sections (responses) in the 1-5 scale indicating no barrier "0", and the remaining sections (responses) indicating a barrier "1" [27]. The ASK-12 questionnaire total score and the TBC score have a possible range of 12-60 and 0-12, respectively, with higher scores representing stronger barriers to adherence.

Statistical Analysis

Sample normality was examined using the D'Agostino–Pearson test. Differences in parameters between populations were analyzed for significance using the Wilcoxon rank-sum test. To assess correlations between variables, Spearman rank correlation coefficient was used, where appropriate. Differences were statistically significant when *P* values were .05 or less. Statistical analyses were performed in Graph Pad Prism software (version 6; GraphPad Software, Inc.).

Results

Clinical and Demographic Characteristics

A total of 1624 patients with physician-diagnosed asthma and 120 patients with physician-diagnosed CVA were enrolled in this study. The baseline characteristics of eligible patients are summarized in Table 1 and the geographic distribution of the study participants is shown in Figure 1B (n=1440). The male-to-female ratio was 38.7:61.3, the median age was 33 years (range 20-74 years), and the median duration of asthma was 15 years (range 0-57 years). The age distribution demonstrated that the study participants were predominantly a younger population (Figure 1C). As much as 91.89% (1598/1739) of individuals remained after exclusion of those who did not respond definitively to the question related to atopy (ie, these individuals selected "Not sure") and these individuals had some form of atopic predisposition (Table 1 and Multimedia Appendix 1). The familial predisposition to atopic disorders included asthma, allergic rhinitis, atopic dermatitis, urticaria, hay fever, food allergy, and drug allergy (Table 1 and Multimedia Appendix 3) and 88.30% (1532/1735) of participants had some form of familial atopic predisposition (Table 1 and Multimedia Appendix 1).

The patients' clinical characteristics (Multimedia Appendix 4) showed that 50.97% (889/1744) and 33.03% (576/1744) of the 1744 participants had unscheduled visits and were absent from school or work for asthma symptoms, respectively, even though 62.84% (1096/1744) of participants attended regularly scheduled visits. Overall, 99.43% (1734/1744) of participants reported that their asthma symptoms could be stimulated by a trigger, including house dust, common cold, cold air, and so on (Figure 1D). Moreover, Table 1 shows that 23.14% (402/1737) and 40.75% (705/1730) of participants were current smokers and had pets, respectively. These data suggest that the Zensoku-Log users were patients with poorly controlled asthma which resulted from exposure to environmental factors.

Table 1. Demographic characteristics of Zensoku-Log users.

Characteristics	Values
Diagnosis of asthma, n/N (%)	1624/1744 (93.12)
Diagnosis of cough variant asthma, n/N (%)	120/1744 (6.88)
Sex	
Male, n/N (%)	675/1744 (38.70)
Female, n/N (%)	1069/1744 (61.30)
Age (year), median (range)[a]	33 (20-74)
Body height (m), median (range)[a]	162 (140-196)
Body weight (kg), median (range)[a]	60 (32-135)
BMI (kg/m^2), median (range)[a]	22.5 (12.7-49.9)
Age at asthma onset (year), median (range)[b]	15 (0-61)
Duration of asthma (year), median (range)[c]	15 (0-57)
Number of remissions of asthma symptoms, n/N (%)	571/1580 (36.14)
Duration in remission (year), median (range)[d]	8 (1-43)
Smoking history status (never/ex/current)	
Never, n/N (%)	955/1737 (54.98)
Ex-smoker, n/N (%)	380/1737 (21.88)
Current, n/N (%)	402/1737 (23.14)
Smoking history (pack-year), median (range)[e]	7.5 (0.1-105)
Having animals or birds as pets, n/N (%)	705/1730 (40.75)
Atopic predisposition, n/N (%)	1598/1739 (91.89)
Familial atopic predisposition, n/N (%)	1532/1735 (88.30)

[a]N=1737.
[b]N=1682.
[c]N=1675.
[d]N=406.
[e]N=750.

Comorbidity

Our data demonstrated that 75.29% (1313/1744) of participants with asthma reported that they suffered from allergic rhinitis (Table 2). Patients with allergic rhinitis usually also have allergic conjunctivitis, and atopic eczema frequently precedes allergic rhinitis. In this survey, 91.89% (1598/1739) of participants with asthma reported that they have a predisposition to atopic disorders (ie, allergic rhinitis, atopic dermatitis, urticaria, allergic conjunctivitis, hay fever, and certain drug-induced reactions; Tables 1 and 2). Among our participants with asthma, 8.20% (143/1744) and 1.32% (23/1744) had aspirin-exacerbated respiratory disease and allergic bronchopulmonary mycoses, respectively (Table 2). Although over half of participants had some symptoms of gastroesophageal reflux (GER), including heartburn and burping, only 20.30% (354/1744) of participants actually suffered from GER and gastroesophageal reflux disease (GERD; Table 2 and Multimedia Appendix 5). These data suggest that Zensoku-Log users require treatment of their comorbidities.

Table 2. Comorbidities (N=1744).

Comorbidities	n (%)
Aspirin-exacerbated respiratory disease	143 (8.20)
Allergic rhinitis	1313 (75.29)
Chronic sinusitis	363 (20.81)
Nasal polyp	98 (5.62)
Atopic dermatitis	554 (31.77)
Urticaria	546 (31.31)
Allergic conjunctivitis	614 (35.21)
Pollen hay fever	1039 (59.58)
Food allergy	520 (29.82)
Drug allergy	307 (17.60)
Allergic bronchopulmonary aspergillosis or allergic bronchopulmonary mycoses	23 (1.32)
Chronic obstructive pulmonary disease or emphysema	31 (1.78)
Bronchial ectasia	65 (3.73)
Gastroesophageal reflux disease or gastroesophageal reflux	354 (20.30)
Sleep apnea syndrome	255 (14.62)
Depression or autism	277 (15.88)
History of pneumonia	507 (29.07)
History of tuberculosis	21 (1.20)
Interstitial pneumoniae	18 (1.03)
Lung cancer	1 (0.06)

Asthma Medications

In the asthma treatment section, 1735 participants completed the questionnaire. A daily inhaled corticosteroid (ICS) had been prescribed as asthma controller medicine for 89.45% (1552/1735) of participants and for all excluded individuals (n=1552) that did not respond to this questionnaire (Table 3). Although a daily oral corticosteroid had been prescribed for 22.07% (383/1735) of participants, a leukotriene receptor antagonist, theophylline, and a long-acting muscarinic antagonist had been prescribed for only 30.66% (532/1735), 15.91% (276/1735), and 4.38% (76/1735) of the participants, respectively (Table 3). These results, taken together with demographic and clinical characteristics shown in Table 1 and Multimedia Appendix 4, suggest that Zensoku-Log users were patients with poorly controlled asthma who require more treatment for asthma control.

Table 3. Asthma treatments (N=1735).

Treatment	Value
Oral corticosteroids, n (%)	383 (22.07)
ICS[a], n (%)	1552 (89.45)
Daily dose of ICS (FP[b] equivalent dose, μg), median (range)[c]	500 (50-2500)
ICS/LABA[d] combination products, n (%)	1298 (74.81)
Single-agent ICS products, n (%)	293 (16.89)
Tulobuterol tape, n (%)	320 (18.44)
Leukotriene receptor antagonist, n (%)	532 (30.66)
Theophylline, n (%)	276 (15.91)
Long-acting muscarinic antagonist, n (%)	76 (4.38)
Disodium cromoglycate, n (%)	78 (4.50)
Omalizumab, n (%)	26 (1.50)

[a]ICS: inhaled corticosteroid.

[b]FP: fluticasone propionate.

[c]N=1216.

[d]LABA: long-acting β adrenoceptor agonist.

Asthma Control

A total of 1727 participants initially answered the Q5 section (asthma control section) of the questionnaire; of these, 1159 (67.11%) participants reported that they experienced good asthma control (Multimedia Appendix 6). However, over half of participants had some symptoms related to asthma or rhinitis and 58.48% (1010/1727) of participants had poor asthma control during the 4 weeks before scoring the MACQ survey (Figure 1E and Multimedia Appendix 6). Table 4 shows the number of each type of response for each item in the MACQ survey; the left column in Figure 1E demonstrates that 28.26% (488/1727) of participants were under complete and total control (Table 4 and Figure 1E). In the 421 participants who repeated the MACQ survey, the MACQ score for the last round was significantly increased as compared with the first round (P=.002). Figure 1E demonstrates that the frequency of participants with total and well control of asthma in the last round of the MACQ was increased as compared with the initial round. These results suggest that most Zensoku-Log users were patients with poorly controlled asthma, and that repeated use of the Zensoku-Log app may contribute to achieving better asthma control among users of the app.

Table 4. Monthly Mobile Asthma Control Questionnaire (N=1727).

Question	Yes, n (%)		No, n (%)
	Two or more times a week	Once a week or less	
1. Have you had any daytime asthma symptoms in the past 4 weeks?	545 (31.56)	495 (28.66)	687 (39.78)
2. Have you had any asthma symptoms at nighttime during the past 4 weeks?	475 (27.50)	409 (23.68)	843 (48.81)
3. Have you limited your activities, including exercising, because of your asthma in the past 4 weeks?	217 (12.57)	205 (11.87)	1305 (75.56)
4. Have you used your rescue medication to relieve asthma symptoms in the past 4 weeks?	376 (21.77)	342 (19.80)	1009 (58.43)
5. Have you had an asthma attack in the past 4 weeks?	827 (47.89)	—[a]	900 (52.11)

[a]Only yes or no.

Adherence to Asthma Medication

A previous study reported that patients reporting low adherence to ICS had poorer asthma control than better adherers [30]. In our survey, 61.53% (966/1570) of participants answered that they have stopped taking ICS because 81.1% (784/967) of them had felt better or 81.6% (787/965) of them had improved asthma attacks (Multimedia Appendix 7). Takemura et al. [28] reported that the optimal cut-off value of the ASK-12 total score for discriminating nonadherent patients was 23. The ASK-12 total score and the TBC score in 1569 participants were 28.7 and 4.2, respectively (Multimedia Appendix 8), suggesting that Zensoku-Log users were patients with poor adherence. Among the 306 participants who repeated the ASK-12 questionnaire, the TBC score, but not the ASK-12 total score, was significantly decreased (P=.001) in the last round as compared with the first

round of answering ASK-12 (Table 5 and Multimedia Appendix 8). Although the scores of Q1 (P<.001), Q2 (P<.001), Q6 (P=.01), and Q7 (P=.003) in ASK-12 in the last round of answering were significantly decreased as compared with the initial round, the scores of Q8 (P=.03), Q9 (P<.001), Q11 (P<.001), and Q12 (P=.004) of the ASK-12 in the last round of answering were significantly increased as compared with the initial round (Multimedia Appendix 8). These results, nevertheless, suggest that repeated use of the Zensoku-Log app may contribute to reducing medication adherence barriers.

Table 5. A summary of the ASK-12 questionnaire (N=306).[a]

Question	TBC[b], Mean (SD)		P value[c]
	First	Last	
1. I just forget to take my medicines some of the time.	0.56 (0.50)	0.43 (0.50)	<.001
2. I run out of my medicine because I don't get refills on time.	0.49 (0.50)	0.40 (0.49)	.002
3. Taking medicines more than once a day is inconvenient.	0.54 (0.05)	0.51 (0.50)	.28
4. I feel confident that each one of my medicines will help me.	0.17 (0.37)	0.17 (0.37)	>.99
5. I know if I am reaching my health goals.	0.49 (0.50)	0.47 (0.50)	.49
6. I have someone who I can call with questions about my medicines.	0.16 (0.37)	0.11 (0.32)	.04
7. My doctor/nurse and I work together to make decisions.	0.32 (0.47)	0.25 (0.47)	.03
8. Taken a medicine more or less often than prescribed?	0.41 (0.49)	0.45 (0.50)	.23
9. Skipped or stopped taking a medicine because you didn't think it was working?	0.12 (0.32)	0.14 (0.35)	.30
10. Skipped or stopped taking medicine because it made you feel bad?	0.07 (0.25)	0.08 (0.27)	.59
11. Skipped, stopped, not refilled, or taken less medicine because of the cost?	0.10 (0.31)	0.15 (0.35)	.07
12. Not had medicine with you when it was time to take it?	0.16 (0.37)	0.24 (0.43)	.006
Total	3.59 (2.32)	3.17 (2.40)	.001

[a]We used the Japanese version of ASK-12 in this study.

[b]TBC: total barrier count.

[c]Comparisons performed by Wilcoxon signed-rank test.

Discussion

Principal Results

This observational study utilized an mHealth app to survey Japanese patients with asthma and CVA in a real-world setting. To the best of our knowledge, such a study has not been performed previously. This clinical observational study demonstrated that the Zensoku-Log app was useful for remote recruitment of Japanese patients, using an electronic informed consent form, with a wide geographic distribution across Japan, via their iPhones and without requiring intervention by medical staff. This app was more efficient than traditional survey studies that used conventional postal mail, telephone, and web page-based surveys, which are both costly and time-consuming. Different proportions of app downloads were investigated in previous mHealth studies conducted in different parts of the world; for example, in the United States, in the Asthma Mobile Health Study for asthma patients, 19% of downloads were investigated, whereas in the mPower study for Parkinson disease, 20% of downloads were analyzed. In Japan, the DryEyeRhythm study for patients with dry eye evaluated 23% of downloads and the GlucoNote study for patients with type 2 diabetes analyzed 31% of downloads [17,21,31,32]. Although our app was downloaded 7855 times, given the relative ease of app downloading, we attributed the moderate enrollment rate (22.20% [1744/7855] of all downloads), with some drop-off, to the huge volume of questionnaires to be filled in. This app allowed collection of a range of asthma-related data, including symptoms, comorbidities, quality of life, medications, asthma control, and adherence. These data suggested that the Zensoku-Log users were patients with poorly controlled asthma who require more treatment for asthma and their comorbidities, and who need to improve adherence. The geographic distribution of participants (Figure 1B) suggested that patients in metropolitan cities may have higher medical literacy or poorer asthma control or may more frequently be iPhone users. Furthermore, Figure 1C demonstrated that the age distribution in this study was biased toward a younger population. This age distribution may be the cause of the poor asthma control in the real-world setting, suggesting that the mHealth app may reach participants that are typically not reached adequately by traditional studies. Prevalence of allergic rhinitis in the United States is about 15%, based on physician diagnosis, and reaches 30% based on self-reported symptoms [33]. Over 42% of patients with asthma have allergic rhinitis, and up to 40% of patients with allergic rhinitis have or will have asthma [33,34]. Among the Zensoku-Log users, 91.89% (1598/1739) of participants with asthma reported that they have a predisposition to atopic predisposition. These findings and the finding of fewer patients with chronic obstructive pulmonary disease were also associated with a younger population bias in this study as compared with a previous review of comorbidities in severe asthma [35]. Moreover, because the number of patients with GER and GERD was less than the number expected from the

results in Multimedia Appendix 5, it is possible that patients themselves underestimate GER. It may be considered that poor asthma control is due to inadequate asthma treatment, poor adherence, inadequate environmental management (eg, current smoking), and inadequate management of comorbidities. Therefore, for asthma management, further education is necessary, including guidelines and development and implementation of a mobile app that can improve a patient's self-management [36].

Three previous systematic reviews showed that mHealth app-based interventions, including providing reminders, have the potential to improve asthma control and medication adherence in patients with asthma [37-39]. Therefore, it is expected that the app-based intervention with the reminder will be more effective, such as facilitating the delivery of care and connecting patients to their medical staff. The COVID-19 (coronavirus disease 2019) pandemic requires the transformation of traditional medical practices. New digital health innovations will help optimize the efficiency of health care systems and improve patient outcomes. In addition to smartphone therapeutic apps, other components such as sensors, video, social media platforms, wearables, or a combination of these can help enable health care delivery and overcome distance, location, and time constraints. Digital treatment is evolving and it is necessary to involve government authorities/institutions in the approval process. In this study, the repeated use of this app appeared to contribute to achieving better asthma control and to reducing medication adherence barriers (Figure 1E and Table 5). It suggested that even an app designed for observational research, such as ours, which does not involve any app-based interventions, may be a useful educational medium. However, the findings based on the self-reported clinical data in this study may not be representative of the general asthma population and consequently, there is a need for further studies in this area.

Limitations

Our study had several limitations. First, only Japanese iPhone users could be enrolled, increasing the risk of selection and socioeconomic bias, because it has been shown in the United States that iPhone owners have higher education and income levels than users of other smartphones [17]. However, it is uncertain whether this applies to Japanese users, because according to a StatCounter report [20], the share of iPhones in the Japanese smartphone market is around 60%. However, the proportion of people with low education and low income levels among this group currently remains unknown. Moreover, this study has a potential selection bias toward a younger population. Although app developers should pay attention to usability aspects of mHealth apps during development and release [40], this study paid little attention to these aspects for users of other smartphones and users unfamiliar with using such apps. These biases also present vast differences in the ability of patients to find the health information they need on the internet, a phenomenon also known as "digital divide," which is one of the most important issues related to the use of mHealth technology. To reduce digital divide across populations disproportionately impacted by health information, it is necessary to make efforts to refine mHealth apps as a study tool (eg, by implementing usability testing to establish how well the app works and serves its intended purpose). All data analyzed in this study were entirely based on participants' self-motivation and self-reporting, with no way to validate these data by medical personnel; however, such a subjective self-reporting can be prone to bias, with patients likely to overestimate or underestimate the actual scenario [23]. Because these biases are also reported by similar studies using mobile apps, a randomized controlled trial should be performed to confirm the findings of this study. In addition, the validity of the study data collected within the app was difficult to be determined objectively. Our survey questionnaire is developed solely for this app, and thus may limit comparison with studies using standardized instruments. However, the validity of the MACQ score was supported by concordance between this score at baseline and another questionnaire for assessing asthma control. For example, participants' daily symptoms were all found to be significantly associated with the MACQ score (Multimedia Appendix 9). Moreover, the MACQ score correlated with the TBC score and medication adherence barriers (Multimedia Appendix 9).

Although we requested participants to enter the values of fractional exhaled nitric oxide and peak expiratory flow as objective parameters (Q2 and Q5), these data were missing from their profiles, possibly because participants were unfamiliar with these terms and this information was self-reported (ie, without any assistance of a medical staff). The ResearchKit platform offers the positioning advantage of completely remote recruitment and enrollment, but lack of human communication may cause less motivation for participants to continue with the questionnaire, as compared with studies conducted face to face [17-19,31]. In this study, the proportion of participants repeating the MACQ survey was only 24.38% (421/1727 participants). Therefore, further studies are required to improve retention rates with long-term use. Moreover, although mHealth offers some advantages over traditional face-to-face methods, data collection and interview for delivering behavioral interventions in face-to-face methods are more precise and accurate than the mHealth method. Besides, the face-to-face methods also report better treatment adherence than studies using the mHealth methods.

Conclusions

We have reported on the use of an mHealth app for Japanese patients with asthma and CVA in a real-world setting. To the best of our knowledge, such a study has not been performed previously. These app users were patients with poorly controlled asthma, who may need more treatment for asthma and their comorbidities. Moreover, repeated use of this app may have improved asthma control. This new mHealth app has novel possibilities for data collection and can reach participants that traditional studies may fail to represent adequately. However, it will be necessary to address the challenges associated with this technology, including selection bias, potential reporting bias, data security, usability issues, digital divides, and low user-retention rate [17].

Acknowledgments

The Zensoku-Log was developed by Medical Logue, Inc. (Tokyo, Japan) using Apple's ResearchKit platform (Apple; Cupertino, CA, USA), and a set of web services for data storage and security developed and operated by C2, Inc. (Tokyo, Japan). We thank Editage (www.editage.com) for English language editing.

Authors' Contributions

NH, JI, RA, SaH, and KT participated in the design of the study and drafted the manuscript. NH, SoH, JI, RA, and SaH performed the statistical analysis and interpretation of the results. All authors have read and approved the final manuscript.

Conflicts of Interest

NH reports personal fees from AstraZeneca outside the submitted work and has a patent pending (Japanese Patent Application 2018-097070). SaH reports receiving a grant from Fujitsu Limited during the conduct of the study. KT reports grants and personal fees from AstraZeneca, Boehringer Ingelheim, Chugai, Eli Lilly, KYORIN, MSD, Nobelpharma, Novartis, ONO, Pfizer, TEIJIN, and grants from Astellas, GSK, Kyowa Hakko Kirin, MiZ, MOCHIDA, Nipro, NIPPON SHINYAKU, TAIHO, TOYAMA CHEMICAL, TSUMURA, Sanofi, SHIONOGI, TORII, and personal fees from Bristol-Myers, Eisai, Meiji Seika, Mitsubishi Tanabe, Otsuka, PAREXEL, Sumitomo Dainippon, outside the submitted work. All other authors have no conflict of interest to declare.

Multimedia Appendix 1
Survey Questionnaire.
[XLSX File (Microsoft Excel File), 21 KB - jmir_v22i8e19006_app1.xlsx]

Multimedia Appendix 2
Monthly mobile asthma control questionnaire (MACQ).
[XLSX File (Microsoft Excel File), 15 KB - jmir_v22i8e19006_app2.xlsx]

Multimedia Appendix 3
Family history of atopy.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v22i8e19006_app3.xlsx]

Multimedia Appendix 4
Clinical characteristics of Zensoku-Log users.
[XLSX File (Microsoft Excel File), 15 KB - jmir_v22i8e19006_app4.xlsx]

Multimedia Appendix 5
Questionnaire for the symptoms of gastroesophageal reflux.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v22i8e19006_app5.xlsx]

Multimedia Appendix 6
Questionnaire for assessing asthma control.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v22i8e19006_app6.xlsx]

Multimedia Appendix 7
Questionnaire for assessing adherence to ICS.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v22i8e19006_app7.xlsx]

Multimedia Appendix 8
Summary of ASK-12 questionnaire.
[XLSX File (Microsoft Excel File), 16 KB - jmir_v22i8e19006_app8.xlsx]

Multimedia Appendix 9
Correlation coefficients for the association of the MACQ score and another questionnaire for assessing asthma control.
[XLSX File (Microsoft Excel File), 15 KB - jmir_v22i8e19006_app9.xlsx]

References

1. Papi A, Brightling C, Pedersen SE, Reddel HK. Asthma. Lancet 2018 Feb 24;391(10122):783-800. [doi: 10.1016/S0140-6736(17)33311-1] [Medline: 29273246]
2. Ichinose M, Sugiura H, Nagase H, Yamaguchi M, Inoue H, Sagara H, Japanese Society of Allergology. Japanese guidelines for adult asthma 2017. Allergol Int 2017 Apr;66(2):163-189 [FREE Full text] [doi: 10.1016/j.alit.2016.12.005] [Medline: 28196638]
3. Adachi M, Ohta K, Morikawa A, Nishima S, Tokunaga S, Disantostefano RL. [Changes in asthma insights and reality in Japan (AIRJ) in 2005 since 2000]. Arerugi 2008 Feb;57(2):107-120. [Medline: 18349585]
4. Rabe KF, Vermeire PA, Soriano JB, Maier WC. Clinical management of asthma in 1999: the Asthma Insights and Reality in Europe (AIRE) study. Eur Respir J 2000 Nov;16(5):802-807 [FREE Full text] [Medline: 11153575]
5. Pike KC, Levy ML, Moreiras J, Fleming L. Managing problematic severe asthma: beyond the guidelines. Arch Dis Child 2018 Apr;103(4):392-397. [doi: 10.1136/archdischild-2016-311368] [Medline: 28903951]
6. Bukstein DA. Patient adherence and effective communication. Ann Allergy Asthma Immunol 2016 Dec;117(6):613-619. [doi: 10.1016/j.anai.2016.08.029] [Medline: 27979018]
7. Arditi C, Burnand B, Peytremann-Bridevaux I. Adding non-randomised studies to a Cochrane review brings complementary information for healthcare stakeholders: an augmented systematic review and meta-analysis. BMC Health Serv Res 2016 Oct 21;16(1):598 [FREE Full text] [doi: 10.1186/s12913-016-1816-5] [Medline: 27769236]
8. Papi A, Ryan D, Soriano JB, Chrystyn H, Bjermer L, Rodríguez-Roisin R, et al. Relationship of Inhaled Corticosteroid Adherence to Asthma Exacerbations in Patients with Moderate-to-Severe Asthma. J Allergy Clin Immunol Pract 2018;6(6):1989-1998.e3. [doi: 10.1016/j.jaip.2018.03.008] [Medline: 29627457]
9. Haktanir Abul M, Phipatanakul W. Severe asthma in children: Evaluation and management. Allergol Int 2019 Apr;68(2):150-157 [FREE Full text] [doi: 10.1016/j.alit.2018.11.007] [Medline: 30648539]
10. Atsuta R, To Y, Sakamoto S, Mukai I, Kobayashi A, Kinoshita A, et al. Assessing usability of the "Adherence Starts with Knowledge 20" (ASK-20) questionnaire for Japanese adults with bronchial asthma receiving inhaled corticosteroids long term. Allergol Int 2017 Jul;66(3):411-417 [FREE Full text] [doi: 10.1016/j.alit.2016.09.001] [Medline: 27712949]
11. Koya T, Hasegawa T, Takasawa J, Yoshimine F, Sakagami T, Hayashi M, Niigata Inhalation Treatment Study Group. Influence of Adherence to Inhaled Corticosteroids and Inhaler Handling Errors on Asthma Control in a Japanese Population. Intern Med 2018 Dec 01;57(23):3357-3363 [FREE Full text] [doi: 10.2169/internalmedicine.0986-18] [Medline: 30101909]
12. Sasaki J, Kawayama T, Yoshida M, Takahashi K, Fujii K, Machida K, the Kyushu Asthma Seminar Investigators. Adherence barriers to inhaled medicines in Japanese older patients with asthma evaluated using the "Adherence Starts with Knowledge 20" (ASK-20) questionnaire. J Asthma 2019 Jun;56(6):632-641. [doi: 10.1080/02770903.2018.1484132] [Medline: 29969920]
13. Toyama T, Kawayama T, Kinoshita T, Imamura Y, Yoshida M, Takahashi K, et al. Differences in Adherence Barriers to Inhaled Medicines between Japanese Patients with Chronic Obstructive Pulmonary Disease and Asthma Evaluated using the "Adherence Starts with Knowledge 20" (ASK-20) Questionnaire. Intern Med 2019 Jan 15;58(2):175-185 [FREE Full text] [doi: 10.2169/internalmedicine.0488-17] [Medline: 30210101]
14. Schmitz H, Howe CL, Armstrong DG, Subbian V. Leveraging mobile health applications for biomedical research and citizen science: a scoping review. J Am Med Inform Assoc 2018 Dec 01;25(12):1685-1695. [doi: 10.1093/jamia/ocy130] [Medline: 30445467]
15. Bousquet J, Arnavielhe S, Bedbrook A, Bewick M, Laune D, Mathieu-Dupas E, et al. MASK 2017: ARIA digitally-enabled, integrated, person-centred care for rhinitis and asthma multimorbidity using real-world-evidence. Clin Transl Allergy 2018 Oct 25;8(1):51 [FREE Full text] [doi: 10.1186/s13601-018-0227-6] [Medline: 31624534]
16. Jardine J, Fisher J, Carrick B. Apple's ResearchKit: smart data collection for the smartphone era? J R Soc Med 2015 Aug;108(8):294-296. [doi: 10.1177/0141076815600673] [Medline: 26268915]
17. Chan YY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Apr;35(4):354-362 [FREE Full text] [doi: 10.1038/nbt.3826] [Medline: 28288104]
18. Dorsey ER, Yvonne CY, McConnell MV, Shaw SY, Trister AD, Friend SH. The Use of Smartphones for Health Research. Acad Med 2017 Feb;92(2):157-160. [doi: 10.1097/ACM.0000000000001205] [Medline: 27119325]
19. Crouthamel M, Quattrocchi E, Watts S, Wang S, Berry P, Garcia-Gancedo L, et al. Using a ResearchKit Smartphone App to Collect Rheumatoid Arthritis Symptoms From Real-World Participants: Feasibility Study. JMIR Mhealth Uhealth 2018 Sep 13;6(9):e177. [doi: 10.2196/mhealth.9656] [Medline: 30213779]
20. StatCounter GlobalStats. Mobile Vendor Market Share Japan, Feb 2016 - Feb 2018. URL: https://gs.statcounter.com/vendor-market-share/mobile/japan/#monthly-201602-201802 [accessed 2020-07-11]
21. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
22. Chan YY, Bot BM, Zweig M, Tignor N, Ma W, Suver C, et al. The asthma mobile health study, smartphone data collected using ResearchKit. Sci Data 2018 Dec 22;5:180096 [FREE Full text] [doi: 10.1038/sdata.2018.96] [Medline: 29786695]
23. McConnell MV, Shcherbina A, Pavlovic A, Homburger JR, Goldfeder RL, Waggot D, et al. Feasibility of Obtaining Measures of Lifestyle From a Smartphone App: The MyHeart Counts Cardiovascular Health Study. JAMA Cardiol 2017 Jan 01;2(1):67-76. [doi: 10.1001/jamacardio.2016.4395] [Medline: 27973671]

24. Global Initiative for Asthma (GINA). Global Strategy for Asthma Management and Prevention. 2006. URL: http://www.ginasthma.org/ [accessed 2020-07-11]
25. Nathan RA, Sorkness CA, Kosinski M, Schatz M, Li JT, Marcus P, et al. Development of the asthma control test: a survey for assessing asthma control. J Allergy Clin Immunol 2004 Jan;113(1):59-65. [doi: 10.1016/j.jaci.2003.09.008] [Medline: 14713908]
26. Schatz M, Sorkness CA, Li JT, Marcus P, Murray JJ, Nathan RA, et al. Asthma Control Test: reliability, validity, and responsiveness in patients not previously followed by asthma specialists. J Allergy Clin Immunol 2006 Mar;117(3):549-556. [doi: 10.1016/j.jaci.2006.01.011] [Medline: 16522452]
27. Matza LS, Yu-Isenberg KS, Coyne KS, Park J, Wakefield J, Skinner EP, et al. Further testing of the reliability and validity of the ASK-20 adherence barrier questionnaire in a medical center outpatient population. Curr Med Res Opin 2008 Nov;24(11):3197-3206. [doi: 10.1185/03007990802463642] [Medline: 18922213]
28. Takemura M, Nishio M, Fukumitsu K, Takeda N, Ichikawa H, Asano T, et al. Optimal cut-off value and clinical usefulness of the Adherence Starts with Knowledge-12 in patients with asthma taking inhaled corticosteroids. J Thorac Dis 2017 Aug;9(8):2350-2359 [FREE Full text] [doi: 10.21037/jtd.2017.06.115] [Medline: 28932539]
29. Hahn SR, Park J, Skinner EP, Yu-Isenberg KS, Weaver MB, Crawford B, et al. Development of the ASK-20 adherence barrier survey. Curr Med Res Opin 2008 Jul;24(7):2127-2138. [doi: 10.1185/03007990802174769] [Medline: 18554431]
30. Clatworthy J, Price D, Ryan D, Haughney J, Horne R. The value of self-report assessment of adherence, rhinitis and smoking in relation to asthma control. Prim Care Respir J 2009 Dec;18(4):300-305 [FREE Full text] [doi: 10.4104/pcrj.2009.00037] [Medline: 19562233]
31. Yamaguchi S, Waki K, Nannya Y, Nangaku M, Kadowaki T, Ohe K. Usage Patterns of GlucoNote, a Self-Management Smartphone App, Based on ResearchKit for Patients With Type 2 Diabetes and Prediabetes. JMIR Mhealth Uhealth 2019 Apr 24;7(4):e13204 [FREE Full text] [doi: 10.2196/13204] [Medline: 31017586]
32. Inomata T, Iwagami M, Nakamura M, Shiang T, Fujimoto K, Okumura Y, et al. Association between dry eye and depressive symptoms: Large-scale crowdsourced research using the DryEyeRhythm iPhone application. Ocul Surf 2020 Apr;18(2):312-319. [doi: 10.1016/j.jtos.2020.02.007] [Medline: 32113987]
33. Wheatley LM, Togias A. Clinical practice. Allergic rhinitis. N Engl J Med 2015 Jan 29;372(5):456-463 [FREE Full text] [doi: 10.1056/NEJMcp1412282] [Medline: 25629743]
34. Pawankar R, Bunnag C, Khaltaev N, Bousquet J. Allergic Rhinitis and Its Impact on Asthma in Asia Pacific and the ARIA Update 2008. World Allergy Organ J 2012 Apr;5(Suppl 3):S212-S217 [FREE Full text] [doi: 10.1097/WOX.0b013e318201d831] [Medline: 23268481]
35. Porsbjerg C, Menzies-Gow A. Co-morbidities in severe asthma: Clinical impact and management. Respirology 2017 May;22(4):651-661 [FREE Full text] [doi: 10.1111/resp.13026] [Medline: 28328160]
36. Tinschert P, Jakob R, Barata F, Kramer J, Kowatsch T. The Potential of Mobile Apps for Improving Asthma Self-Management: A Review of Publicly Available and Well-Adopted Asthma Apps. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e113 [FREE Full text] [doi: 10.2196/mhealth.7177] [Medline: 28768606]
37. Farzandipour M, Nabovati E, Sharif R, Arani MH, Anvari S. Patient Self-Management of Asthma Using Mobile Health Applications: A Systematic Review of the Functionalities and Effects. Appl Clin Inform 2017 Oct;8(4):1068-1081 [FREE Full text] [doi: 10.4338/ACI-2017-07-R-0116] [Medline: 29241254]
38. Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, Mackenzie AM, et al. Digital asthma self-management interventions: a systematic review. J Med Internet Res 2014;16(2):e51 [FREE Full text] [doi: 10.2196/jmir.2814] [Medline: 24550161]
39. Hui CY, Walton R, McKinstry B, Jackson T, Parker R, Pinnock H. The use of mobile applications to support self-management for people with asthma: a systematic review of controlled studies to identify features associated with clinical effectiveness and adherence. J Am Med Inform Assoc 2017 May 01;24(3):619-632. [doi: 10.1093/jamia/ocw143] [Medline: 27694279]
40. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for Evaluating the Content, Usability, and Efficacy of Commercial Mobile Health Apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]

Abbreviations

ASK-12: Adherence Starts with Knowledge-12
CVA: cough variant asthma
FP: fluticasone propionate
GER: gastroesophageal reflux
GERD: gastroesophageal reflux disease
ICS: inhaled corticosteroid
LABA: long-acting β-adrenoceptor agonist
MACQ: monthly Mobile Asthma Control Questionnaire
mHealth: mobile health
TBC: total barrier count

Edited by G Eysenbach; submitted 31.03.20; peer-reviewed by I El Akkary, D Dalan; comments to author 22.04.20; revised version received 05.05.20; accepted 14.05.20; published 14.08.20.

Please cite as:
Harada N, Harada S, Ito J, Atsuta R, Hori S, Takahashi K
Mobile Health App for Japanese Adult Patients With Asthma: Clinical Observational Study
J Med Internet Res 2020;22(8):e19006
URL: https://www.jmir.org/2020/8/e19006
doi:10.2196/19006
PMID:32795993

Original Paper

Development and Feasibility of a Mobile Health–Supported Comprehensive Intervention Model (CIMmH) for Improving the Quality of Life of Patients With Esophageal Cancer After Esophagectomy: Prospective, Single-Arm, Nonrandomized Pilot Study

Chao Cheng[1*], MD, PhD; Rainbow Tin Hung Ho[2,3*], PhD; Yan Guo[4,5,6*], PhD; Mengting Zhu[4], BS; Weixiong Yang[1], MD; Yiran Li[4], MB; Zhenguo Liu[1], MD; Shuyu Zhuo[7], MD; Qi Liang[8], MD; Zhenghong Chen[8], BS; Yu Zeng[4], MB; Jiali Yang[1], MD; Zhanfei Zhang[1], MD; Xu Zhang[9], MD; Aliza Monroe-Wise[10], MSc, MD; Sai-Ching Yeung[11], MD

[1]Department of Thoracic Surgery, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[2]Department of Social Work & Social Administration, The University of Hong Kong, Hong Kong, China

[3]Centre on Behavioral Health, The University of Hong Kong, Hong Kong, China

[4]Department of Medical statistics, School of Public Health, Sun-Yat-Sen University, Guangzhou, China

[5]Sun Yat-sen Center for Migrant Health Policy, Guangzhou, China

[6]Sun Yat-sen Center for Global Health, Guangzhou, China

[7]Department of Clinical Nutrition, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[8]Department of Rehabilitation Medicine, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[9]Department of Thoracic Surgery, Sun Yat-Sen University Cancer Center, Guangzhou, China

[10]Department of Global Health, University of Washington, Seattle, WA, United States

[11]Department of Emergency Medicine, Division of Internal Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

[*]these authors contributed equally

Corresponding Author:
Yan Guo, PhD
Department of Medical statistics
School of Public Health
Sun-Yat-Sen University
No. 74 2nd Zhongshan Road, Guangzhou 510080, China
Guangzhou, 510080
China
Phone: 86 0208733420
Email: guoy8@mail.sysu.edu.cn

Abstract

Background: Patients with esophageal cancer often experience clinically relevant deterioration of quality of life (QOL) after esophagectomy owing to malnutrition, lack of physical exercise, and psychological symptoms.

Objective: This study aimed to evaluate the feasibility, safety, and efficacy of a comprehensive intervention model using a mobile health system (CIMmH) in patients with esophageal cancer after esophagectomy.

Methods: Twenty patients with esophageal cancer undergoing the modified McKeown surgical procedure were invited to join the CIMmH program with both online and offline components for 12 weeks. The participants were assessed before surgery and again at 1 and 3 months after esophagectomy. QOL, depressive symptoms, anxiety, stress, nutrition, and physical fitness were measured.

Results: Of the 20 patients, 16 (80%) completed the program. One month after esophagectomy, patients showed significant deterioration in overall QOL (P=.02), eating (P=.005), reflux (P=.04), and trouble with talking (P<.001). At the 3-month follow-up, except for pain (P=.02), difficulty with eating (P=.03), dry mouth (P=.04), and trouble with talking (P=.003), all other QOL

dimensions returned to the preoperative level. There were significant reductions in weight (P<.001) and BMI (P=.02) throughout the study, and no significant changes were observed for physical fitness measured by change in the 6-minute walk distance between baseline and the 1-month follow-up (P=.22) or between baseline and the 3-month follow-up (P=.52). Depressive symptoms significantly increased 1 month after surgery (P<.001), while other psychological measures did not show relevant changes. Although there were declines in many measures 1 month after surgery, these were much improved at the 3-month follow-up, and the recovery was more profound and faster than with traditional rehabilitation programs.

Conclusions: The CIMmH was feasible and safe and demonstrated encouraging efficacy testing with a control group for enhancing recovery after surgery among patients with esophageal cancer in China.

Trial Registration: Chinese Clinical Trial Registry (ChiCTR-IPR-1800019900); http://www.chictr.org.cn/showprojen.aspx?proj=32811.

(J Med Internet Res 2020;22(8):e18946) doi:10.2196/18946

KEYWORDS

esophageal cancer; quality of life; nutrition; physical exercise; psychological support; mobile health; mHealth

Introduction

Esophageal cancer is the third most common cancer and the fourth most common cause of cancer death in China [1]. Esophagectomy is the major curative treatment option and is often performed in combination with neoadjuvant chemotherapy or chemoradiotherapy [2]. The surgery is considered extensive and entails a more than 40% risk of postoperative complications [3,4]. For example, anastomotic leakage is one of the most common postoperative complications, and it occurs in 5%-20% of patients with esophageal cancer [5,6]. Complications such as this may increase hospital stay, delay oral feeding, lead to malnutrition, increase psychological burden, cause poor quality of life (QOL), and subsequently worsen the long-term survival of patients [7-10].

Enhanced recovery after surgery (ERAS) is a patient-centered, evidence-based multimodal and multidisciplinary approach for promoting early recovery and reducing complications among patients after surgery [11]. Previous studies have shown the beneficial effects of interventions based on ERAS guidelines to improve the nutrition and physical status of patients with head, neck, and breast cancer after surgery [12,13]. For patients with esophageal cancer, existing interventions may be effective, but each intervention program usually focuses on one specific aspect of health such as nutrition or exercise [14-16]. To meet different needs, such as overcome postoperative complications and malnutrition, patients need to meet different professionals and to return to the hospital frequently for different appointments, creating potential obstacles for those who have been discharged from the hospital, especially those residing in rural or remote areas. Moreover, individual intervention programs may lack coherence. A comprehensive intervention program tailored to individual patients and designed to support patients holistically in all aspects, including nutrition, physical exercise, and psychosocial support, is thus urgently warranted for patients with esophageal cancer after esophagectomy.

In addition, effective information delivery and adherence to follow-up with health care professionals are of high priority in cancer care and are key elements of successful implementation of ERAS. In the past, most interventions were delivered face-to-face in either individual or group settings [12,17]. However, a face-to-face approach may not be easy for patients who live far away from the hospital or those who have physical difficulties in travelling. Therefore, a home-based supportive care intervention is warranted for discharged patients with esophageal cancer after esophagectomy [18,19]. An offer of timely guidance of home-based supportive care to patients can help reduce symptom distress or anxiety and prevent complications, and thus promote physical rehabilitation after surgery. In recent years, the wide adoption of mobile technology (eg, smartphones and mobile apps) offers a promising platform for efficient and accessible intervention delivery [20]. Mobile health (mHealth) refers to health care or health-related services delivered by mobile or other wireless devices, such as smartphones and tablets [21,22]. Several studies have demonstrated the efficacy of mHealth interventions for improving overall QOL in patients with endometrial, breast, or lung cancer [23-25]. However, there has been no mHealth-based intervention to improve QOL in patients with esophageal cancer after esophagectomy. Using an mHealth system on a mobile platform may be ideal for patients who have difficulty in making frequent follow-up visits to a hospital, which can be particularly challenging in China as hospitals are always situated in city centers. As 98.5% of people aged 50 to 80 years use the WeChat platform in China [26], an mHealth program delivered via WeChat would reach a substantial percentage of patients with cancer in China to support them at home.

Therefore, we designed the first comprehensive intervention model supported by mHealth (CIMmH) delivered on the WeChat platform, providing nutrition, exercise, and psychological support for patients with esophageal cancer after esophagectomy. This prospective pilot study aimed to examine the feasibility and safety of a 12-week CIMmH. The study will support the development of future programs for those patients with cancer who may not be able to visit the hospital frequently or who live in rural areas in China.

Methods

Study Design

This prospective, single-arm, nonrandomized pilot study was conducted at the First Affiliated Hospital of Sun Yat-sen University in Guangzhou, China, which has 2850 beds serving

4.9 million patients each year. The Department of Thoracic Surgery cares for more than 300 patients with esophageal cancer each year. The study was registered at the Chinese Clinical Trial Registry (ChiCTR-IPR-1800019900) and was approved by the ethics committee of the First Affiliated Hospital of Sun Yat-sen University according to the Declaration of Helsinki.

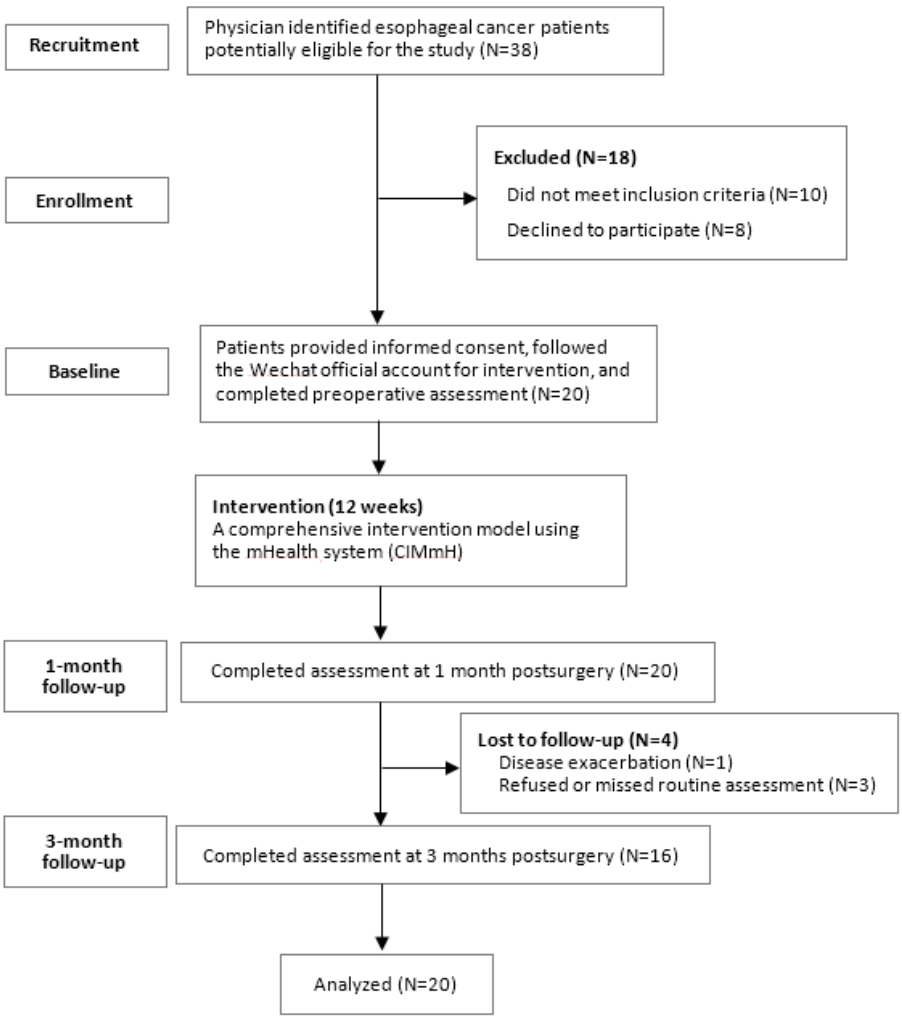
Participants

Patients diagnosed with esophageal cancer and scheduled for esophageal radical resection were referred by thoracic oncologists in the inpatient department of the hospital from December 2018 to October 2019. Those who met the eligibility criteria were invited to join the study. The inclusion criteria of the study were a diagnosis of esophageal cancer, suitability for the modified McKeown procedure (thoracoscopic esophageal mobilization three-incision esophagectomy) [27] and jejunostomy before surgery, age between 18 and 75 years with an expected survival of 12 months or longer, normal preoperative gastrointestinal function, Karnofsky performance scores of ≥90 before surgery, ability to walk continuously for 6 minutes or longer before surgery, own WeChat account or an account among family members, cognitive capability to understand Chinese and the study procedures, and ability to provide written informed consent. Individuals were excluded if they had esophageal carcinoma with distant metastases, were unable to engage in physical exercise owing to medical comorbidities, were HIV seropositive, were pregnant or lactating, were unable to finish the questionnaire owing to mental problems or other reasons, did not undergo R_0 resection, failed to have a jejunostomy feeding tube fitted, or had other serious medical, psychiatric, or cognitive illnesses that would interfere with their participation. Interested participants had appointments with trained research staff to receive further information about the study, signed the informed consent form, and completed the baseline assessment.

Study Flow

Figure 1 depicts the flow chart of the study. Of 38 patients diagnosed with esophageal cancer and scheduled for resection with the modified McKeown surgical approach, 10 did not meet the inclusion criteria and 8 declined to take part in the study. Personal reasons for rejection included concerns for patient privacy from family members and patients' unstable postoperative condition. Thus, the final sample included 20 patients. The overall response rate was 52.6% (20/38). Of the 20 patients enrolled, 16 (80%) completed the study and 4 (20%) dropped out at the 3-month follow-up owing to disease exacerbation, unwillingness to continue, or missing routine assessments.

Figure 1. CONSORT flow chart of the study.



Intervention Protocol

A 3-month CIMmH program was delivered to participants after surgery by specialists in the hospital (offline) and through the enhanced WeChat platform (online). The program included general guidelines on postsurgery recovery, strategies to cope with postoperative complications, nutrition guidelines, physical exercise promotion, and psychological support courses. Details of the CIMmH are provided in Table 1. The components of the CIMmH are presented below.

Table 1. Components of the 3-month CIMmH program.

Category	Presurgery	Postsurgery[a]			
	Week 1-2	Week 1 (before discharge)	Week 2-3 (nasogastric tube removed)	Week 4-6 (nasogastric tube feeding tube removed)	Week 7-12
General introduction	Introduction of the CIMmH[b] program (video by a doctor)	Introduction of the strategies to cope with common postoperative complications	Introduction of the strategies to cope with common postoperative complications	Introduction of the strategies to cope with common postoperative complications	Introduction of the strategies to cope with common postoperative complications
Nutrition	N/A[c]	Standard postoperative nutrition support in the hospital	Rehabilitation guidance of the jejunostomy feeding enteral nutrition period (video by a doctor and nurse) Home jejunostomy feeding enteral nutrition guidance (article) Home enteral nutrition guidance for the use of the feeding tube (video)	Rehabilitation guidance of the transitional period between enteral nutrition and ONS[d] (video by a doctor and nurse) Home nutrition guidance for the transitional period (article) Nutrition prescription by a nutritionist (face-to-face)	Rehabilitation guidance of ONS and oral intake (video by a doctor and nurse) Home oral nutrition guidance (article) Nutrition prescription by a nutritionist (face-to-face)
Physical exercise	Inspiratory muscle training	Walk promotion	Walk promotion	Walk promotion Baduanjin Qigong (video)	Walk promotion Baduanjin Qigong (video)
Psychological courses	N/A	N/A	N/A	Adapted MBCR[e] courses (articles and audio)	Adapted MBCR courses (articles and audio)
Data collection	Data collection at baseline (ie, about 1 week before surgery)	N/A	N/A	Data collection at 1 month after surgery	Data collection at 3 months after surgery

[a]Postoperative day (POD) 1: commence 20 mL/h water via jejunostomy feeding; POD 2: commence 20 mL/h jejunostomy feed enteral nutrition suspension; POD 3 to the date of discharge: gradually increase jejunostomy feed to the rate that meets the individual daily energy plan.

[b]CIMmH: comprehensive intervention model using the mobile health system.

[c]N/A: not applicable.

[d]ONS: oral nutrition supply.

[e]MBCR: mindfulness-based cancer recovery.

Nutrition Guidelines

Individual nutrition plans were developed by clinical nutritionists and cardiothoracic surgeons based on European Society of Clinical Nutrition and Metabolism guidelines (energy, 30 kcal/kg; protein, 1.5 g/kg ideal body weight) [28]. Patients and/or family caregivers were trained in jejunostomy feeding and used the WeChat platform at home. There were three different periods of nutrition support for patients with esophageal cancer who underwent surgery after discharge as follows: (1) home total enteral nutrition (TEN) on postoperative days (PODs) 8 to 21; (2) partial enteral nutrition (PEN) + oral nutrition supply (ONS) on PODs 22 to 42; and (3) ONS + oral intake on PODs 43 to 90. Detailed individual nutrition plans of each period were delivered by a nutritionist in a face-to-face meeting (offline) at POD 7, POD 21, and POD 42 in the hospital. In addition, three educational readings on general nutrition guidelines for each period (1000 words and 5 minutes of reading on average) and two instructional videos (7 minutes on average) on home-based enteral nutrition were sent to the participants or their caregivers via the WeChat (online) platform.

Physical Exercise

The physical exercise protocol consisted of inspiratory muscle training (chest mobilization exercise, flow-oriented incentive spirometry, deep breathing, and coughing exercise), walking exercise [29,30], and Baduanjin qigong [31], which is a mild form of muscular exercise from China involving eight movements. The participants were trained by rehabilitation therapists to perform inspiratory muscle exercises once before surgery lasting about 30 minutes (offline) and were also asked to use these techniques after surgery. Guidance on walking was individually tailored to the participants' fitness levels measured by the 6-minute walk distance (6MWD) before discharge

(around POD 7) and adjusted according to the 6MWD measured at 1 month after discharge.

Baduanjin qigong has been shown to have positive effects on patients with cancer, including alleviating sleep disturbances, strengthening immune function, and improving QOL [32,33]. A video of the Baduanjin qigong exercise was sent to the patients every day from 1 month to 3 months after surgery (online). Patients were encouraged to complete at least one cycle of about 15 minutes every day during this period.

Psychological Support

A psychological support program was adapted from mindfulness-based cancer recovery (MBCR) courses [34,35], which have demonstrated effectiveness in reducing stress and depressive symptoms in patients with cancer [36,37]. The adapted MBCR program consisted of four articles on meditation and coping with stress (400 words for each article and average reading time of 3 minutes) and 14 audio clips on meditation and stress reduction (10 minutes on average for each clip). The participants received the program 6 days a week for 2 months from 1 month to 3 months after surgery, with one rest day per week.

mHealth Intervention Development

The enhanced WeChat platform was developed by the research team, with three enhanced functions, including automatic intervention delivery, progress monitoring of patient engagement, and personalized feedback with community support.

Online Intervention Delivery

Through the enhanced WeChat platform, intervention materials were delivered to the participants and their family caregivers. In order to deliver the targeted intervention at different stages of recovery, the 3-month intervention was divided into the following five stages: (1) enrollment to presurgery; (2) after surgery and before discharge (POD 8); (3) after discharge until before removal of the nasogastric tube (POD 21); (4) before removal of the feeding tube (POD 42); and (5) after removal of the feeding tube until completion (POD 84). Researchers preset the stages for each patient on the online platform, and the corresponding intervention materials for each stage were automatically sent to the users.

Progress Monitoring

Patient engagement was tracked and monitored by the enhanced mHealth system, which showed whether the participants had switched on the program and the length of time they stayed on it. In addition, the participants were asked to report their nutrition intake, duration of walking, frequency of practicing Baduanjin qigong exercise, and mood every day on WeChat. The patients received instant and automatic feedback through WeChat and phone calls when needed, to discuss how they had completed the CIMmH program and whether their intake met the nutrition needs.

Support Community

An online support community was developed to offer social support through a chat feature. All participants were invited to join, and they were able to post their questions or comments on the WeChat group or via private chat to seek help or share their experiences. Researchers could also respond to messages instantly.

Data Collection and Measures

The participants were assessed at the following three time points: baseline (about 1 week before surgery) and 1 month and 3 months after surgery. The assessments were conducted in the hospital with the use of tablets and were assisted by trained research staff. Sociodemographic characteristics were collected, and they included age, gender, marital status, education, employment, and income. For measuring patient QOL, the European Organization for Research and Treatment of Cancer-Quality of life Question-Core (EORTC-QLQ-C30, version 3.0) and Oesophageal Cancer Module (EORTC-QLQ-OES-18) questionnaires [38] were used. EORTC-QLQ-C30 is a 30-item measure that includes an overall QOL scale, five functional scales (physical, role, emotional, cognitive, and social), three symptom subscales (fatigue, nausea & vomiting, and pain), and six single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties). The scores of EORTC-QLQ-C30 range from 0 to 100, with higher scores indicating a higher QOL. The Chinese version of the scale has been used in Chinese patients with cancer and has high validity and reliability [39].

EORTC-QLQ-OES-18 is a supplement of the disease-specific module for patients with esophageal cancer [40]. It consists of four symptom subscales (dysphagia, difficulty with eating, reflux, and pain) and six single items (trouble swallowing saliva, choking at swallowing, dry mouth, trouble with tasting, coughing, and talking). The Chinese version of the scale is reliable and acceptable to measure the health-related QOL of patients with esophageal cancer in China [41]. In addition to QOL, the patients' body weight and physical fitness measured by the 6MWD were assessed by medical doctors. Psychological status was measured using the Chinese versions of Patient Health Questionnaire-9 (PHQ-9) for depressive symptoms, General Anxiety Disorder-7 (GAD-7) for anxiety, and Perceived Stress Scale-10 (PSS-10) for stress. All these measures have high validity and reliability in Chinese populations [42-44].

Data Management and Statistical Analyses

Descriptive analyses of the sociodemographic characteristics and health outcomes were conducted. Means and SDs were used to describe normally distributed continuous variables, while medians and IQRs were used for continuous variables that were not normally distributed, and proportions were used for categorical variables. Pre-post comparisons of outcomes between baseline and the 1-month follow-up and between baseline and the 3-month follow-up were conducted. Paired Student *t* tests were used for normally distributed continuous variables, Wilcoxon signed-rank tests were used for nonnormally distributed variables, and chi-square tests were used for categorical variables. The rates of depressive symptoms and anxiety in the participants were calculated using the cutoff scores of at least 10 and 7 for PHQ-9 and GAD-7, respectively [45,46]. The analyses were performed using SAS version 9.4 (SAS Institute, Inc).

Results

Patient Engagement

In total, 95% (19/20) of the participants used the online program. Moreover, participants used the online program for an average of 71 minutes in total during the study period. Participants viewed on average 84% (3.38/4) of the online video intervention content and completed on average 14% (3.20/23) and 34% (9.44/28) of the online audio and article content, respectively. Participants completed on average 63% (5.01/8), 100.00% (1/1), and 24% (10.89/46) of the online nutrition, physical exercise, and psychological intervention content, respectively. There was no serious adverse event in any of the participants.

Patients' Characteristics

Participants' sociodemographic and clinical characteristics are summarized in Table 2. The mean age of the participants was 62.2 years (SD 7.1 years). The majority (18/20, 90%) of the participants were male, and more than half lived in rural regions (12/20, 60%) and did not complete high school (11/20, 55%). Most participants (13/20, 65%) had not received neoadjuvant chemotherapy or radiotherapy, and half (10/20, 50%) had a tumor in the middle thoracic area.

Table 2. Demographic and clinical characteristics of the study participants.

Characteristic	Participants (N=20), mean (SD) or n (%)
Age, years	62.20 (7.10)
Gender	
Male	18 (90)
Female	2 (10)
Region	
City	8 (40)
Rural	12 (60)
Educational status	
Less than high school	11 (55)
High school or greater	9 (45)
Marital status	
Married	20 (100)
Unmarried	0 (0)
Parenting	
Yes	19 (95)
No	1 (5)
Occupation	
Retired	12 (60)
Employed	8 (40)
Family monthly income, yuan (¥)[a]	
<3000	11 (55)
≥3000	9 (45)
Smoking	17 (85)
Drinking	9 (45)
Kungfu tea drinking	10 (50)
Regular exercise	10 (50)
Cancer history	9 (45)
Neoadjuvant chemotherapy/radiotherapy	
Not performed	13 (65)
Performed	7 (35)
Tumor location	
Upper thoracic area	4 (20)
Middle thoracic area	10 (50)
Lower thoracic area	6 (30)
Pathological stage	
I	4 (20)
II	6 (30)
III	9 (45)

[a]¥1 = US $0.14.

Outcomes of the CIMmH Program

Table 3 presents the scores of all outcome variables at baseline, the 1-month follow-up, and the 3-month follow-up.

Quality of Life

Overall QOL decreased significantly (P=.02) from baseline to the 1-month follow-up, and symptoms, including fatigue (P<.001), pain (P=.004), dyspnea (P<.001), difficulty with eating (P=.005), trouble with coughing (P=.02), trouble with talking (P<.001), and reflux (P=.04), were aggravated. From baseline to the 3-month follow-up, most of the QOL dimensions returned to the preoperative level, except for pain (P=.02), diarrhea (P=.04), difficulty with eating (P=.03), and trouble with talking (P=.003). Compared with baseline findings, the symptom of dry mouth was significantly alleviated at the 3-month follow-up (P=.04).

Nutrition Status

Participants' nutrition status worsened after esophagectomy. Analyses of pre-post changes showed a significant decrease in weight (P<.001) and BMI (P=.02) from baseline to the 1-month follow-up. Similarly, there was a significant decrease in weight (P<.001) and BMI (P=.02) from baseline to the 3-month follow-up.

6-Minute Walk Distance

There was no significant change in the 6MWD between baseline and the 1-month follow-up (P=.22) or between baseline and the 3-month follow-up (P=.52).

Psychological Outcomes

There was a significant increase in depressive symptoms from baseline to the 1-month follow-up (P<.001). From baseline to the 3-month follow-up, the change in depressive symptoms was not statistically significant (P=.08). There were also no significant changes in anxiety (from baseline to the 1-month follow-up: P=.48; from baseline to the 3-month follow-up: P=.59) and perceived stress levels (from baseline to the 1-month follow-up: P=.06; from baseline to the 3-month follow-up: P=.78) throughout the study. Based on the cutoff scores of the measures, 15% (3/20) of patients developed depressive symptoms 1 month after surgery, while 6% (1/16) still had depressive symptoms at the 3-month follow-up. With regard to anxiety, 20% (4/20) of patients had anxiety at baseline, 20% (4/20) had it at the 1-month follow-up, and 12% (2/16) still had it at the 3-month follow-up.

Table 3. Results of the outcome variables.

Outcome variables	Baseline score or value[a] (N=20)	1-month follow-up score or value[a] (N=20)	3-month follow-up score or value[a] (N=16)
Quality of life			
EORTC-QLQ-C30 [b]			
Overall quality of life scale[c]	76.70 (17.40)	65.40 (16.10)[d]	69.80 (12.10)
Functioning scale [e]			
Physical functioning	93.70 (12.30)	84.00 (17.30)	90.80 (8.20)
Role functioning	80.00 (29.80)	72.50 (23.70)	74.00 (20.60)
Emotional Functioning	86.70 (13.20)	76.70 (26.00)	85.40 (20.70)
Cognitive functioning	94.20 (11.80)	90.00 (18.10)	92.70 (11.40)
Social functioning	71.70 (25.80)	70.80 (22.90)	65.60 (25.90)
General symptom scale [f]			
Fatigue	10.00 (12.30)	36.10 (19.70)[g]	20.80 (23.10)
Nausea and vomiting scale	5.00 (11.60)	12.50 (22.90)	5.20 (15.90)
Pain	3.30 (6.50)	15.80 (18.20)[g]	12.50 (15.70)[d]
Dyspnea	3.30 (14.20)	25.00 (17.40)[g]	20.80 (27.70)
Insomnia	23.30 (32.70)	30.00 (32.40)	22.90 (35.60)
Appetite loss	6.70 (22.10)	20.00 (26.00)	20.80 (27.70)
Constipation	11.70 (21.30)	13.30 (21.60)	16.70 (28.00)
Diarrhea	6.70 (13.00)	13.30 (19.00)	12.50 (15.70)
Financial difficulties	36.70 (28.90)	28.30 (33.00)	29.20 (25.20)
EORTC-QLQ-OES18 [h]			
General functional scale [e]			
Dysphagia	68.30 (33.80)	62.80 (27.50)	68.10 (29.60)
General symptom scale [f]			
Trouble swallowing saliva	36.70 (44.70)	31.70 (39.10)	26.70 (37.60)
Choked when swallowing	16.70 (28.20)	23.30 (25.40)	28.90 (33.10)
Eating	6.70 (14.70)	25.80 (18.20)[g]	21.70 (16.40)[d]
Dry mouth	20.00 (28.00)	23.30 (25.40)	6.70 (12.90)[d]
Trouble with taste	5.00 (21.30)	11.70 (23.60)	2.20 (8.10)
Trouble with coughing	11.70 (18.60)	38.30 (33.00)[d]	15.60 (20.00)
Trouble talking	0.00 (0.00)	43.30 (29.30)[g]	17.80 (20.00)[g]
Reflux	5.00 (14.60)	23.30 (26.90)[d]	25.60 (31.70)
Pain	8.30 (14.10)	10.60 (12.10)	10.40 (13.30)
Nutrition status			
Weight (kg)	60.00 (8.70)	56.30 (7.80)[g]	55.00 (8.00)[g]
BMI (kg/m^2)	21.50 (3.30)	20.50 (2.60)[d]	20.00 (2.60)[d]
Physical fitness			
6MWD[i] (m)	506 (330.00-558.00)	469 (276.00-612.00)	486 (343.00-682.00)
6MWD change[j]	N/A[k]	0.95 (0.67-1.43)	1.03 (0.83-1.24)

Outcome variables	Baseline score or value[a] (N=20)	1-month follow-up score or value[a] (N=20)	3-month follow-up score or value[a] (N=16)
Psychological measures			
PHQ-9[l]	1.11 (1.33)	5.00 (4.61)[d]	2.81 (3.56)
GAD-7[m]	3.50 (4.16)	4.20 (4.54)	2.65 (3.52)
PSS-10[n]	10.30 (4.54)	12.60 (6.61)	10.65 (7.03)
Depressed, n (%)	0 (0)	3 (15)	1 (6)
Anxiety, n (%)	4 (20)	4 (20)	2 (12)

[a]Data are presented as mean (SD), n (%), or median (range).

[b]EORTC-QLQ-C30: European Organization for Research and Treatment of Cancer-Quality of life Question-Core-30.

[c]Higher scores indicate better health.

[d]$P<.05$.

[e]Higher scores indicate better function.

[f]Higher scores indicate worse symptoms.

[g]$P<.01$.

[h]EORTC-QLQ-OES-18: European Organization for Research and Treatment of Cancer-Quality of life Question-Oesophageal Cancer Module-18.

[i]6MWD: 6-minute walk distance.

[j]6MWD change was calculated using follow-up 6MWD values divided by baseline 6MWD values.

[k]N/A: not applicable.

[l]PHQ-9: Patient Health Questionnaire-9.

[m]GAD-7: General Anxiety Disorder-7.

[n]PSS-10: Perceived Stress Scale-10.

Qualitative Feedback

At the end of the study, participants were asked about their experiences of the intervention through the WeChat platform. Of the 20 participants, 18 (90%) reported that they were satisfied with the intervention program. Some suggested that the intervention interface and content could be designed in a more interesting and attractive way. One patient suggested that more intervention programs should be delivered in video format. Eight patients reported that they would like to receive more information about effective strategies to cope with postoperative complications.

Discussion

Principal Results

To the best of our knowledge, this prospective pilot study is the first attempt to develop and test the feasibility of an mHealth-based comprehensive intervention with nutrition, exercise, and psychological support through an online platform to help promote the ERAS program for patients with esophageal cancer.

For patients with cancer, ERAS is critical and challenging, as they experience both physical and mental complications. Tailor-made comprehensive interventions are needed for patients with specific cancers, as different types of cancers have different needs. Patients with esophageal cancer, for example, need special attention for nutrition intake and rehabilitation of respiratory movement after esophagectomy. The CIMmH is a comprehensive intervention addressing poor nutrition, physical inactivity, and intensified mental health symptoms within a single program for patients with esophageal cancer after surgery. The findings from this study indicate that the CIMmH is feasible and safe with no serious adverse effects for patients. The relevant decrease in overall QOL and increases in symptoms like fatigue, pain, dyspnea, difficulty with eating, trouble with coughing, trouble with talking, and reflux at the 1-month follow-up were expected, as patients were still in the recovery period after the surgery and were using feeding tubes. The results indicated that at the 3-month follow-up, except for pain, difficulty with eating, dry mouth, and trouble with talking, most of the QOL measures returned to the levels at the preoperative stage, indicating that recovery in these dimensions occurred 3 months after surgery. Compared with this study, previous studies that used traditional postoperative rehabilitation programs reported greater decreases in most functional dimensions of QOL and more serious deterioration of symptoms at 1 month and 3 months after surgery [47-49]. The CIMmH facilitated more comprehensive recovery for patients undergoing esophagectomy by restoring their declining functions faster and easing the symptoms caused by surgical injury. Nevertheless, patients still reported more problems with talking, coughing, and eating at 3 months after the surgery compared with the preoperative stage. These issues might stem from neurological injury during surgery, requiring a longer recovery time. These findings therefore highlight the need for routine evaluation and assessment of the postoperative functional status in patients. Careful consideration of the effects of possible complications on functional outcomes after surgery is thus needed when providing counseling services to patients before they make their decisions to undergo surgery and to plan rehabilitation.

As surgical injury often worsens the nutritional status of patients with esophagectomy, decreases in body weight and BMI are expected and have been well documented [8,50]. However, the drop in BMI was smaller in this study (1.0 and 1.5 at the 1- and 3-month follow-ups, respectively) than in a previous report (1.9 and 2.3 at the 1- and 3-month follow-ups, respectively) [50]. The minor drop in nutritional measures in this study indicates improved nutritional outcomes compared with the previous study [50].

Results of the 6MWD test at the 3-month follow-up demonstrated physical status comparable to that at baseline, indicating the effects of the CIMmH with regard to buffering the deterioration of physical fitness in patients after esophagectomy. Previous studies involving traditional postoperative rehabilitation showed a greater decrease in the 6MWD in the third month after surgery when compared with the finding in this study [51,52]. Lastly, with respect to psychological outcomes, depressive symptoms greatly increased at the 1-month follow-up, while anxiety and stress did not change greatly across all time points. Increased scores for depressive symptoms were expected as patients were greatly affected by surgical injuries at 1 month and were experiencing difficulties in eating, speaking, and even breathing in some cases. Nevertheless, the rates of depressive symptoms and anxiety in the present sample at 3 months after surgery were 6% (1/16) and 12% (2/16), respectively, whereas in a previous study, more than 40% of patients had depressive symptoms and anxiety after surgery [53], showing the potential of the CIMmH to support patients in coping with their conditions.

In this study, the mHealth system yielded a unique opportunity to provide much needed postsurgical care for the included patients. The functions of automatic monitoring, timely interventions, and feedback through the online mHealth system helped improve intervention adherence, as 80% (16/20) of participants completed the intervention. Although no formal qualitative data were collected in this study, some participants reported that they liked the video talks given by the medical doctors on how to take care of themselves and found the information on nutrition, exercise, and symptom management useful and helpful. In addition, professionals in the hospital reflected that the comprehensive intervention model of combining online (mHealth) and offline (face-to-face) services in this study was cost-effective and easier to incorporate into existing clinical practice and health care services, as less professional time was required and patients could receive tailor-made and timely interventions at home [54]. Given the large number of patients with esophageal cancer in low- and middle-income countries, the CIMmH has the potential to be a feasible cost-effective therapeutic option for improving postsurgery recovery in patients with esophageal cancer after esophagectomy.

Patient engagement data indicated that online intervention content in video format was more popular than audio or written materials. One possible reason might be that most participants in this study were elderly people who might have found it easier to understand vivid videos compared with audio content and articles. Future interventions may consider using more intervention materials in video format. Moreover, the completed proportions of the online nutrition and exercise intervention content were much higher than the completed proportion of the psychological intervention content. One possible explanation could be that the mental health status of the participants at baseline was better than that during follow-ups, so it was very likely that participants paid more attention to coping with postoperative complications than mental health–related issues. Another reason might be that there was insufficient emphasis on the importance of mental health at the beginning of the program. Mental health is an important problem in patients with esophageal cancer, as many of these patients experience depressive symptoms and anxiety after surgery [53,55]. Future interventions should emphasize the importance of mental health issues and educate patients to pay attention to their mental wellness after surgery.

Implications

The CIMmH has yielded a unique opportunity to provide much needed postsurgical care in patients with esophageal cancer for the likely improvement of postoperative nutrition and the physical and mental status. This pilot study has shown that the CIMmH approach is a feasible and well-received option for ERAS in patients with esophageal cancer. Experiences of the CIMmH pilot study may help future development of a large randomized controlled trial or similar programs for patients with esophageal cancer or other cancers, especially those who are not able to visit the hospital frequently or who reside in rural areas. For example, patient adherence to the program needs to be enhanced in future interventions for better treatment effects, especially in the component of psychological intervention. More intervention programs should be delivered in video format, as video talks by medical doctors on patient self-care are particularly well received by patients.

Limitations

Despite the positive outcomes, there were several limitations in this study. First, the sample size was small and there was no comparison group; thus, caution is needed to avoid over interpreting the findings. Second, some outcome data were missing owing to the drop-out of several patients at follow-up assessments. Third, effects of the CIMmH might be influenced by adherence and complications after surgery, which differ from patient to patient. Future studies should adopt a larger sample size and preferably use a randomized controlled trial design. The subjective experience of the participants should also be explored by collecting qualitative feedback throughout the study.

Conclusions

The CIMmH is the first mHealth-based comprehensive intervention developed and tested in patients with esophageal cancer after esophagectomy. Our results show that the CIMmH is usable, feasible, and safe among patients with esophageal cancer after surgery in China. Future studies with a more rigorous design and larger samples are needed to establish efficacy in patients with esophageal cancer and those with other types of cancers.

Acknowledgments

The authors would like to thank the patients, their family members, and the research staff at the hospital and research sites involved.

Conflicts of Interest

None declared.

References

1. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66(2):115-132 [FREE Full text] [doi: 10.3322/caac.21338] [Medline: 26808342]
2. Stahl M, Mariette C, Haustermans K, Cervantes A, Arnold D, ESMO Guidelines Working Group. Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2013 Oct;24 Suppl 6:vi51-vi56 [FREE Full text] [doi: 10.1093/annonc/mdt342] [Medline: 24078662]
3. Daly JM, Fry WA, Little AG, Winchester DP, McKee RF, Stewart AK, et al. Esophageal cancer: results of an American College of Surgeons Patient Care Evaluation Study. J Am Coll Surg 2000 May;190(5):562-72; discussion 572. [doi: 10.1016/s1072-7515(00)00238-6] [Medline: 10801023]
4. Viklund P, Lindblad M, Lu M, Ye W, Johansson J, Lagergren J. Risk factors for complications after esophageal cancer resection: a prospective population-based study in Sweden. Ann Surg 2006 Feb;243(2):204-211. [doi: 10.1097/01.sla.0000197698.17794.eb] [Medline: 16432353]
5. Seesing MF, Gisbertz SS, Goense L, van Hillegersberg R, Kroon HM, Lagarde SM, et al. A Propensity Score Matched Analysis of Open Versus Minimally Invasive Transthoracic Esophagectomy in the Netherlands. Ann Surg 2017 Nov;266(5):839-846. [doi: 10.1097/SLA.0000000000002393] [Medline: 28796016]
6. Kassis ES, Kosinski AS, Ross P, Koppes KE, Donahue JM, Daniel VC. Predictors of anastomotic leak after esophagectomy: an analysis of the society of thoracic surgeons general thoracic database. Ann Thorac Surg 2013 Dec;96(6):1919-1926. [doi: 10.1016/j.athoracsur.2013.07.119] [Medline: 24075499]
7. Jacobs M, Macefield RC, Elbers RG, Sitnikova K, Korfage IJ, Smets EM, et al. Meta-analysis shows clinically relevant and long-lasting deterioration in health-related quality of life after esophageal cancer surgery. Qual Life Res 2014 May;23(4):1155-1176. [doi: 10.1007/s11136-013-0576-5] [Medline: 24293086]
8. Baker M, Halliday V, Williams RN, Bowrey DJ. A systematic review of the nutritional consequences of esophagectomy. Clin Nutr 2016 Oct;35(5):987-994 [FREE Full text] [doi: 10.1016/j.clnu.2015.08.010] [Medline: 26411750]
9. Hellstadius Y, Lagergren J, Zylstra J, Gossage J, Davies A, Hultman CM, et al. A longitudinal assessment of psychological distress after oesophageal cancer surgery. Acta Oncol 2017 May;56(5):746-752. [doi: 10.1080/0284186X.2017.1287945] [Medline: 28447567]
10. Markar S, Gronnier C, Duhamel A, Mabrut J, Bail J, Carrere N, FREGAT (French Eso-Gastric Tumors) working group, FRENCH (Fédération de Recherche EN CHirurgie),AFC (Association Française de Chirurgie). The Impact of Severe Anastomotic Leak on Long-term Survival and Cancer Recurrence After Surgical Resection for Esophageal Malignancy. Ann Surg 2015 Dec;262(6):972-980. [doi: 10.1097/SLA.0000000000001011] [Medline: 26469952]
11. Ljungqvist O, Scott M, Fearon KC. Enhanced Recovery After Surgery: A Review. JAMA Surg 2017 Mar 01;152(3):292-298. [doi: 10.1001/jamasurg.2016.4952] [Medline: 28097305]
12. Capozzi LC, McNeely ML, Lau HY, Reimer RA, Giese-Davis J, Fung TS, et al. Patient-reported outcomes, body composition, and nutrition status in patients with head and neck cancer: Results from an exploratory randomized controlled exercise trial. Cancer 2016 Apr 15;122(8):1185-1200 [FREE Full text] [doi: 10.1002/cncr.29863] [Medline: 26828426]
13. Kwiatkowski F, Mouret-Reynier M, Duclos M, Bridon F, Hanh T, Van Praagh-Doreau I, et al. Long-term improvement of breast cancer survivors' quality of life by a 2-week group physical and educational intervention: 5-year update of the 'PACThe' trial. Br J Cancer 2017 May 23;116(11):1389-1393 [FREE Full text] [doi: 10.1038/bjc.2017.112] [Medline: 28427084]
14. Takesue T, Takeuchi H, Ogura M, Fukuda K, Nakamura R, Takahashi T, et al. A Prospective Randomized Trial of Enteral Nutrition After Thoracoscopic Esophagectomy for Esophageal Cancer. Ann Surg Oncol 2015 Dec;22 Suppl 3:S802-S809. [doi: 10.1245/s10434-015-4767-x] [Medline: 26219242]
15. van Vulpen JK, Siersema PD, van Hillegersberg R, Nieuwenhuijzen GA, Kouwenhoven EA, Groenendijk RP, et al. Physical ExeRcise Following Esophageal Cancer Treatment (PERFECT) study: design of a randomized controlled trial. BMC Cancer 2017 Aug 18;17(1):552 [FREE Full text] [doi: 10.1186/s12885-017-3542-8] [Medline: 28821284]
16. Inoue T, Ito S, Ando M, Nagaya M, Aso H, Mizuno Y, et al. Changes in exercise capacity, muscle strength, and health-related quality of life in esophageal cancer patients undergoing esophagectomy. BMC Sports Sci Med Rehabil 2016;8:34 [FREE Full text] [doi: 10.1186/s13102-016-0060-y] [Medline: 27822378]
17. Lawn S, Zrim S, Leggett S, Miller M, Woodman R, Jones L, et al. Is self-management feasible and acceptable for addressing nutrition and physical activity needs of cancer survivors? Health Expect 2015 Dec;18(6):3358-3373 [FREE Full text] [doi: 10.1111/hex.12327] [Medline: 25545411]

18. Ren JY, Zhong JD, Yuan J, Zhang JE, Li CZ, Wei WJ. Unmet supportive care needs and associated factors among Chinese discharged patients with esophageal cancer after esophagectomy: A cross-sectional study. Eur J Oncol Nurs 2020 May 21;46:101767. [doi: 10.1016/j.ejon.2020.101767] [Medline: 32504878]
19. Henselmans I, Jacobs M, van Berge Henegouwen MI, de Haes HC, Sprangers MA, Smets EM. Postoperative information needs and communication barriers of esophageal cancer patients. Patient Educ Couns 2012 Jul;88(1):138-146 [FREE Full text] [doi: 10.1016/j.pec.2011.12.004] [Medline: 22244819]
20. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
21. Mobasheri MH, Johnston M, King D, Leff D, Thiruchelvam P, Darzi A. Smartphone breast applications - what's the evidence? Breast 2014 Oct;23(5):683-689 [FREE Full text] [doi: 10.1016/j.breast.2014.07.006] [Medline: 25153432]
22. Kelli HM, Witbrodt B, Shah A. THE FUTURE OF MOBILE HEALTH APPLICATIONS AND DEVICES IN CARDIOVASCULAR HEALTH. Euro Med J Innov 2017 Jan;2017:92-97 [FREE Full text] [Medline: 28191545]
23. Haggerty AF, Hagemann A, Barnett M, Thornquist M, Neuhouser ML, Horowitz N, et al. A Randomized, Controlled, Multicenter Study of Technology-Based Weight Loss Interventions among Endometrial Cancer Survivors. Obesity (Silver Spring) 2017 Nov;25 Suppl 2:S102-S108 [FREE Full text] [doi: 10.1002/oby.22021] [Medline: 29086522]
24. Galiano-Castillo N, Cantarero-Villanueva I, Fernández-Lao C, Ariza-García A, Díaz-Rodríguez L, Del-Moral-Ávila R, et al. Telehealth system: A randomized controlled trial evaluating the impact of an internet-based exercise intervention on quality of life, pain, muscle strength, and fatigue in breast cancer survivors. Cancer 2016 Oct 15;122(20):3166-3174 [FREE Full text] [doi: 10.1002/cncr.30172] [Medline: 27332968]
25. Porter LS, Keefe FJ, Garst J, Baucom DH, McBride CM, McKee DC, et al. Caregiver-assisted coping skills training for lung cancer: results of a randomized clinical trial. J Pain Symptom Manage 2011 Jan;41(1):1-13 [FREE Full text] [doi: 10.1016/j.jpainsymman.2010.04.014] [Medline: 20832982]
26. Smith C. Amazing WeChat Statistics and Facts (2020). DMR. URL: https://expandedramblings.com/index.php/wechat-statistics/ [accessed 2020-03-06]
27. Zhou J, Chen H, Lu JJ, Xiang J, Zhang Y, Hu H, et al. Application of a modified McKeown procedure (thoracoscopic esophageal mobilization three-incision esophagectomy) in esophageal cancer surgery: initial experience with 30 cases. Dis Esophagus 2009;22(8):687-693. [doi: 10.1111/j.1442-2050.2009.00963.x] [Medline: 19392854]
28. Weimann A, Braga M, Carli F, Higashiguchi T, Hübner M, Klek S, et al. ESPEN guideline: Clinical nutrition in surgery. Clin Nutr 2017 Jun;36(3):623-650. [doi: 10.1016/j.clnu.2017.02.013] [Medline: 28385477]
29. Park S, Lee J, Min K. Comparison of the effects of core stabilization and chest mobilization exercises on lung function and chest wall expansion in stroke patients. J Phys Ther Sci 2017 Jul;29(7):1144-1147 [FREE Full text] [doi: 10.1589/jpts.29.1144] [Medline: 28744034]
30. Lunardi AC, Porras DC, Barbosa RC, Paisani DM, Marques da Silva CC, Tanaka C, et al. Effect of volume-oriented versus flow-oriented incentive spirometry on chest wall volumes, inspiratory muscle activity, and thoracoabdominal synchrony in the elderly. Respir Care 2014 Mar;59(3):420-426 [FREE Full text] [doi: 10.4187/respcare.02665] [Medline: 23983269]
31. Zou L, Yeung A, Quan X, Boyden SD, Wang H. A Systematic Review and Meta-Analysis of Mindfulness-Based (Baduanjin) Exercise for Alleviating Musculoskeletal Pain and Improving Sleep Quality in People with Chronic Diseases. Int J Environ Res Public Health 2018 Jan 25;15(2):206 [FREE Full text] [doi: 10.3390/ijerph15020206] [Medline: 29370149]
32. Zeng Y, Luo T, Xie H, Huang M, Cheng AS. Health benefits of qigong or tai chi for cancer patients: a systematic review and meta-analyses. Complement Ther Med 2014 Feb;22(1):173-186. [doi: 10.1016/j.ctim.2013.11.010] [Medline: 24559833]
33. Chen M, Liu H, Huang H, Chiou A. The effect of a simple traditional exercise programme (Baduanjin exercise) on sleep quality of older adults: a randomized controlled trial. Int J Nurs Stud 2012 Mar;49(3):265-273. [doi: 10.1016/j.ijnurstu.2011.09.009] [Medline: 21963235]
34. Carlson L, Speca M. Mindfulness‐Based Cancer Recovery: A Step‐By‐Step MBSR Approach to Help You Cope With Treatment and Reclaim Your Life. California, USA: New Harbinger Publications; 2010.
35. Chan JS, Li A, Ng S, Ho RT, Xu A, Yao T, et al. Adiponectin Potentially Contributes to the Antidepressive Effects of Baduanjin Qigong Exercise in Women With Chronic Fatigue Syndrome-Like Illness. Cell Transplant 2017 Mar 13;26(3):493-501 [FREE Full text] [doi: 10.3727/096368916X694238] [Medline: 27938498]
36. Zou L, SasaKi JE, Wang H, Xiao Z, Fang Q, Zhang M. A Systematic Review and Meta-Analysis Baduanjin Qigong for Health Benefits: Randomized Controlled Trials. Evid Based Complement Alternat Med 2017;2017:4548706 [FREE Full text] [doi: 10.1155/2017/4548706] [Medline: 28367223]
37. Schell LK, Monsef I, Wöckel A, Skoetz N. Mindfulness-based stress reduction for women diagnosed with breast cancer. Cochrane Database Syst Rev 2019 Mar 27;3:CD011518 [FREE Full text] [doi: 10.1002/14651858.CD011518.pub2] [Medline: 30916356]
38. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 03;85(5):365-376. [doi: 10.1093/jnci/85.5.365] [Medline: 8433390]

39. Zhao H, Kanda K. Testing psychometric properties of the standard Chinese version of the European Organization for Research and Treatment of Cancer Quality of Life Core Questionnaire 30 (EORTC QLQ-C30). J Epidemiol 2004 Nov;14(6):193-203 [FREE Full text] [doi: 10.2188/jea.14.193] [Medline: 15617393]
40. Blazeby JM, Conroy T, Hammerlid E, Fayers P, Sezer O, Koller M, European Organisation for ResearchTreatement of Cancer GastrointestinalQuality of Life Groups. Clinical and psychometric validation of an EORTC questionnaire module, the EORTC QLQ-OES18, to assess quality of life in patients with oesophageal cancer. Eur J Cancer 2003 Jul;39(10):1384-1394. [doi: 10.1016/s0959-8049(03)00270-3] [Medline: 12826041]
41. Dai Z, Lang W, Yang H, Tian J, Sun W, Pekbay B, et al. Validation of EORTC QLQ-OES18 for Chinese patients with esophageal cancer. Dis Esophagus 2017 Oct 01;30(10):1-7. [doi: 10.1093/dote/dox046] [Medline: 28859386]
42. Xia N, Lin J, Ding S, Dong F, Shen J, Du Y, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire 9 (C-PHQ-9) in patients with epilepsy. Epilepsy Behav 2019 Jun;95:65-69. [doi: 10.1016/j.yebeh.2019.03.049] [Medline: 31026785]
43. Garabiles MR, Lao CK, Yip P, Chan EW, Mordeno I, Hall BJ. Psychometric Validation of PHQ-9 and GAD-7 in Filipino Migrant Domestic Workers in Macao (SAR), China. J Pers Assess 2019 Jul 30:1-12. [doi: 10.1080/00223891.2019.1644343] [Medline: 31361153]
44. Sun Y, Gao L, Kan Y, Shi B. The Perceived Stress Scale-10 (PSS-10) is reliable and has construct validity in Chinese patients with systemic lupus erythematosus. Lupus 2019 Feb;28(2):149-155. [doi: 10.1177/0961203318815595] [Medline: 30518288]
45. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 2012 Feb 21;184(3):E191-E196 [FREE Full text] [doi: 10.1503/cmaj.110829] [Medline: 22184363]
46. Esser P, Hartung TJ, Friedrich M, Johansen C, Wittchen H, Faller H, et al. The Generalized Anxiety Disorder Screener (GAD-7) and the anxiety module of the Hospital and Depression Scale (HADS-A) as screening tools for generalized anxiety disorder among cancer patients. Psychooncology 2018 Jun;27(6):1509-1516. [doi: 10.1002/pon.4681] [Medline: 29473255]
47. Wang H, Feng M, Tan L, Wang Q. Comparison of the short-term quality of life in patients with esophageal cancer after subtotal esophagectomy via video-assisted thoracoscopic or open surgery. Dis Esophagus 2010 Jul;23(5):408-414. [doi: 10.1111/j.1442-2050.2009.01025.x] [Medline: 19930404]
48. Nafteux P, Moons J, Coosemans W, Decaluwé H, Decker G, De Leyn P, et al. Minimally invasive oesophagectomy: a valuable alternative to open oesophagectomy for the treatment of early oesophageal and gastro-oesophageal junction carcinoma. Eur J Cardiothorac Surg 2011 Dec;40(6):1455-63; discussion 1463. [doi: 10.1016/j.ejcts.2011.01.086] [Medline: 21514837]
49. Barbour AP, Cormack OM, Baker PJ, Hirst J, Krause L, Brosda S, et al. Long-term Health-related Quality of Life Following Esophagectomy: A Nonrandomized Comparison of Thoracoscopically Assisted and Open Surgery. Ann Surg 2017 Jun;265(6):1158-1165. [doi: 10.1097/SLA.0000000000001899] [Medline: 27429022]
50. Koizumi M, Hosoya Y, Dezaki K, Yada T, Hosoda H, Kangawa K, et al. Postoperative weight loss does not resolve after esophagectomy despite normal serum ghrelin levels. Ann Thorac Surg 2011 Apr;91(4):1032-1037. [doi: 10.1016/j.athoracsur.2010.11.072] [Medline: 21440118]
51. Otani T, Ichikawa H, Hanyu T, Ishikawa T, Kano Y, Kanda T, et al. Long-Term Trends in Respiratory Function After Esophagectomy for Esophageal Cancer. J Surg Res 2020 Jan;245:168-178. [doi: 10.1016/j.jss.2019.07.040] [Medline: 31421359]
52. Xu Y, Cheng JC, Lee J, Huang P, Huang G, Chen CC. A Walk-and-Eat Intervention Improves Outcomes for Patients With Esophageal Cancer Undergoing Neoadjuvant Chemoradiotherapy. Oncologist 2015 Oct;20(10):1216-1222 [FREE Full text] [doi: 10.1634/theoncologist.2015-0178] [Medline: 26341759]
53. Bouras G, Markar SR, Burns EM, Huddy JR, Bottle A, Athanasiou T, et al. The psychological impact of symptoms related to esophagogastric cancer resection presenting in primary care: A national linked database study. Eur J Surg Oncol 2017 Feb;43(2):454-460. [doi: 10.1016/j.ejso.2016.10.010] [Medline: 27919514]
54. Kelli HM, Witbrodt B, Shah A. The future of mobile health applications and devices in cardiovascular health. Euro Med J Innov 2017 Jan;2017:92-97 [FREE Full text] [Medline: 28191545]
55. Pinto E, Cavallin F, Scarpa M. Psychological support of esophageal cancer patient? J Thorac Dis 2019 Apr;11(Suppl 5):S654-S662 [FREE Full text] [doi: 10.21037/jtd.2019.02.34] [Medline: 31080642]

Abbreviations

6MWD: 6-minute walk distance
CIMmH: comprehensive intervention model using the mobile health system
EORTC-QLQ-C30: European Organization for Research and Treatment of Cancer-Quality of life Question-Core-30
EORTC-QLQ-OES-18: European Organization for Research and Treatment of Cancer-Quality of life Question-Oesophageal Cancer Module-18
ERAS: enhanced recovery after surgery

GAD-7: General Anxiety Disorder-7
MBCR: mindfulness-based cancer recovery
mHealth: mobile health
ONS: oral nutrition supply
PEN: partial enteral nutrition
PHQ-9: Patient Health Questionnaire-9
POD: postoperative day
PSS-10: Perceived Stress Scale-10
QOL: quality of life
TEN: total enteral nutrition

Edited by G Eysenbach; submitted 29.03.20; peer-reviewed by YA Hong, M Khazaee-Pool; comments to author 12.06.20; revised version received 14.06.20; accepted 25.06.20; published 18.08.20.

Please cite as:
Cheng C, Ho RTH, Guo Y, Zhu M, Yang W, Li Y, Liu Z, Zhuo S, Liang Q, Chen Z, Zeng Y, Yang J, Zhang Z, Zhang X, Monroe-Wise A, Yeung SC
Development and Feasibility of a Mobile Health–Supported Comprehensive Intervention Model (CIMmH) for Improving the Quality of Life of Patients With Esophageal Cancer After Esophagectomy: Prospective, Single-Arm, Nonrandomized Pilot Study
J Med Internet Res 2020;22(8):e18946
URL: http://www.jmir.org/2020/8/e18946/
doi:10.2196/18946
PMID:32808933

Original Paper

Food Communication and its Related Sentiment in Local and Organic Food Videos on YouTube

Xanat Vargas Meza[1], MA, PhD; Toshimasa Yamanaka[2], PhD

[1]Department of International Industrial Engineering, National Institute of Technology Ibaraki College, Hitachinaka, Japan

[2]Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Japan

Corresponding Author:
Toshimasa Yamanaka, PhD
Graduate School of Comprehensive Human Sciences
University of Tsukuba
Laboratory of Advanced Research D
1-1-1 Tennodai
Tsukuba, 305-8577
Japan
Phone: 81 298532999
Email: tyam@geijutsu.tsukuba.ac.jp

Abstract

Background: Local and organic foods have shown increased importance and market size in recent years. However, attitudes, sentiment, and habits related to such foods in the context of video social networks have not been thoroughly researched. Given that such media have become some of the most important venues of internet traffic, it is relevant to investigate how sustainable food is communicated through such video social networks.

Objective: This study aimed to explore the diffusion paths of local and organic foods on YouTube, providing a review of trends, coincidences, and differences among video discourses.

Methods: A combined methodology involving webometric, framing, semantic, and sentiment analyses was employed.

Results: We reported the results for the following two groups: organic and local organic videos. Although the content of 923 videos mostly included the “Good Mother” (organic and local organic: 282/808, 34.9% and 311/866, 35.9%, respectively), “Natural Goodness” (220/808, 27.2% and 253/866, 29.2%), and “Undermining of Foundations” (153/808, 18.9% and 180/866, 20.7%) frames, organic videos were more framed in terms of “Frankenstein” food (organic and local organic: 68/808, 8.4% and 27/866, 3.1%, respectively), with genetically modified organisms being a frequent topic among the comments. Organic videos (N=448) were better connected in terms of network metrics than local organic videos (N=475), which were slightly more framed regarding “Responsibility” (organic and local organic: 42/808, 5.1% and 57/866, 6.5%, respectively) and expressed more positive sentiment (M ranks for organic and local organic were 521.2 and 564.54, respectively, Z=2.15, *P*=.03).

Conclusions: The results suggest that viewers considered sustainable food as part of a complex system and in a positive light and that food framed as artificial and dangerous sometimes functions as a counterpoint to promote organic food.

(J Med Internet Res 2020;22(8):e16761) doi:10.2196/16761

KEYWORDS

social networks; framing; semantic analysis; sentiment analysis; organic; local; food; YouTube

Introduction

Background

Sustainability has been receiving global attention with the United Nations promotion of Sustainable Development Goals. Goal 2 “Zero Hunger” involves food security and agriculture development [1]. In these regards, organic and local foods have been proposed as alternatives to industrialized patterns of food production, consumption, and distribution, with the potential to overcome the structural issues of mainstream sustainable consumption practices [2]. However, sustainable food markets still need to address a lack of information on resource efficiency and eating habits [3].

The global organic food market size was estimated at US $90 billion in 2016, with the United States, Germany, and France being the main consumers and India, Uganda, and Mexico being

the main producers [4]. Lee and Yun [5] provided a model to explain organic food consumer behavior based on stimuli (nutrition, nature, ecological welfare, sensory appeal, and price), organism (utilitarian and hedonic attitudes), and response (buying intention). In the European context, animal welfare, regional production, and fair price were important in purchase decisions [6].

On the other hand, the marketing of food products showed a change in the 2000s from industrialized processes to artisanal, small, and locally based processes [7]. In 2007, the term "locavores" was used for those who eat food produced within a 100-mile radius [8]. Darby et al [9] found that a consumer's will to pay for local foods is independent from values associated with product freshness and farm size. Additionally, local food is linked to authenticity in tourism contexts [10]. These findings suggest that "organic" and "local" are conceptually different enough to be distinguishable for consumers, although the terms frequently overlap in food descriptions. However, it is important to understand consumers' definitions, attitudes, and behaviors related to sustainable food in the context of modern technologies, such as social networking sites.

There is evidence that sensationalist and erroneous content is being fueled by social media search engines owing to their business focus [11,12]. How this impacts the communication of sustainable food products is largely unknown. Thus, this study aimed to explore organic and local food products on YouTube, with the following objectives:

- To find diffusion paths of local and organic food products on YouTube by collecting information on related videos and comparing their network levels with social network analysis.
- To review trends and differences among discourses through framing analysis on video content.
- To explore the opinions, attitudes, behaviors, and emotions expressed by viewers through semantic and sentiment analyses on comments extracted from the videos.

Literature Review

With the advent of internet 2.0, online collaboration and activism increased, transforming the internet into a conversational space through the rise of social networking sites. Internet 3.0 incorporated location and real-time aware devices and apps, prompting more personalization of products and services.

Social networks can reduce the communication gap between producers, consumers, and other interested people. For example, business engagement on Twitter is related to consumers' web-based word-of-mouth communication, and its influence reaches consumers with a second-degree relationship to brands [13]. However, Rutsaert et al [14] argued that social media can contain inaccurate, incorrect, or misleading information and that there is a delicate balance between fast communication in a food crisis and amplification of risk perceptions.

Previous social media studies on food communication through text included two-way communication by public organizations related to food safety and nutrition, and it was found that the main themes were queries and complaints, benefits of social media in query and complaint services, content redesign driven by social media use, and social media to learn about consumers [15]. Pantelidis [16] examined restaurant reviews, arguing that costumers considered food, service, ambience, price, menu, and décor regardless of their economic conditions. Moreover, direct-to-consumer marketing strategies of farmers on Facebook were positively related to their social capital and farm revenue [17].

Structural and social factors of web-based communication channels affect their roles in the image construction process of organic food brands [18]. Studies on organic food communication and eating habits expressed by Koreans and Mexicans on Twitter revealed that (1) Mexicans focused on basic food products in street markets; (2) Koreans highlighted promotions and firms related to processed products; and (3) both countries showed family orientation/domesticity and sentiment/emotion [19,20].

Comments from Mexico's Starbucks Facebook page (a shop chain that sells organic coffee) reflected that people interacted more through happiness, but anger and longing were often used to generate participation [21]. Predicted effects of comments and likes were found when comments in Facebook posts related to organic food were perceived as useful, and the number of likes had an effect on negative emotions and willingness to pay [22]. A study involving organic local food suggested that such products can gain popularity and overcome purchasing barriers (eg, high price, low awareness, and low availability) through integration in consumers' daily experiences with Facebook [23].

However, video networks related to organic or local food have not been thoroughly researched. Video social networks account for over half of the internet traffic when measured in bytes [24], and YouTube is an integrated video sharing platform that tends to form small-world networks [25]. An analysis of 76 food safety–related videos posted on YouTube found that artful quality was important to attract viewers and that the intended purpose of the videos can differ from the viewers' perception of them [26]. Moreover, a study in Europe argued that YouTube was the least used media for consultation in case of food risks, although its general usage was high, particularly among younger generations [27]. In summary, communities related to sociotechnical food systems can be observed through video social networks in order to understand their communication patterns and discourses.

Methods

Operational Definitions

There is no consensus on what is organic or local food. A definition adapted from previous reports [28,29], establishes that any foods produced, processed, and managed through methods that do not implicate harmful synthetic inputs or additives, irradiation, or genetically modified organisms in compliance with standards generally set by the country in which they are sold, are considered organic.

With regard to local food, the most recognized feature is the marketing arrangement [30]. Thus, local food can be defined as food sold by producers to consumers and food sold by producers directly to retailers at markets or other regional areas..

Data Collection

According to applicable institutional and national guidelines and regulations, ethics approval was not required for this study, as we focused on publicly available YouTube data. Video data were extracted in 2015 with YouTube Data Tools [31], which use the YouTube Data application program interface (API). The following two modules were employed: Video Network and Video Info. The first module builds a network starting with a query (one or more keywords), thus retrieving related videos and their metrics (number of views, likes, dislikes, comments, and categories). It also creates a graph file readable through the Gephi software [32]. As for the second module, video identifiers obtained through the first module were used to extract video comments.

Classification of Data

The following three queries were used to extract video data: "organic food," "local food," and "local organic food." The resulting three files were appended into one. Videos that appeared only with the "organic food" query were labelled as "organic food videos." There were no videos that appeared exclusively with the "local food" query; thus, all videos that were not labelled as "organic food videos" were labelled as "local organic food videos." Videos were watched by an investigator, and in case the content was not in an understandable language, the investigator requested a native speaker to interpret the content. Content that clearly was unrelated to food was discarded, reducing a total of 964 videos to 923 videos. The videos were also classified according to country and uploader. The types of uploaders considered were as follows:

- Business: It included businesses related to food production, processing, and distribution. Businesses linked to health, tourism, and banking were added to this category as well.
- Community: It included citizens and communities.
- Education: It included citizens disclosed as professors, researchers, students, and lecturers; research institutions; universities; and informal education-related accounts.
- Media: It included both traditional and internet-based media.
- Others: It included government, politicians, and celebrities.
- Undisclosed: It included all accounts that did not fit in the previous categories.

Framing Analysis

Video content was also classified according to food frames. Framing is the action of using images and words to influence how audiences interpret a message, promoting specific versions of reality. They are useful tools to infer what people think is important.

Food-related framing usually falls into the corporative-political category [33] or into the social activism category [34]. Nonetheless, this study employed van Gorp and van der Goot [35] framing categories, which are used by stakeholders of sustainable food products and are based on archetypes (Table 1). The "Natural Goodness" frame and the "Good Mother" frame are used frequently and are strongly interlinked [35]. Therefore, most of the videos were classified in two or more frame categories, taking into account discursive/textual and visual cues. Only videos that reflected at least one frame were employed in further analysis.

Table 1. Condensed frame packages for sustainable food products.

Frame	Emotional basis	Key concepts	Visual cues	Textual cues
Responsibility	Endearment	Accountability and vulnerability	Children, fragile plants, and young animals	Caring, future generations, our children, and vulnerability
Undermining of foundations	Alarm and concern	Balance, base, complex systems, and links	Interconnections between elements in the ecosystem	Fragile balance, mutual dependency, and unstable
Frankenstein	Anxiety and unscrupulousness	Apocalypse, Pandora's box, and sorcerer's apprentice	Monsters and skulls	Frankenstein food, poison, and risks
Natural goodness	Admiration and astonishment	Authenticity, good taste, health, and purity	Idyllic nature and products	Natural, pure, and taste
Progress	Trust	Modernization and progress	High-tech tools	A better world, constant striving, and technology
Good mother	Gratitude, enjoyment, and love	Freedom of choice and great variety of products	Pleasure of shopping and rich harvest	Friendliness and product range

Social Network Analysis

This type of analysis involves theories, methods, and techniques to study social relations and their structures [36]. A social network can be defined as a group of nodes and ties, where a node represents an entity (a YouTube video in this case) and a tie represents a relationship. In this study, ties connect nodes to related videos. Social network measurements were made with the Gephi software package.

Sampling and Cleaning of Comments

Based on the study by Tsou et al [37], a random sample of comments made on "average" videos was considered to conduct semantic and sentiment analyses. In other words, comments from videos that had a maximum of 500 and a minimum of 5 comments were sampled. Thus, a total of 213 videos (155 from the organic network and 58 from the local organic network) were selected for sampling. The number of total extracted comments for each video was identified and included in a

random number generator [38]. The outcome was five numbers that were used to select the comments, resulting in a total of 1065 comments. If a comment was not in English, the random number generator was employed again until a comment in English was selected. Moreover, text belonging to 13 comments not retrieved completely by the YouTube API was added by hand. URLs (web addresses) were shortened, as they were not the focus of this study. Information on replies to Google+ and Twitter accounts was also omitted.

Semantic Analysis

This is a type of network study where the unit of analysis is keywords. The software employed was TI, an open source tool that generates a word frequency list, a word-occurrence matrix, a word co-occurrence matrix, a normalized co-occurrence matrix, and a word list from a set of short texts [39]. In order to cope with semantic ambiguities, coding notes were employed and refined over multiple test rounds to adjust the data, especially in the case of frequently found synonyms and plurals. Because TI does not recognize emoticons or punctuation symbols, these features were later included in the sentiment analysis.

Sentiment Analysis

The sentiment analysis aimed to determine the polarity of text through natural language processing. Although most sentiment studies on social networks do not consider emoticons, this tendency has reverted in recent years, as their inclusion increases accuracy. Thus, the value of emoticons (positive, neutral, or negative) was assigned based on the SentiStrength software package [40] and in the context of comments sampled for this study.

Results

Classification of Videos and YouTube Metrics

Among the 923 videos related to organic and local food, 448 were included in the organic food video list and 475 in the local organic food video list. Overall, 606 videos disclosed location (47 countries). As the keywords employed were in English, the lists contained videos mostly from the United States (n=393). The second most frequent location was undisclosed (n=317), followed by India (n=25) for organic food videos and Canada (n=39) for local organic food videos. Overall, organic food videos had higher metrics in terms of views, likes, dislikes, and comments (Table 2).

As for types of uploaders, media-related YouTube channels were the most common for both lists (Table 3). In the case of organic food videos, the second most frequent uploader was “Education,” whereas in the case of local organic food videos, it was “Community.” However, the differences were not significant (P=.42).

Table 2. Nonparametric test of YouTube metrics.

Metric and group	N	Mean rank	Sum of ranks	Z
Views	923			11.056[a]
Organic	448	561.90	251729.00	
Local organic	475	367.78	174697.00	
Likes	905			12.655[a]
Organic	441	565.08	249199.50	
Local organic	464	346.48	160765.50	
Dislikes	905			8.319[a]
Organic	441	515.27	227234.00	
Local organic	464	393.82	182731.00	
Comments	899			9.304[a]
Organic	436	528.26	230323.00	
Local organic	463	376.30	174227.00	

[a]P<.001.

Table 3. Types of uploaders of organic and local organic food videos.

Uploader	Organic[a] (N=448)	Local organic[a] (N=475)
Business	37	87
Community	78	94
Education	93	74
Media	194	157
Others	7	17
Undisclosed	39	46

[a]χ^2_1 (N=923)=4.15; *P*=.42.

The top organic-related videos in terms of views and likes were uploaded mostly by media channels and contained short educative facts about food products. However, some of them were sensationalist. The top video was "Grocery Store Wars," a stop-motion parody of the movie franchise Star Wars, contrasting organic food products and conventional food products in a supermarket. On the other hand, the top videos related to local organic food were business oriented and sometimes employed humor. The video with the highest number of views and likes was a three-dimensional animation commercial, which was part of a campaign by Chipotle Mexican Grill, an American restaurant chain. This video contrasted chemically treated and mechanically processed food products with food produced and processed by farmers. It can be concluded that the top videos in both cases were story-telling driven and showed a contrast between sustainable and nonsustainable foods.

Frames in Organic and Local Organic Food Videos

Based on van Gorp and van der Goot [35], the Good Mother and Natural Goodness frames were expected to be found most frequently in video images and narratives. The third most used frame was Undermining of Foundations in both video groups. A difference was found for the fourth most used frame, which was Frankenstein in the case of organic food videos and Responsibility in the case of local organic food videos (Table 4).

Table 4. Types of frames in organic and local organic food videos.

Frame	Organic[a] (N=808), n (48.26%)	Local organic[a] (N=866), n (51.73%)
Good mother	282 (34.9%)	311 (35.9%)
Natural goodness	220 (27.2%)	253 (29.2%)
Undermining of foundations	153 (18.9%)	180 (20.7%)
Frankenstein	68 (8.4%)	27 (3.1%)
Responsibility	42 (5.1%)	57 (6.5%)
Progress	43 (5.3%)	38 (4.3%)

[a]χ^2_1 (N=1674/923)=4.84; *P*=.03.

Comparison of Network Metrics and Structures

In order to visualize how different is the structure of the two video groups, their network metrics were compared. Because both lists share ties, metrics for the entire video network were also calculated. Local organic food videos had higher connected components, modularity, diameter, and average path lengths (Table 5). Density, number of shortest paths, degree, clustering coefficient, betweenness, and closeness were lower. Such results indicated that the local organic food network was less consolidated than the organic food network. On adding videos from the organic food list, the network showed a slight decrease in its diameter and became better connected, suggesting a broadcast typology for the organic food video network. This can be seen in Figure 1, which is drawn with NodeXL [41].

Table 5. Network centralities.

Centrality name	Description	Organic network	Local organic network	Organic and local organic network
Weakly connected components	Subgroups of nodes that can be reached from every other node in the group.	48	155	197
Density	Total number of ties divided by the number of all possible ties that can exist within a network.	0.038	0.005	0.007
Modularity	The strength of the division between subgroups in a network.	0.197	0.471	0.329
Diameter	Average of the maximum distance between the nodes of a network.	12	16	15
Path length	Average of the distance between the nodes of a network	3.902	4.640	4.652
Number of nodes	N/A[a]	448	475	923
Number of shortest paths	N/A	95,908	28,967	263,249
Degree	Average number of direct connections a node has to other nodes.	9.67	2.50	6.75
Clustering Coefficient	Measure of how close a node is to be part of a group.	0.191	0.081	0.142
Closeness	Average number of steps to access all the other nodes in a network.	3.272	2.509	2.287
Betweenness	Number of shortest paths that connect other nodes in the network by passing through a specific node.	621.457	222.01	532.013

[a]N/A: not applicable.

Figure 1. The local organic food video network and the organic food video network tied together. The organic food video network is presented as a star with central videos reaching the most distant videos within the network.

The top videos in terms of betweenness and in-degree centrality explained the basics of organic and local food and were predominantly uploaded by media channels. In contrast, videos with high out-degree centrality were uploaded mostly by

businesses and individuals. A few organic-related videos identified the food as expensive, while local organic-related videos usually presented a specific area where such food products were available.

Spearman correlation analyses between YouTube metrics and network-related metrics were performed to find how much the popularity features are related to communication patterns in the network. The number of views, likes, dislikes, and comments were moderately correlated with degree, modularity class, eigenvector, and betweenness centralities in the organic food video network (Table 6). As for the local organic network, there were less correlations and the ranks were lower (Table 7). This supports the assumption that the organic network has more consolidated features.

Table 6. Spearman correlations for the organic network.

Variable	Degree	Modularity class	Clustering coefficient	Betweenness
Views	0.209[a]	0.174[a]	−0.185[a]	0.218[a]
Likes	0.187[a]	0.192[a]	−0.201[a]	0.182[a]
Dislikes	0.240[a]	0.120 (*P*=.01)	−0.180[a]	0.193[a]
Comments	0.182[a]	0.157 (*P*=.001)	0.191[a]	0.164 (*P*=.001)

[a]*P*<.001.

Table 7. Spearman correlations for the local organic network.

Variable	Degree	Modularity class	Clustering coefficient	Betweenness
Views	0.156 (*P*=.001)	0.116 (*P*=.01)	0.015 (*P*=.75)	0.287[a]
Likes	0.087 (*P*=.06)	0.075 (*P*=.11)	−0.030 (*P*=.52)	0.239[a]
Dislikes	0.086 (*P*=.07)	0.039 (*P*=.41)	0.029 (*P*=.54)	0.177[a]
Comments	0.143 (*P*=.002)	0.081 (*P*=.08)	0.045 (*P*=.34)	0.257[a]

[a]*P*<.001.

Semantic Analysis of Comments

The Gephi software was used to visualize the semantic networks corresponding to the two video lists. The 107 most frequent words found in the comments sample from the organic food videos were represented with nodes (Figure 2).

The word “organic” was the most frequent in these comments, with the term “food” closely related to it. Verbs connected to “organic” were “grow,” “know,” “like,” “need,” “say,” and “think.” Frame-related words are presented in Table 1. Textual cues associated with the Good Mother frame were “love,” “product,” and “thanks.” Words connected to the Natural Goodness frame were “health,” “healthy,” “raw,” and “taste.” Textual cues associated with the Frankenstein frame were “chemical,” “gmo” (genetically modified organism [GMO]), “hormone,” “Monsanto,” and “pesticide.” All terms related to this frame were connected to “organic.” Products included fruits, vegetables, fish, meat, cheese, and milk. Positive words included “awesome,” “cool,” “great,” “hope,” “nice,” and laughing textual representations like “haha” and “LOL.” The 98 most frequent terms found in the comments sample extracted from the local organic food videos were mapped (Figure 3). The word “food” was the most frequent in these comments, with the words “good,” “organic,” and “video” closely related to it. The verbs connected to “local” were “eating,” “love,” “need,” “produce,” “think,” and “know.” Textual cues associated with the Good Mother frame were “love,” “product,” and “thanks.” Words related to the Natural Goodness frame included “fresh,” “health,” “healthy,” “natural,” and “taste.” In particular, the terms health and healthy were found less frequently than in the organic-related comments. Textual cues associated with the Frankenstein frame were “gmo,” “pesticide,” and “toxic.” Although both the words “local” and “organic” were connected to “gmo,” only the word “organic” was connected to “pesticide” and “toxic,” suggesting a more frequent usage of this frame regarding organic food. Products included fruits, vegetables, corn, and soy. The positive terms were “awesome,” “cool,” “nice,” and laughing textual representations like “LOL.”

Figure 2. Semantic network for organic food video comments. The size reflects the word frequency. Ties show which words were found in the same comment, with tie thickness reflecting the frequency of such relationships.

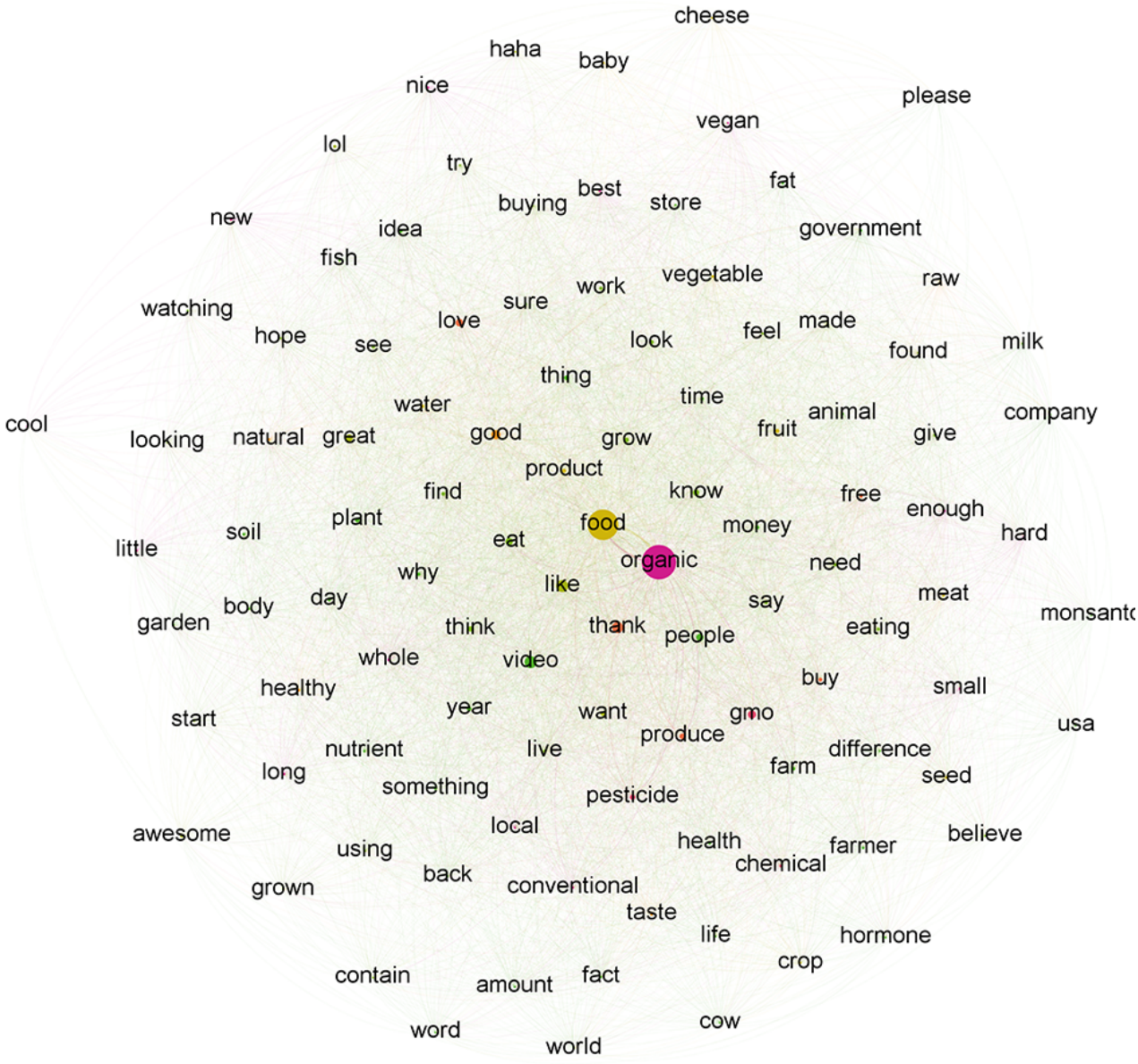
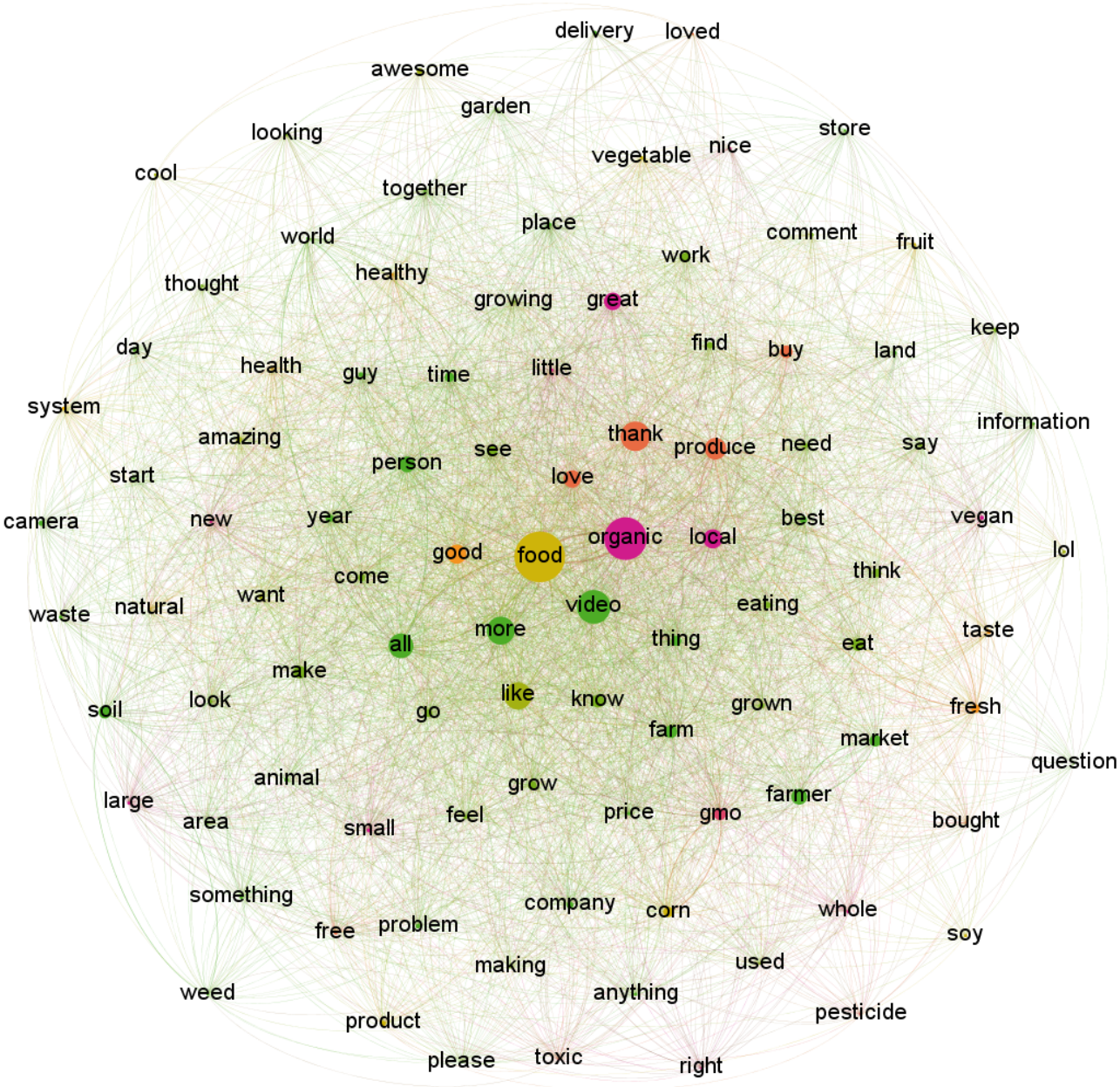


Figure 3. Semantic network for local organic food video comments.



In summary, while both groups of comments featured fruits and vegetables, organic food videos featured dairy and proteins, and local organic food videos featured grains and cereals. Moreover, the organic food network had a greater variety of Frankenstein frame–related words.

Sentiment Analysis of Comments

There was no relevance with regard to negativity valence. However, comments made on local organic videos had a higher rank for positive valence in comparison with comments made on organic videos (Table 8).

Table 8. Nonparametric test: sentiment valence of comments made on the videos.

Valence and group	Mean rank	Sum of ranks	Z	*P*
Positive valence			2.159	.03[a]
Organic	521.2	403927		
Local organic	564.54	163718		
Negative valence			0.15	.88[a]
Organic	532.27	4125409		
Local organic	534.95	155136		

[a]N_1=775, N_2=290.

Discussion

Principal Findings and Comparison With Prior Work

Organic and local food products were communicated on YouTube by several actors articulating efforts to educate the public about food. In particular, the position of organic food videos in the network strengthened the diffusion paths, with a better structural capital than local organic videos, whereas the high in-degree (ties directed to a video) suggested a broadcast network. The relationship between network metrics and negative reactions (number of dislikes and less positive sentiment in comments) implied that negativity might play a relevant role in the diffusion of videos. This coincides with the theory that negative relationships might explain social network outcomes better than positive relationships [42], although most of the video content and user reactions fell in the positive category. Other factors like psychological traits, which influence food choice [43], should also be taken into account. The interplay between psychological traits and the communication of negative/sensationalist video content related to sustainable food should be investigated further in future studies.

Another relevant finding was the words related to “organic” and “local.” The frames have been summarized in Table 1. Although both terms were used in contrast with “gmo,” the amount of risk and artificial-related words linked to “organic” was larger. Taking into account that “organic” and “local” are viewed as good food alternatives (in congruency with the frequent usage of the Natural Goodness and Good Mother frames), a clear contrast between what is considered dangerous and artificial can be seen in some video-related comments. Further, there was more emphasis on health among these comments than in the local organic–related text, reflecting a relationship with subjective well-being, and it is moderated by health concern that has been documented in scientific literature previously [44].

Words like “love” and “thanks” were used often in the case of comments for local organic videos. As they also employed the Responsibility frame more frequently, it implies a more human and social dimension for the word “local.” It also has implications for activating the viewer in areas besides food consumption, as images involving future generations and nonhuman living beings are part of this frame. This partly explains why comments for local organic videos were more positive than those for organic videos.

Finucane et al [45] argued that individuals who have favorable opinions about GMOs are more likely to endorse statements reflecting hierarchical views. The communities analyzed in this study might then represent an egalitarianism perspective linked to organic and local food. Commenters shared their own research on the topics in an articulated, generous, and respectful voice. Thus, usage of the words “know,” “think,” “please,” and “thanks” was frequent.

The persistence of uncomfortable feelings among the public towards GMOs, which are considered as “monsters” according to the Frankenstein frame, points to a failure in scientific communication. This is particularly true regarding information on pesticides, soil depletion, and nutrition. Such communication patterns could be improved by closing the gap between scientists and the public, as attempted by some of the videos uploaded by education channels. This could potentially bring more transparency and trust to the food production chain.

Dissimilar ideas of what is food influence the outcomes of guidelines and food policies. There is little knowledge on the effect of YouTube’s ranking criteria in areas other than politics and entertainment. Hence, a multimethod analysis of the communication of basic human needs, such as food, can provide us with more understanding of the consequences of algorithm usage in rich social media and its interplay with human users. Moreover, we uncovered communication patterns and specific visual and textual cues, providing more comprehensive insights of a complex ecosystem of actors involved in food production, distribution, and consumption. The case of organic and local food intertwines consumerism with environmental concerns that have the potential to impact public health.

Limitations

As the language of analysis was English, there was partially limited access to videos from non-English speaking cultures. Sentiment can be influenced by many contextual features, such as weather [46] and geographical location [47], which were not taken into account for this study.

Conclusions

This study explored the communication paths, discourses, opinions, attitudes, behaviors, and emotions related to sustainable food products in YouTube video networks. Based on our objectives, we can make the following conclusions:

- The organic video network was more consolidated than the local organic video network and was driven mostly by media and business YouTube channels.
- The "Good Mother" and "Natural Goodness" frames were most frequently employed in the videos, followed by "Undermining of Foundations."
- The concept "organic" has become consolidated among both specialists and the public, while the term "local" is in the process of acquiring a formal definition. Nevertheless, the term "organic" was slightly more associated with health risks and negative feelings, while the term "local" was perceived as more human/social and more positive.

Further studies could incorporate more languages, as well as a larger data set. Time-based analysis and segmentation of semantic analysis across different geographical locations would also deepen the present understanding of the diffusion of both organic and local food products. Another area for further exploration is GMOs, as this technology will continue to transform food production and consumption patterns in contrast with traditional agricultural methods.

Acknowledgments

The authors are grateful to Deepak Vijay, Priyanka Ram, Reach Syden, and Sittiphan Jiyavorananda for their help with translation. A previous version of this article was presented at the 6th International Kansei Engineering and Emotion Research Conference, UK in 2016.

Conflicts of Interest

None declared.

References

1. Sustainable Development Goals. United Nations. 2018. URL: https://www.un.org/sustainabledevelopment/sustainable-development-goals [accessed 2020-07-11]
2. Seyfang G. Ecological citizenship and sustainable consumption: Examining local organic food networks. Journal of Rural Studies 2006 Oct;22(4):383-395. [doi: 10.1016/j.jrurstud.2006.01.003]
3. Arbenz M. It is time for Organic 3.0!. In: Willer H, Lernoud J, editors. The world of organic agriculture. Statistics & emerging trends 2014. Switzerland: FiBL and IFOAM; 2014.
4. Willer H, Lernoud J, editors. The World of Organic Agriculture. Statistics & Emerging Trends 2018. Switzerland: FiBL and IFOAM; 2018.
5. Lee H, Yun Z. Consumers' perceptions of organic food attributes and cognitive and affective attitudes as determinants of their purchase intentions toward organic food. Food Quality and Preference 2015 Jan;39:259-267. [doi: 10.1016/j.foodqual.2014.06.002]
6. Zander K, Hamm U. Consumer preferences for additional ethical attributes of organic food. Food Quality and Preference 2010 Jul;21(5):495-503. [doi: 10.1016/j.foodqual.2010.01.006]
7. Nosi C, Zanni L. Moving from "typical products" to "food - related services". British Food Journal 2004 Oct;106(10/11):779-792. [doi: 10.1108/00070700410561388]
8. Thilmany D, Bond CA, Bond JK. Going Local: Exploring Consumer Behavior and Motivations for Direct Food Purchases. American Journal of Agricultural Economics 2008 Dec;90(5):1303-1309. [doi: 10.1111/j.1467-8276.2008.01221.x]
9. Darby K, Batte M, Ernst S, Roe B. Decomposing Local: A Conjoint Analysis of Locally Produced Foods. American Journal of Agricultural Economics 2008 May;90(2):476-486. [doi: 10.1111/j.1467-8276.2007.01111.x]
10. Sims R. Food, place and authenticity: local food and the sustainable tourism experience. Journal of Sustainable Tourism 2009 May 19;17(3):321-336. [doi: 10.1080/09669580802359293]
11. Tufecki Z. We're building a dystopia just to make people click on ads. YouTube. 2017. URL: https://www.youtube.com/watch?v=iFTWM7HV2UI [accessed 2020-07-11]
12. Lewis P. Fiction is outperforming reality: how YouTube's algorithm distorts truth. The Guardian. 2018. URL: https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth [accessed 2020-07-11]
13. Zhang M, Jansen B, Chowdhury A. Business engagement on Twitter: a path analysis. Electron Markets 2011 Aug 2;21(3):161-175. [doi: 10.1007/s12525-011-0065-z]
14. Rutsaert P, Regan A, Pieniak Z, McConnon A, Moss A, Wall P, et al. The use of social media in food risk and benefit communication. Trends in Food Science & Technology 2013 Mar;30(1):84-91. [doi: 10.1016/j.tifs.2012.10.006]
15. Shan L, Panagiotopoulos P, Regan A, De Brún A, Barnett J, Wall P, et al. Interactive communication with the public: qualitative exploration of the use of social media by food and health organizations. J Nutr Educ Behav 2015;47(1):104-108. [doi: 10.1016/j.jneb.2014.09.004] [Medline: 25449827]
16. Pantelidis IS. Electronic Meal Experience: A Content Analysis of Online Restaurant Comments. Cornell Hospitality Quarterly 2010 Aug 11;51(4):483-491. [doi: 10.1177/1938965510378574]

17. Abrams KM, Sackmann A. Are Alternative Farmers Yielding Success with Online Marketing and Communication Tools for Their Social Capital and Business Viability? Journal of Applied Communications 2014 Sep 01;98(3):48-62. [doi: 10.4148/1051-0834.1085]
18. Gilpin D. Organizational Image Construction in a Fragmented Online Media Environment. Journal of Public Relations Research 2010 Jul 02;22(3):265-287. [doi: 10.1080/10627261003614393]
19. Vargas Meza X. An analysis of cultural differences and similarities in Twitter communication using webometrics method: the case of organic products in Mexico and South Korea. Gyeongsan, South Korea: Yeungnam University; 2014.
20. Meza X, Park HW. Organic Products in Mexico and South Korea on Twitter. J Bus Ethics 2015 Jan 31;135(3):587-603. [doi: 10.1007/s10551-014-2345-y]
21. Ponce L, Cordelier B. Emotions on Facebook: a content analysis of Mexico's Starbucks page. In: Proceedings of the 2015 International Conference on Social Media & Society. 2015 Presented at: 2015 International Conference on Social Media & Society; July 2015; Toronto, Ontario, Canada p. 1-7. [doi: 10.1145/2789187.2789207]
22. Hilverda F, Kuttschreuter M, Giebels E. The Effect of Online Social Proof Regarding Organic Food: Comments and Likes on Facebook. Front. Commun 2018 Aug 3;3:30 [FREE Full text] [doi: 10.3389/fcomm.2018.00030]
23. Kane K, Chiru C, Ciuchete S. Exploring the eco-attitudes and buying behaviour of Facebook users. Amfiteatru Economic 2012;14(31):157-171.
24. Kholeif O, McCormack S, Butler E. Electronic Superhighway: From Experiments in Art and Technology to Art After the Internet. London, UK: Whitechapel Gallery; 2016.
25. Cheng X, Liu J, Dale C. Understanding the Characteristics of Internet Short Video Sharing: A YouTube-Based Measurement Study. IEEE Trans. Multimedia 2013 Aug;15(5):1184-1194. [doi: 10.1109/TMM.2013.2265531]
26. Rhoades E, Ellis J. Food Tube: Coverage of food safety issues through video. Journal of Food Safety 2010;30(1):162-176. [doi: 10.1111/j.1745-4565.2009.00198.x]
27. Kuttschreuter M, Rutsaert P, Hilverda F, Regan A, Barnett J, Verbeke W. Seeking information about food-related risks: The contribution of social media. Food Quality and Preference 2014;37:10-18. [doi: 10.1016/j.foodqual.2014.04.006]
28. Allen GJ. In: Albala K, editor. The Business of Food: Encyclopedia of the Food and Drink Industries. Westport, CT: Greenwood; 2007.
29. International Federation of Organic Agriculture Movements. The IFOAM norms for organic production and processing. Germany: Die Deutsche Bibliothek; 2012.
30. Martinez S. Local Food Systems Concepts, Impacts, and Issues. United States Department of Agriculture. 2010. URL: https://www.ers.usda.gov/webdocs/publications/46393/7054_err97_1_.pdf?v=42265 [accessed 2020-07-18]
31. Rieder B. YouTube Data Tools, Ver. 1. Digital Methods. 2015. URL: https://tools.digitalmethods.net/netvizz/youtube/ [accessed 2020-07-18]
32. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. 2009 Presented at: International AAAI Conference on Weblogs and Social Media; 2009; San Jose, California, US.
33. Coffman M, Umemoto K. The triple-bottom-line: framing of trade-offs in sustainability planning practice. Environ Dev Sustain 2009 Oct 9;12(5):597-610. [doi: 10.1007/s10668-009-9213-4]
34. Vicari S. Measuring collective action frames: A linguistic approach to frame analysis. Poetics 2010 Oct;38(5):504-525. [doi: 10.1016/j.poetic.2010.07.002]
35. Van Gorp B, van der Goot MJ. Sustainable food and agriculture: Stakeholder's frames. Communication, Culture & Critique 2012;5(2):127-148. [doi: 10.1111/j.1753-9137.2012.01135.x]
36. Prell C, editor. Social Network Analysis History, Theory and Methodology. London: SAGE; 2012:1.
37. Tsou A, Thelwall M, Mongeon P, Sugimoto C. A community of curious souls: an analysis of commenting behavior on TED talks videos. PLoS One 2014;9(4):e93609 [FREE Full text] [doi: 10.1371/journal.pone.0093609] [Medline: 24718634]
38. RANDOM.ORG. URL: https://www.random.org/ [accessed 2020-07-18]
39. Leydesdorff L. TI.exe for Co-Word Analysis. Leydesdorff. 2013. URL: http://www.leydesdorff.net/software/ti/index.htm [accessed 2020-07-18]
40. Thelwall M, Buckley K, Paltoglou G. Sentiment strength detection for the social web. J. Am. Soc. Inf. Sci 2011 Oct 13;63(1):163-173. [doi: 10.1002/asi.21662]
41. Smith M, Ceni A, Milic-Frayling N, Shneiderman B, Mendes Rodrigues E, Leskovec J, et al. NodeXL. Social Media Research Foundation. 2010. URL: https://www.smrfoundation.org/nodexl/ [accessed 2013-04-29] [WebCite Cache ID 6GFF0fwWa]
42. Labianca G, Brass DJ. Exploring the Social Ledger: Negative Relationships and Negative Asymmetry in Social Networks in Organizations. AMR 2006 Jul;31(3):596-614. [doi: 10.5465/amr.2006.21318920]
43. Cliceri D, Spinelli S, Dinnella C, Prescott J, Monteleone E. The influence of psychological traits, beliefs and taste responsiveness on implicit attitudes toward plant- and animal-based dishes among vegetarians, flexitarians and omnivores. Food Quality and Preference 2018 Sep;68:276-291. [doi: 10.1016/j.foodqual.2018.03.020]
44. Apaolaza V, Hartmann P, D'Souza C, López CM. Eat organic – Feel good? The relationship between organic food consumption, health concern and subjective wellbeing. Food Quality and Preference 2018 Jan;63:51-62. [doi: 10.1016/j.foodqual.2017.07.011]

45. Finucane ML, Slovic P, Mertz CK, Flynn J, Satterfield TA. Gender, race, and perceived risk: The 'white male' effect. Health, Risk & Society 2000 Jul;2(2):159-172. [doi: 10.1080/713670162]
46. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
47. Gore RJ, Diallo S, Padilla J. You Are What You Tweet: Connecting the Geographic Variation in America's Obesity Rate to Twitter Content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]

Abbreviations

API: application program interface
GMO: genetically modified organism

Edited by G Eysenbach; submitted 22.10.19; peer-reviewed by H Lawrence, R Gore; comments to author 17.12.19; revised version received 24.03.20; accepted 30.03.20; published 10.08.20.

Please cite as:
Vargas Meza X, Yamanaka T
Food Communication and its Related Sentiment in Local and Organic Food Videos on YouTube
J Med Internet Res 2020;22(8):e16761
URL: https://www.jmir.org/2020/8/e16761
doi:10.2196/16761
PMID:32773370

Review

When Public Health Research Meets Social Media: Knowledge Mapping From 2000 to 2018

Yan Zhang[1*], PhD; Bolin Cao[1*], PhD; Yifan Wang[1], BA; Tai-Quan Peng[2], PhD; Xiaohua Wang[1], PhD

[1]School of Media and Communication, Shenzhen University, Shenzhen, China

[2]Department of Communication, Michigan State University, East Lansing, MI, United States

[*]these authors contributed equally

Corresponding Author:
Bolin Cao, PhD
School of Media and Communication
Shenzhen University
Rm809, Building L7, South Campus, Shenzhen University
Nanhai Avenue, Nanshan District
Shenzhen
China
Phone: 86 13247393943
Email: caobolin@szu.edu.cn

Abstract

Background: Social media has substantially changed how people confront health issues. However, a comprehensive understanding of how social media has altered the foci and methods in public health research remains lacking.

Objective: This study aims to examine research themes, the role of social media, and research methods in social media–based public health research published from 2000 to 2018.

Methods: A dataset of 3419 valid studies was developed by searching a list of relevant keywords in the Web of Science and PubMed databases. In addition, this study employs an unsupervised text-mining technique and topic modeling to extract research themes of the published studies. Moreover, the role of social media and research methods adopted in those studies were analyzed.

Results: This study identifies 25 research themes, covering different diseases, various population groups, physical and mental health, and other significant issues. Social media assumes two major roles in public health research: produce substantial research interest for public health research and furnish a research context for public health research. Social media provides substantial research interest for public health research when used for health intervention, human-computer interaction, as a platform of social influence, and for disease surveillance, risk assessment, or prevention. Social media acts as a research context for public health research when it is mere reference, used as a platform to recruit participants, and as a platform for data collection. While both qualitative and quantitative methods are frequently used in this emerging area, cutting edge computational methods play a marginal role.

Conclusions: Social media enables scholars to study new phenomena and propose new research questions in public health research. Meanwhile, the methodological potential of social media in public health research needs to be further explored.

(J Med Internet Res 2020;22(8):e17582) doi:10.2196/17582

KEYWORDS

social media; public health; infodemiology; infoveillance; topic modeling; research theme; research method

Introduction

Social media has deeply penetrated people's lives in many aspects. In developing and developed societies, social media has played a significant role in health management and disease control [1]. Social media is integrated into empirical examinations of the prevention and control of various types of diseases, including emerging, infectious, and chronic diseases [2-4]. Social media has been employed to study health phenomena among different populations or social groups such as children, pregnant women, and older adults [5,6]; different genders [7,8]; and individuals in various social classes [9].

Agencies widely use social media to fulfill different health purposes. For the general public, social media is used to satisfy

its orientation for health information, linking with health services and communication with others who share the same health interests [10]. For public health professionals and organizations, social media (eg, Facebook, Grindr, mobile apps) serves as a multifunctional tool to launch interventions to reach a wide array of the population efficiently [11-14]. The large volume of mobility and discourse data on social media (eg, Twitter) can be conducive for public health management, including disease surveillance, assessment, and control [15-17].

These studies have demonstrated the increase of scholarly interest in empirical research conducted on social media platforms with public health goals, including the social media–based public health research in this study [18-21]. Although many scholars in social science and public health have contributed to this field, the overview about how social media has been integrated into public health research remains limited. Prior systematic reviews on social media–based public health research often focused on certain domains or topics. Many reviews systematically investigated the effectiveness of social media interventions for varied specific health outcomes, such as the promotion of safe sexual health behaviors [22], vaccine uptake [23], noncommunicable disease management [24], and HIV prevention [19]. Scholars are often dedicated to one or two domains, neglecting the fact that using social media in one field may shed light on another. Meanwhile, focusing on one particular area, these reviews often face challenges to identify similar patterns across domains and capture an integrated picture about social media–based public health research. In addition, most existing reviews included a limited number of original articles [25,26]. Even a review of systematic reviews only extracted few studies [18]. In this emerging and fast-growing subject field, the limited literature being included may fail to provide a panoramic description of social media–based public health research.

Furthermore, most prior systematic reviews have adopted a top-down approach and therefore may have narrowed the view by overlooking certain nuances and novelties that have been emerging. This study adopts a bottom-up approach [27] to understand the growth of social media–based public health research and remain open to map the intellectual landscape in this area. Specifically, the study aims to address the following research questions:

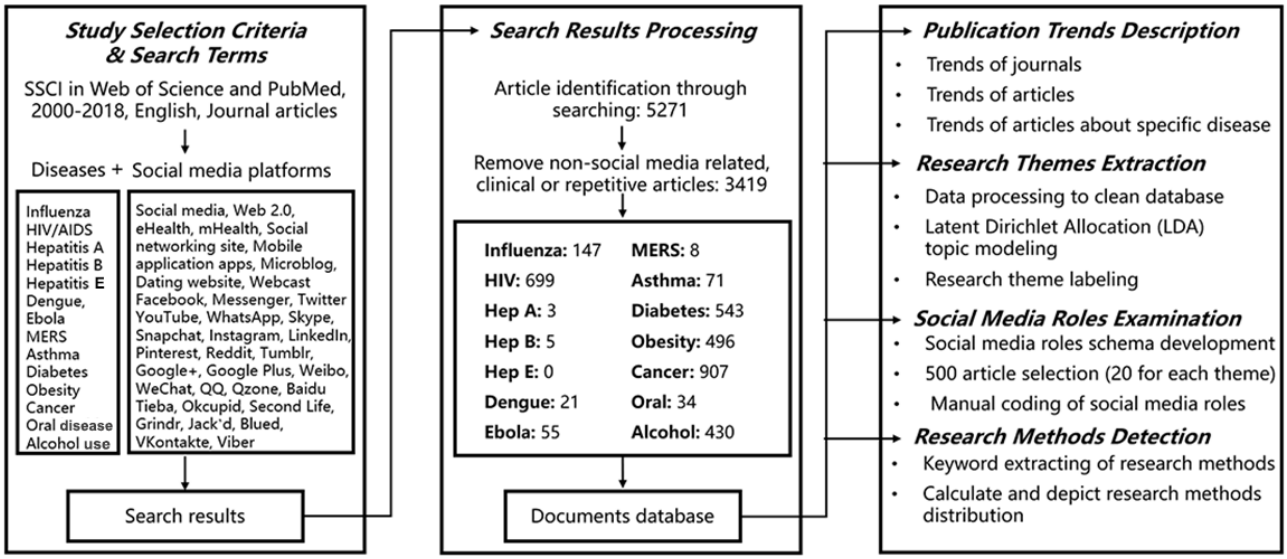
- RQ 1: What are the major publication trends of social media–based public health research since 2000?
- RQ 2: What are the major research themes in social media–based public health research since 2000?
- RQ 3: What role does social media play in social media–based public health research since 2000?
- RQ 4: What are the major research methods adopted in social media–based public health research since 2000?

Methods

Data Collection

To examine how social media has been adopted and integrated into public health research, a list of terms was identified and the Web of Science (Clarivate Analytics) and PubMed databases were searched (see Figure 1). Lists of keywords about social media and disease were established. This study focused on emerging, infectious, and chronic diseases. Specifically, 14 diseases were selected that are of high prevalence among the population or pose major public health threats according to the World Health Organization [28]: influenza, HIV, hepatitis A, hepatitis B, hepatitis E, dengue, Ebola, Middle East respiratory syndrome (MERS), asthma, diabetes, obesity, cancer, oral disease, and alcohol use. A list of keywords was constructed. Then a list of social media keywords, including the general social media categories and specific social media platforms, was established.

Figure 1. Knowledge-mapping workflow of social media–based public health research from 2000 to 2018.



The lists of disease and social media keywords were combined pairwise and submitted to search titles, abstracts, and keywords of published studies in the databases of the Web of Science and PubMed. The publication period was limited to between 2000 and 2018. The article language was limited to English, and document type was limited to scholarly journal articles.

Considering the PubMed database includes numerous medical and clinical studies that are beyond the research scope of this study, the search results were refined by setting the broad subject terms as related categories, such as public health and medical informatics to reduce the noise in the search results. This search strategy led to the identification of 5271 articles in the two databases. Then, another round of data checking to remove unqualified studies such as non–social media–related articles, duplicate records, and clinical studies was implemented. Finally, a dataset of 3419 articles was collected for further analysis. Ultimately, document level information of the 3419 articles from the databases, including authors, article title, journal title, abstract, author keywords, and cited references, was retrieved.

Data Analysis

An automatic text-mining approach was adopted to extract research themes in the field of social media–based public health research. As the abstracts of published studies conveyed the themes or foci of the articles [29], the article abstracts were mined through latent Dirichlet allocation (LDA) topic modeling, which is a popular unsupervised text-mining technique in computational social science. LDA topic modeling helps recognize the structure of research development, current trends, and interdisciplinary landscapes of research [27]. The LDA topic modeling [30] was implemented with the tm package in the R software (R Foundation for Statistical Computing). Data preprocessing, such as removal of stop words and numbers, was performed before the LDA topic modeling.

Numerous LDA topic models were estimated with various numbers of topics. These models were evaluated on the basis of three main criteria: (1) a substantial proportion of articles exists under each topic, (2) themes show independence with one another and the lists of top terms of topics are not highly overlapped or not relevant, (3) models differing in theme number are compared to identify the nuanced differences and determine the best theme extraction by assuring that each term list is coherent. Finally, a topic model with 25 topics, which presented adequate discrimination between topics and convergence within a topic, was selected. The articles were classified into the research theme with which they had the greatest probability scores.

To understand how social media has been integrated into public health research with different thematic foci, a manual content analysis was conducted among a randomly selected sample of 500 abstracts (20 from articles in each theme) to understand the role of social media. A coding scheme was developed by two authors of this study. The two authors first separately coded a subsample of 60 randomly selected abstracts to construct the coding scheme. After several rounds of exploration and discussion, they achieved a satisfying intercoder reliability (as measured with a Cohen kappa). The role of social media is categorized into two main types. First, social media provides a substantial research interest for public health research, which includes the use of social media for intervention, as human-computer interaction characteristics, as platforms of social influence, and for risk assessment or disease prevention. Second, social media is employed as a context in public health research with social media as mere reference, as participant recruitment tool, or as data source. Categories and their definitions are further illustrated in the analytical findings section (also see Multimedia Appendix 1).

Finally, the research methods adopted by social media–based public health research were identified by searching a list of keywords associated with various research methods among the titles, abstracts, and keywords of the retrieved studies. Figure 1 summarizes the study workflow.

Results

Publication Trends in Social Media–based Public Health Research

Publication trends in the field of social media–based public health research in the past two decades were presented in 3 dimensions: growth of overall publications, growth of publications by specific diseases, and growth of journal outlets.

Empirical studies in this area were relatively limited in the first decade (2000 to 2010), which demonstrated a minimal increase as shown in Figure 2. A significant annual increase was observed from 2011 to 2018. Such trends intersected with the advancement of the internet, especially that of social media. Although popular social media platforms, such as Facebook, Twitter, and Instagram, were launched before 2010, they have been widely accepted worldwide since 2010. This implies that social media–based public health research is a study area responsive to technological development. Prior research also demonstrated similar findings that internet research evolved along with technological development [29].

Social media has been increasingly incorporated into the studies of certain types of diseases in the past decades. Dramatic increases in research occurred on cancer, HIV, diabetes, obesity, and alcohol use after 2010. Other diseases, such as influenza, hepatitis A, hepatitis B, dengue, Ebola, MERS, asthma, and oral disease, showed a relatively slow growth rate that remained quite stable from 2000 to 2018.

For the journal outlets, a total of 799 journals published studies in these areas (see Figure 3). Table 1 reports the 15 most visible journals in this area. Among them, Journal of Medical Internet Research, which published 331 articles, is the most visible one accounting for 9.68% of the total publications. Moreover, JMIR sister journals, such as JMIR mHealth and uHealth (165 publications), JMIR Research Protocols (114 publications), and JMIR Public Health and Surveillance (49 publications) also showed great interest in this domain. Other high-ranked journals included PLoS One, BMC Public Health, Studies in Health Technology and Informatics, AIDS and Behavior, and BMJ Open, each occupy more than 1.5% of publication in this field.

Figure 2. Total number of articles for each disease in the Social Sciences Citation Index from 2000 to 2018.

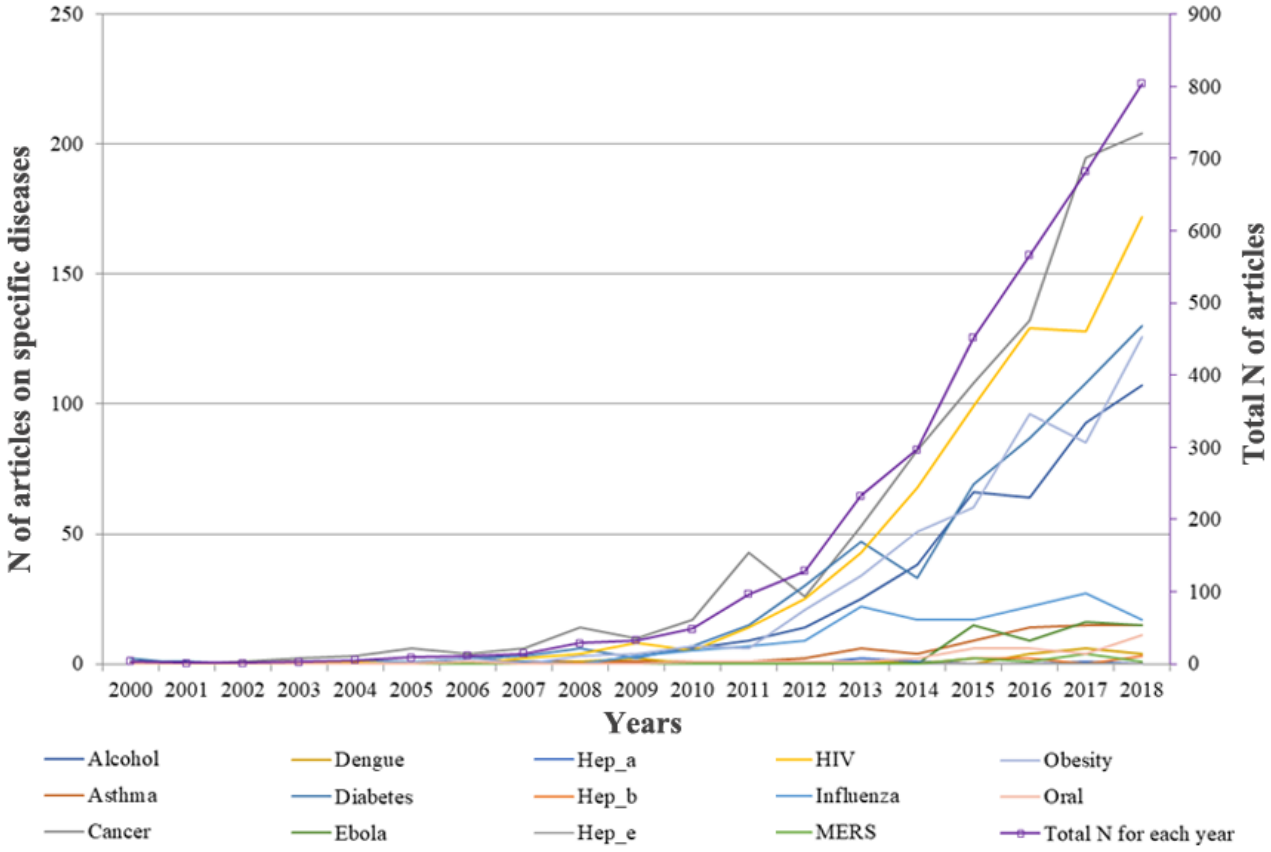


Figure 3. Number of unique outlets that published output of social media–based public health research from 2000 to 2018.

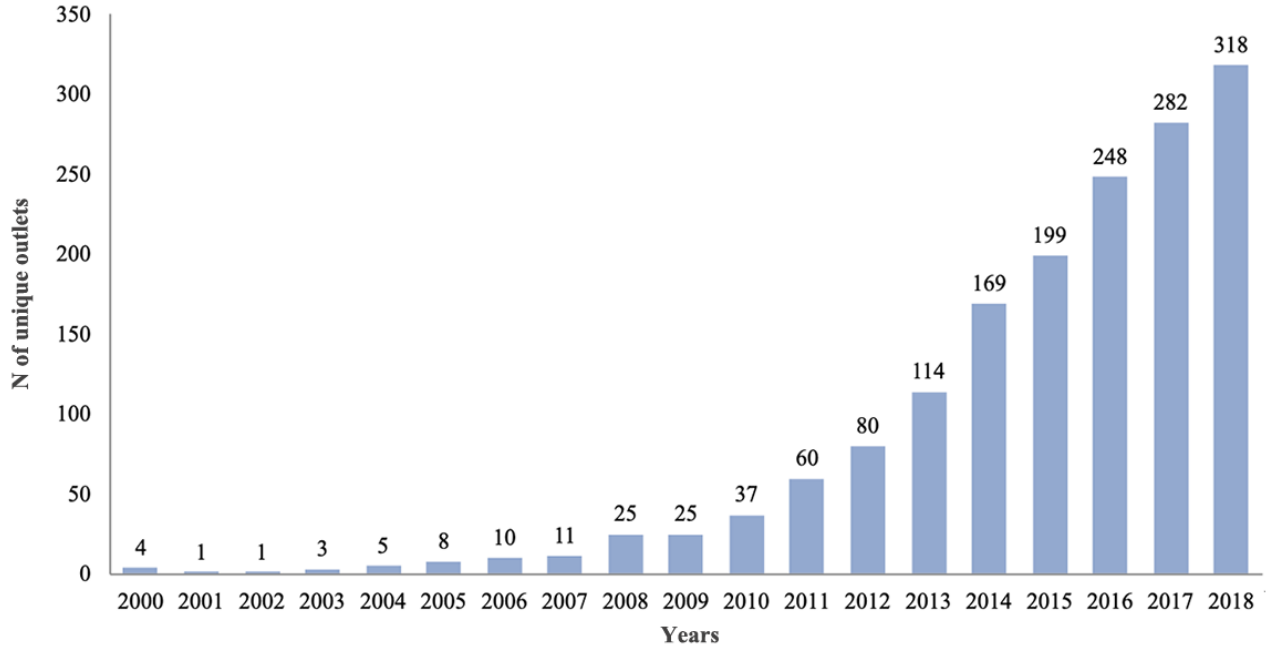


Table 1. Top journals in social media–based public health research.

Number	Journal name	Publications, n (%)
1	Journal of Medical Internet Research	331 (9.7)
2	JMIR mHealth and uHealth	165 (4.8)
3	PLoS One	123 (3.6)
4	JMIR Research Protocols	114 (3.3)
5	BMC Public Health	72 (2.1)
6	Studies in Health Technology and Informatics	68 (2.0)
7	AIDS and Behavior	53 (1.5)
8	BMJ Open	52 (1.5)
9	JMIR Public Health and Surveillance	49 (1.4)
10	Journal of Health Communication	49 (1.4)
11	International Journal of Medical Informatics	37 (1.1)
12	Journal of Diabetes Science and Technology	37 (1.1)
13	Health Communication	34 (1.0)
14	Computers in Human Behavior	33 (1.0)
15	International Journal of Environmental Research and Public Health	31 (0.9)

Research Themes in Social Media–Based Public Health Research

The 25 extracted research themes were labeled on the basis of the top 15 most frequently used terms associated with each theme and the articles assigned to the theme. Table 2 presents the lists of terms under each theme. A network graph of theme-word probability of the 25 research themes is provided in Multimedia Appendix 2. Multimedia Appendix 3 displays a typical study under each theme.

The percentages in Table 2 reveal the article distribution across research themes. The articles under each theme varied greatly from 2.22% (76/3419) to 7.93% (271/3419; Table 2), with men and HIV occupying the largest number of articles and reproductive cancers the least. Among them, the mHealth family, the themes about mHealth (themes 1, 2, 3, 4), contained a large body of 653 articles. Themes about substance use (themes 6, 7, 8) comprised 409 articles. Another big cluster was cancer (themes 10, 12, 13), which consisted of 385 articles.

The 25 research themes were further grouped into 6 research clusters on the basis of similar concerns and associations. The first cluster was on health education, which comprises 4 themes: health education–school and students, health education–family and oral/dental health, mHealth and medical decisions, and pregnancy. Health education aims to prevent diseases through improving people's knowledge and health efficacy. School and family, as the main scenes for the students to learn health beliefs and behaviors, have been the foci of health education. Sexual health education on condom use and pregnancy have also attracted increasing scholarly attention.

The second cluster was on health management with the help of mHealth. The themes mHealth and weight control, mHealth and diabetes management, digital campaigns in targeted populations, social media and alcohol drinking, substance usage and cessation, food and asthma, and vaccination and immunization all fell into this cluster. This indicates the functional attributes of social media to help manage health problems. Social media use is used to intervene in certain unhealthy behaviors and promote healthy behavior adoption.

The third cluster, cancer studies, includes women's cancer, reproductive cancer, cancer survivor, and caregiving on social media. Cancer is one of the world's largest health problems and a significant cause of death. Thus, continuous attention has been paid to cancer studies.

The fourth cluster, infectious diseases, includes HIV as a key topic: mHealth and HIV; men and HIV; and infectious disease, health campaign, and stigma belong to this cluster. In this line of research, social media provides breakthrough channels to reach risky subgroups and focuses more attention on campaigns to reduce the stigma surrounding infectious diseases.

The fifth cluster was on mental health issues. This cluster consists of two themes: mental health and substance use and mental health–depression and digital technology. Mental health problems have been prominent in modern society. Digital technology is considered a cause and a solution to mental health issues.

The sixth cluster was on extended health research empowered by social media: health and human mobility, health marketing, surveillance and Twitter, and eHealth–miscellaneous. These research areas have flourished due to the availability of geographical information, mass user behavioral data, and extensive online discourse on social media platforms.

Table 2. Research themes and the top 15 keywords under each theme.

Number	Research themes	Top 15 keywords	n (%)
1	mHealth and weight management	App, weight, loss, selfmonitor, usabl, mHealth, adher, exercise, download, dietary, BMI, Fit, Android, coach, mainten	208 (6.08)
2	mHealth and diabetes management	diabet, selfmanag, glucos, usabl, mHealth, adolesc, HbA_{1c}, TDM, young, selfcar, older, glycem, insulin, selfefficacy, cardiovascular	192 (5.62)
3	mHealth and medical decisions	decis, mHealth, intent, peer, consum, screen, trust, cell, doctor, choic, privacy, navig, leader, read, worker	103 (3.01)
4	mHealth and HIV	adher, mhealth, HIV, literacy, SMS, ART, portal, selfmanag, PLWH, retent, digit, RCTs, beta, viral, nurs	150 (4.39)
5	Men and HIV	HIV, men, sexual, MSM, partner, PrEP, AOR, gay, condom, drug, young, STI, YMSM, websit, Latino	271 (7.93)
6	Alcohol drinking and social media	alcohol, drink, young, consumpt, student, Facebook, post, colleg, alcoholrel, SNS, exposur, norm, adolesc, market, peer	174 (5.09)
7	Substance use and cessation	smoke, cessat, smoker, drug, tobacco, quit, marijuana, cigarett, EMA, abstin, substanc, addict, ecolog, alcohol, momentary	115 (3.36)
8	Mental health and substance use	symptom, mental, substanc, pain, disord, veteran, drug, cope, distress, adolesc, tan, fatigu, abus, stress, pro	120 (3.51)
9	Mental health–depression and digital technologies	depress, pain, digit, anxiety, adolesc, symptom, suicid, cancer, disord, older, memory, genet, injury, dengu, young	112 (3.28)
10	Women's cancer	cancer, breast, screen, women, vaccin, HPV, campaign, prostat, colorect, cervic, lung, imag, news, cancerrel, papillomavirus	151 (4.42)
11	Pregnancy	women, pregnanc, pregnant, mother, gestat, GDM, matern, worker, CHWS, child, contracept, HCV, mHealth, postpartum, antenat	117 (3.42)
12	Reproductive cancers	cancer, ovarian, prostat, gene, polymorph, cultur, genotyp, postop, cohort, predict, nutrit, genet, PON, BRCA, surgic	76 (2.22)
13	Cancer survivor care	cancer, survivor, emot, breast, psychosoci, oncolog, young, modul, wellb, survivorship, AYA, selfmanag, QOL, depress, consult	158 (4.62)
14	Caregiving on social media	Facebook, post, caregiv, page, blog, comment, emot, CRC, virtual, channel, Twitter, friend, profil, chat, fit	136 (3.98)
15	Vaccination and immunization	vaccin, influenza, predict, coverag, season, event, news, queri, flu, immun, volum, forecast, surveil, websit, outbreak	119 (3.48)
16	Infectious disease, health campaigns and stigma	Ebola, HIV/AIDS, campaign, epidem, stigma, outbreak, audienc, Africa, outreach, facil, news, post, neighbourhood, stori, IBD	110 (3.22)
17	Food and asthma	food, asthma, nutrit, game, children, intak, consumpt, dietari, exposur, veget, infant, eat, weight, beverag, feed	146 (4.27)
18	Health education–family and oral/dental health	parent, websit, children, oral, dental, child, readabl, read, grade, instrument, discern, childhood, rank, pediatr, page	128 (3.74)
19	Health education–school and students	student, nurs, physician, mHealth, skin, school, melanoma, sun, cluster, EMR, hypertens, rural, India, NCDS, CVD	107 (3.13)
20	Health and human mobility	map, street, neighborhood, urban, walk, audit, built, sale, resid, agreement, happi, crowdsourc, hookah, LOS, imag	111 (3.25)
21	Digital campaigns in targeted populations	youth, advertis, rural, Hispan, adolesc, percent, young, homeless, campaign, urban, digit, women, black, cultur, underserv	131 (3.83)
22	Health behavior guidelines	behavior, guidelin, usag, practition, programm, ethic, sedentary, men, websit, GPS, ICT, screen, kingdom, citat, geosoci	86 (2.52)
23	Health marketing and social media	video, YouTub, market, brand, product, girl, adolesc, tobacco, industry, consum, company, ecigarett, boy, cigarett, surgery	100 (2.92)
24	Surveillance and Twitter	tweet, Twitter, surveil, post, sentiment, ILI, influenza, outbreak, detect, drug, opinion, mention, retweet, marijuana, pandem	165 (4.83)
25	eHealth–miscellaneous	eHealth, phase, eat, referr, client, mHealth, telemedicin, uncertainty, COPD, reward, PLHIV, emot, EVD, static, compet	133 (3.89)

Roles of Social Media in Public Health Research

Social Media as Research Context or Substantial Interest

Social media is integrated into public health research by providing a new research context or producing new substantial interest in public health research.

When social media was adopted as a research context, social media was specifically considered as a mere reference, a platform for participant recruitment, and as a data source. When social media was adopted as a mere reference, research mostly used social media as a tool to offer intervention and facilitate the health management of individuals. For the role of a platform for recruitment, research either recruited participants through distributing questionnaires or posting participant recruitment announcements on social media (eg, Facebook) or employed users of certain social media platforms as the study target group (Grindr for the men who have sex with men group). For the role of data source, social media could contribute to collecting data in text, image, video, and app interface formats and collecting published posts and articles for meta-analysis or scope review.

When social media produced substantial interests for public health research, social media was used for intervention; employed to study human-computer interaction characteristics; used as a platform of social influence; and used for disease surveillance, risk assessment, or prevention. Under these 4 broad categories, the role of social media is described as follows:

Intervention

For public health intervention, the 4 subroles of social media in the published studies are as follows: (1) interactive intervention tool targeted at changing personal and environmental risky health factors, (2) intervention information-distributing tool (1-way and not real-time interactive), (3) source for health information seeking, such as YouTube and other platforms, and (4) usability test of social media platforms as intervention instruments.

Human-Computer Interaction Characteristics

Under this role, social media was used to serve the goal of revealing (1) the public's attitudes toward technology and social media for health use, (2) characteristics and behaviors of social media users and groups, (3) factors affecting the health behaviors or attitudes of users on social media, and (4) consequences/influences on health behaviors caused by (popular) social media.

Platform of Social Influence

Social networking and interaction between different individuals and groups on social media could facilitate the change of health behaviors through the following approaches: (1) building online (support) groups for patients, such as cancer patient groups on Facebook, (2) promoting physician-patient communication or information seeker–provider communication, (3) enhancing health-related marketing, such as precision advertising, and (4) changing public health behavior at a macro level. All of these approaches are representations of social influence in online communities.

Disease Surveillance, Risk Assessment, or Prevention

The digital traces of online behaviors and massive online discourse granted opportunities to understand health conditions at a population level. For instance, Google trends and search query records could grant references to predict the possibility of a flu outbreak at an early stage.

Figure 4 presents the percentages of articles in each type of social media role. The results showed that among substantial interest, "an interactive intervention tool targeted at changing personal and environmental risky health factors" accounted for the largest percentage of 19.2%, followed by "usability test of social media platforms as intervention instruments" (10.6%), and "a source for health information seeking" (8.0%). Four of other types, "characteristics and behaviors of social media users and groups" (7.4%), "consequences/influences on health behaviors caused by (popular) social media" (6.4%), "enhancing health-related marketing" (6.4%), and "changing public health behavior as macro influence" (6.4%), also occupied a relatively larger proportion of more than 6%.

Among the dimension of social media as research context (Figure 5), "as a mere reference," which took social media as a research background or a research environment, played the dominant role (47.2%). The second most frequent role that social media plays was content data source (ie, text/picture/video/app data sources, 24.8%). The other three were "social media as article search platform for meta-analysis or literature review" (11.6%), "for participant recruitment" (10.6%), and "as platforms to recruit their users" (1.2%).

Figure 4. Article distributions based on social media as substantial interest.

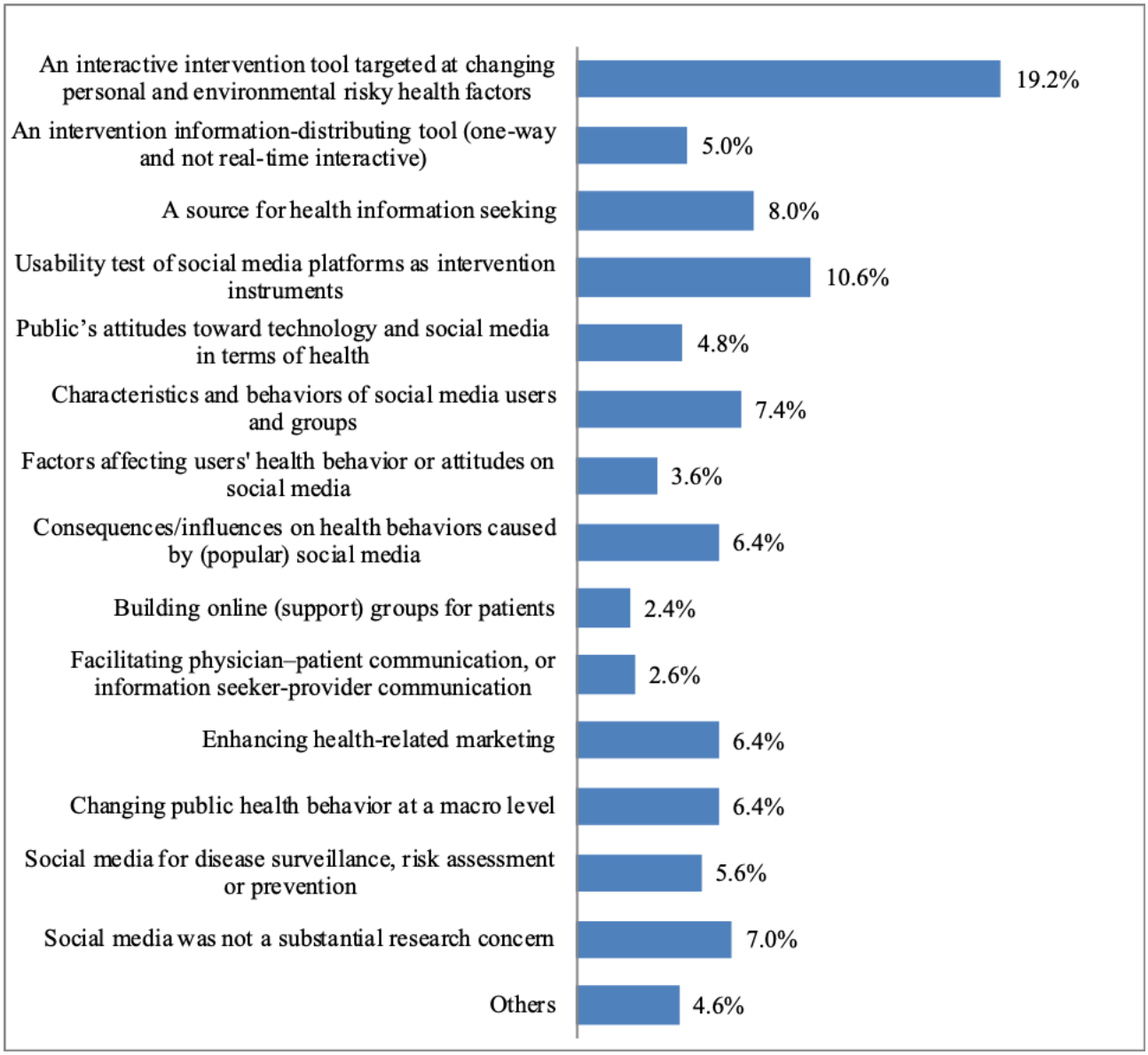
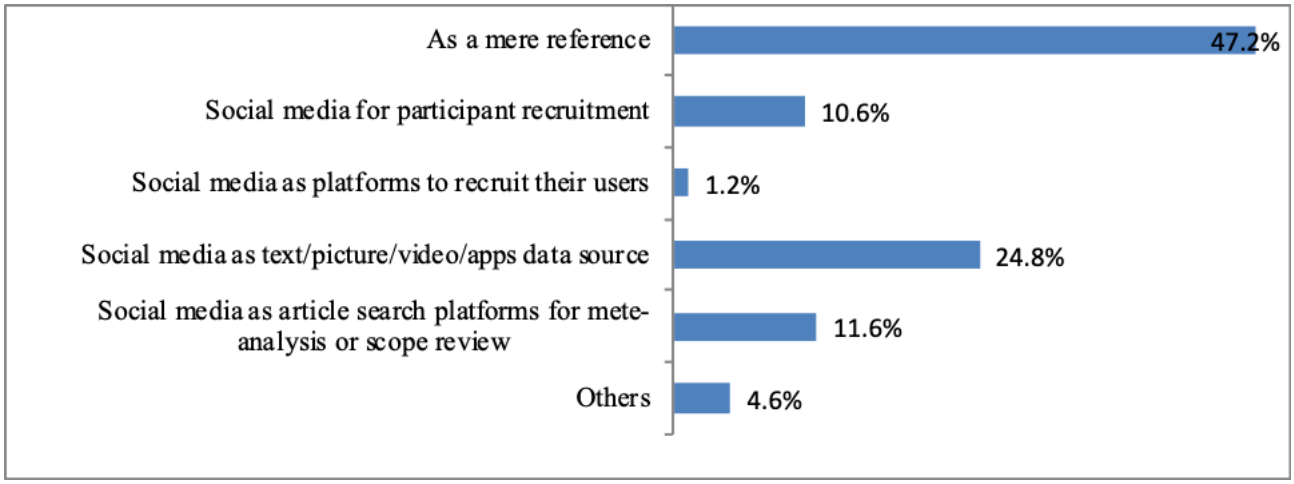


Figure 5. Article distributions based on social media as research context.

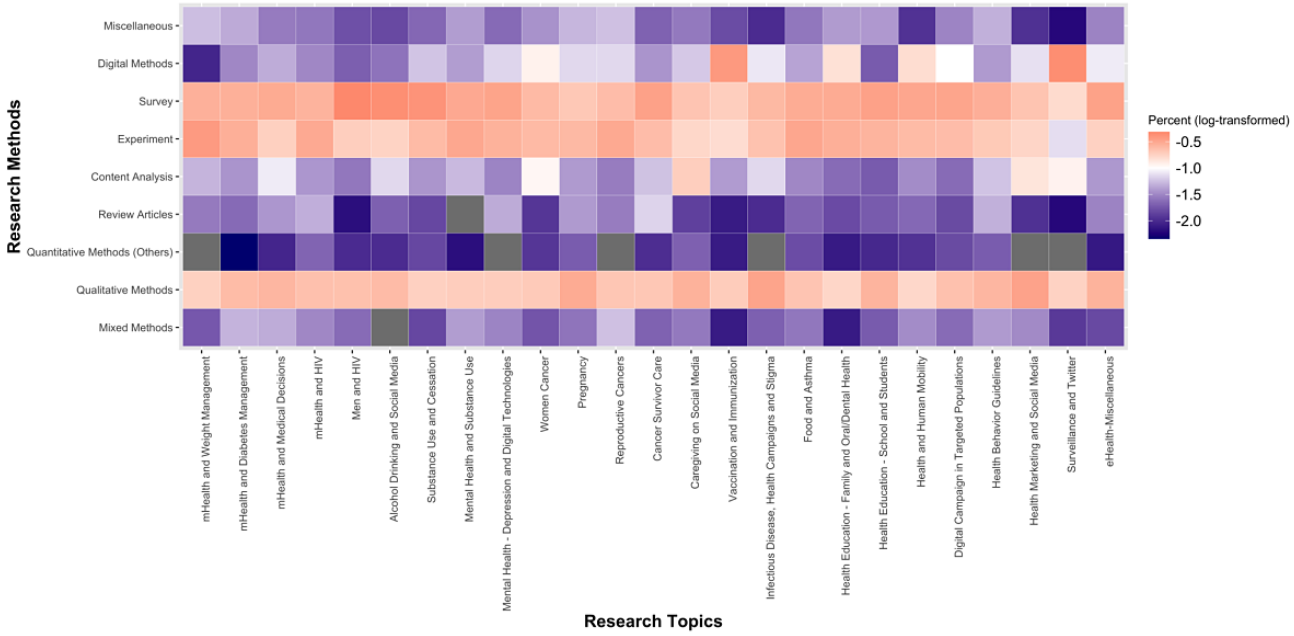


Research Methods in Social Media–Based Public Health Research

Public health research with social media data was dominated by traditional quantitative research methods, whereas cutting edge computational methods played a minor role. Among all the articles, 30.6% employed survey method, 24.0% employed experiment design, 22.7% employed qualitative methods (eg, field observation, in-depth interview, and focus group), 8.3% included employed digital methods (including digital tracks analysis and computational methods, such as text mining, sentiment analysis, agent-based modeling, and network modeling), and 5.6% employed traditional content analysis.

Figure 6 demonstrates that the method distributions under the 25 research themes were similar to the general distribution among the whole body of the studies. Survey and experiment were the two most adopted methods, whereas review article number was relatively small among all themes.

Figure 6. Research methods adopted under research topics.



Intercoder reliabilities were over .80 for both dimensions before each coder independent coding of half items of the randomly extracted sample.

Discussion

Principal Findings

With a bottom-up approach, this study provided a panoramic mapping of the landscape of social media–based public health research. By analyzing publication trends, research themes, roles of social media, and research methods adopted in this emerging research area, this study concluded that (1) social media has penetrated almost all the health-related processes and domains since 2010, showing a dramatic increase in the research body; (2) existing social media–based public health research mainly focuses on 25 themes in 6 clusters; (3) social media generally played two roles in public health research: generating substantial research interest and providing a research context/platform; (4) existing social media–based public health research is dominated by traditional research methods while the share of computational method is on the rise. The panoramic mapping can help scholars understand the state of the art in this research area and what is under- or overstudied in this field. This study can enable scholars across various disciplines to understand each other's needs and contribute jointly to health promotion and disease control. Here, three notable issues that possess theoretical and methodological implications in social media–based public health research are elaborated.

When Public Health Research Meets Social Media: From New Phenomena to New Questions

Social media has infiltrated almost all health-related processes and domains with the rapid advancement of social and mobile media. This dramatic change has entailed many new phenomena to be explored in public health research. The findings of this study are consistent with previous studies that found that almost one-third of internet studies have focused on eHealth and mHealth since 2009 [29] and a trend toward digitization exists in health care [31]. Many traditional public health activities, such as health education, health promotion, and disease surveillance, have taken advantage of social media technologies to become digitized [32-34]. Social media has substantially altered how individuals seek and share health information, discuss health issues, and engage in health behaviors [35]. Social media also provides innovative ways to change health behaviors in various domains, such as smoking cessation, substance use, weight control, HIV prevention, and cancer screening [36-38]. Consistent with previous reviews on social media and public health studies, this study concludes that social media contributes to these public health domains by broadening the reach of health education, providing accessible online professional consultation, and improving the efficacy of access to care and medication uptake, etc [19,39]. Moreover, an upward trend of integrating social media in various public health campaigns exists due to

the instrumental benefits of social media technologies, such as lower intervention cost, higher user engagement, higher efficiency, and better documentation of the process [40].

When public health research meets social media, new topics have emerged and attracted the attention of public health scholars [41]: mHealth and social media–empowered health research. "Digital campaigns in targeted populations" and "surveillance and Twitter" are typical new topics where researchers frequently examine new research questions [42,43]. For example, researchers discuss how to employ user-generated content together with geolocation information to predict an outbreak of an emerging disease or visually map their diffusion routes and locate the risky population [44]. The digital trace on social and mobile media offers many possibilities to study online health behaviors such as online health information–seeking, online social support, and online medical consultation behaviors [45]. In addition, some health topics have attracted burgeoning attention in the era of social media. For instance, mental health problems have been identified as significant concerns among the 25 themes in this study. However, no conclusion has yet been reached whether and how the adoption and use of social media alleviates or exacerbates mental health problems [46].

This relatively new domain calls for in-depth exploration. New phenomena and new questions raised by social media are of practical and theoretical significance for public health research. Timely responses to those new phenomena via scientific research can promote the advancement of the domain to keep up with technology advances and establish a realistic understanding of what social media can and cannot do in public health. Meanwhile, public health research should delve into scientific research questions behind those new phenomena and address them either by exploiting the existing body of knowledge or exploring new methods and knowledge to extend the domain.

When Public Health Research Meets Social Media: Methodological Potential to Be Further Tapped

When public health research meets social media, the dominant research methods are traditional quantitative methods, despite the growing interest in computational methods among public health scholars. The methodological potentials of social media for public health research can and should be further tapped. Specifically, the potential of social media in participant recruitment and measurement development has direct and salient implications to public health research.

Social media substantially facilitates participant recruitment in public health research. Recruiting research participants from specific groups of individuals who have sensitive health issues or are stigmatized in society, such as people living with HIV/AIDS or individuals with mental health issues, remains a significant challenge for public health scholars [19,47]. Given the size and heterogeneity of social media users, recruiting a fairly sizable number of subjects from particular social groups to participate in public health surveys and experiments should be possible. More importantly, participants recruited from online platforms such as Facebook and Amazon's Mechanical Turk can have significant heterogeneity in their demographic characteristics (eg, age, gender, race, cultural background) and other key variables relevant to specific researcher contexts [48]. Nevertheless, it is worth noting here that the representativeness of participants recruited on social media needs to be empirically evaluated in particular contexts. Amazon's Mechanical Turk workers are not a generalizable population with regard to health status and behaviors in the United States [49]. Without an empirical evaluation of representativeness of recruited subjects on social media, researchers should be cautious in the generalizability of their research findings. Moreover, ethical issues involved in participant recruitment via social media platforms have become more prominent and challenging. Due to the anonymity of social media users, it is extremely difficult if not impossible to obtain informed consent beforehand from recruited participants. When users of a social media platform accept the terms of service of the platform, can researchers assume that the users have given an explicit or implicit consent to participate in any type of experiment or intervention conducted on the platform [50]? We do not have a widely accepted ethical guideline in this regard. A collective effort from the scientific community is needed to outline responsible and ethical conduct in this emerging research area.

Social media contributes to public health research by providing refreshed measurements of existing concepts or new observations of emerging phenomena. Rich semantic information in digital traces can provide a social telescope [51] with which to observe or infer what health information is produced, shared, and consumed by ordinary users. Multiple social and interactive relations in digital traces facilitate empirical studies on who connects with whom in various contexts. Voluminous and real-time social media data have been widely employed for epidemic surveillance or tracking emotional contagion [52,53]. A growing number of studies have employed user-generated content on social media to monitor emerging diseases at the breaking-out stage to minimize consequences or track trends in public health issues [54-56]. When public health scholars embrace new measures derived from social media data, empirically assessing and monitoring the quality of the new measures by cross-validating them with established measures is necessary. The parable of Google Flu Trends well illustrates the necessity of such cross-validation. When Google Flu Trends was first released, it outperformed traditional flu surveillance measures adopted by the US Centers for Disease Control and Prevention [57]. However, Google Flu Trends is reported to overestimate flu cases in the United States [58]. Validation of empirical measures is an ongoing process in public health research and beyond [59].

When Public Health Research Meets Social Media: Unequal Status With Detached Concerns?

Social media–based public health research lies in the crossroad between public health studies and social science studies on information and communication technologies (ICTs) [60] and benefits from both perspectives. In the cooperative process, social media–based public health research reaches various levels in elaborating on the two perspectives. Taking the initial perspective of public health interest, many acceptability studies and randomized controlled trials have been documented to examine the effectiveness of social media to reach different public health goals [40,61]. In these studies, social media is

often considered a new functional tool to improve public health. Meanwhile, in research that further examines the influence of ICTs on public health, who used what social media content targeted at whom through which social media platforms with what health effects is the core concern [62]. In this line of research, studies typically focus on the transmission of health information, communication between health agencies, the uses of health apps, and so on [63]. The inherent concerns of these studies seem to be detached though not in conflict in that social media facilitates the public health promotion process, and public health outcomes add value to the communication through social media.

From an overview of social media–based public health research, the dominating approach of these published studies considers public health issues as the substantial interests and ultimate outcomes rather than regard social media as an equally important area of concern. Many articles used limited space to describe the use of social media in health promotion campaigns or projects [64,65]. The subordinate role of social media suggests that the potential of ICTs has not been fully realized in the domain of public health [41]. Empirical studies should not only focus on what social media can contribute to public health research but should also examine how and why social media can make an impact in various contexts of public health research. This can substantially improve the understanding of the intended as well as unintended consequences social media can exert on health attitudes and behaviors. This can also enable public health researchers to integrate social media into their research design further.

Limitations

Despite the strengths and contributions, this study has certain limitations. First, the study may suffer from the file drawer effect given that only studies indexed in Web of Science and PubMed were included. Empirical studies published in other outlets were not considered here. Future studies are warranted to expand the pools to conference proceedings and articles indexed in other databases. Second, this study used numerous diseases as search terms in the initial search, but the list remains incomplete. Some important diseases, such as mental disorders, were not incorporated. Despite this, the topic modeling captured mental health as a major theme. Further research is suggested to include mental health keywords as search terms. Third, LDA topic modeling is a well-recognized method to identify related themes through document-word matrices. However, the results of the topic modeling were not as neat as expected. No standard and quantitative thresholds exist for researchers to choose the optimal number of topics. Future studies are encouraged to replicate this study and examine the reliability of such themes.

Conclusions

This study examined research themes, roles of social media, and research methods in social media–based public health research published from 2000 to 2018. This research identifies 25 research themes covering different diseases, various population groups, physical and mental health topics, and other significant issues. Social media assumes two major roles in public health research: one is to produce substantial research interest for public health research and the other is to furnish a research context for public health research. Social media enables scholars to study new phenomena and propose new research questions in public health research. Meanwhile, the methodological potential of social media in public health research needs further exploration.

Acknowledgments

This study was supported by National Social Science Fund of China (19ZDA324 and 18CXW017), Junior Faculty Supporting Fund at Shenzhen University (QNFC1903), and a research grant awarded to TQP at Michigan State University (RC109127).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Social media roles and definitions.
[DOCX File , 20 KB - jmir_v22i8e17582_app1.docx]

Multimedia Appendix 2
Network graph of topic-word probability of the 25 research themes.
[PNG File , 648 KB - jmir_v22i8e17582_app2.png]

Multimedia Appendix 3
Representative articles of the 25 research themes.
[XLSX File (Microsoft Excel File), 36 KB - jmir_v22i8e17582_app3.xlsx]

References

1. Tran BX, Mai HT, Nguyen LH, Nguyen CT, Latkin CA, Zhang MWB, et al. Vietnamese validation of the short version of Internet Addiction Test. Addict Behav Rep 2017 Dec;6:45-50 [FREE Full text] [doi: 10.1016/j.abrep.2017.07.001] [Medline: 29450235]
2. Bernard S, Cooke T, Cole T, Hachani L, Bernard J. Quality and readability of online information about type 2 diabetes and nutrition. JAAPA 2018 Nov;31(11):41-44. [doi: 10.1097/01.JAA.0000546481.02560.4e] [Medline: 30358679]
3. Crawford R, Rutz D, Evans D. Between Combat boots and Birkenstocks: Lessons from HIV/AIDS, SARS, H1N1 and Ebola. Public Health 2016 Dec;141:186-191. [doi: 10.1016/j.puhe.2016.09.018] [Medline: 27931998]
4. Fitzpatrick T, Zhou K, Cheng Y, Chan P, Cui F, Tang W, et al. A crowdsourced intervention to promote hepatitis B and C testing among men who have sex with men in China: study protocol for a nationwide online randomized controlled trial. BMC Infect Dis 2018 Sep 29;18(1):489 [FREE Full text] [doi: 10.1186/s12879-018-3403-3] [Medline: 30268114]
5. Roberts CA, Geryk LL, Sage AJ, Sleath BL, Tate DF, Carpenter DM. Adolescent, caregiver, and friend preferences for integrating social support and communication features into an asthma self-management app. J Asthma 2016 Nov;53(9):948-954. [doi: 10.3109/02770903.2016.1171339] [Medline: 27116540]
6. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16 Suppl 5:290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
7. Almutairi N, Alhabash S, Hellmueller L, Willis E. The effects of Twitter users' gender and weight on viral behavioral intentions toward obesity-related news. J Health Commun 2018;23(3):233-243. [doi: 10.1080/10810730.2018.1423648] [Medline: 29388884]
8. Bogaerts A, Ameye L, Bijlholt M, Amuli K, Heynickx D, Devlieger R. INTER-ACT: prevention of pregnancy complications through an e-health driven interpregnancy lifestyle intervention: study protocol of a multicentre randomised controlled trial. BMC Pregnancy Childbirth 2017 May 26;17(1):154 [FREE Full text] [doi: 10.1186/s12884-017-1336-2] [Medline: 28549455]
9. Lennox J, Emslie C, Sweeting H, Lyons A. The role of alcohol in constructing gender and class identities among young women in the age of social media. Int J Drug Pol 2018 Aug;58:13-21. [doi: 10.1016/j.drugpo.2018.04.009]
10. Zhang MWB, Ho RCM, Hawa R, Sockalingam S. Analysis of the information quality of bariatric surgery smartphone applications using the Silberg scale. Obes Surg 2016 Jan;26(1):163-168. [doi: 10.1007/s11695-015-1890-5] [Medline: 26424704]
11. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May;22(5):413-432. [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
12. Vorderstrasse A, Lewinski A, Melkus GD, Johnson C. Social support for diabetes self-management via eHealth interventions. Curr Diab Rep 2016 Dec;16(7):56. [doi: 10.1007/s11892-016-0756-0] [Medline: 27155606]
13. Menefee HK, Thompson MJ, Guterbock TM, Williams IC, Valdez RS. Mechanisms of communicating health information through Facebook: implications for consumer health information technology design. J Med Internet Res 2016 Aug 11;18(8):e218 [FREE Full text] [doi: 10.2196/jmir.5949] [Medline: 27515151]
14. Dehlin JM, Stillwagon R, Pickett J, Keene L, Schneider JA. #PrEP4Love: an evaluation of a sex-positive HIV prevention campaign. JMIR Public Health Surveill 2019 Jun 17;5(2):e12822 [FREE Full text] [doi: 10.2196/12822] [Medline: 31210141]
15. Kass-Hout TA, Alhinnawi H. Social media in public health. Br Med Bull 2013;108:5-24. [doi: 10.1093/bmb/ldt028] [Medline: 24103335]
16. Thackeray R, Neiger BL, Smith AK, Van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012;12:242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]
17. Zhang MW, Ho CS, Fang P, Lu Y, Ho RC. Usage of social media and smartphone application in assessment of physical and psychological well-being of individuals in times of a major air pollution crisis. JMIR Mhealth Uhealth 2014 Mar 25;2(1):e16 [FREE Full text] [doi: 10.2196/mhealth.2827] [Medline: 25098255]
18. Giustini D, Ali SM, Fraser M, Kamel Boulos MN. Effective uses of social media in public health and medicine: a systematic review of systematic reviews. Online J Public Health Inform 2018;10(2):e215 [FREE Full text] [doi: 10.5210/ojphi.v10i2.8270] [Medline: 30349633]
19. Taggart T, Grewe ME, Conserve DF, Gliwa C, Roman IM. Social media and HIV: a systematic review of uses of social media in HIV communication. J Med Internet Res 2015 Nov 02;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]
20. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
21. Williams G, Hamm MP, Shulhan J, Vandermeer B, Hartling L. Social media interventions for diet and exercise behaviours: a systematic review and meta-analysis of randomised controlled trials. BMJ Open 2014;4(2):e003926 [FREE Full text] [doi: 10.1136/bmjopen-2013-003926] [Medline: 24525388]
22. Swanton R, Allom V, Mullan B. A meta-analysis of the effect of new-media interventions on sexual-health behaviours. Sex Transm Infect 2015 Feb;91(1):14-20. [doi: 10.1136/sextrans-2014-051743] [Medline: 25433051]

23. Odone A, Ferrari A, Spagnoli F, Visciarelli S, Shefer A, Pasquarella C, et al. Effectiveness of interventions that apply new media to improve vaccine uptake and vaccine coverage. Hum Vaccin Immunother 2015;11(1):72-82 [FREE Full text] [doi: 10.4161/hv.34313] [Medline: 25483518]
24. Mita G, Mhurchu CN, Jull A. Effectiveness of social media in reducing risk factors for noncommunicable diseases: a systematic review and meta-analysis of randomized controlled trials. Nutr Rev 2016 Apr;74(4):237-247 [FREE Full text] [doi: 10.1093/nutrit/nuv106] [Medline: 26946250]
25. Cartledge P, Miller M, Phillips B. The use of social-networking sites in medical education. Med Teach 2013 Oct;35(10):847-857. [doi: 10.3109/0142159X.2013.804909] [Medline: 23841681]
26. Willis EA, Szabo-Reed AN, Ptomey LT, Steger FL, Honas JJ, Washburn RA, et al. Do weight management interventions delivered by online social networks effectively improve body weight, body composition, and chronic disease risk factors? A systematic review. J Telemed Telecare 2017 Feb;23(2):263-272. [doi: 10.1177/1357633X16630846] [Medline: 26880695]
27. Tran BX, Nghiem S, Sahin O, Vu TM, Ha GH, Vu GT, et al. Modeling research topics for artificial intelligence applications in medicine: latent Dirichlet allocation application study. J Med Internet Res 2019 Nov 01;21(11):e15511 [FREE Full text] [doi: 10.2196/15511] [Medline: 31682577]
28. WHO publishes list of top emerging diseases likely to cause major epidemics. World Health Organization. 2017. URL: https://www.who.int/medicines/ebola-treatment/WHO-list-of-top-emerging-diseases/en/ [accessed 2020-08-03]
29. Peng T, Zhang L, Zhong Z, Zhu JJ. Mapping the landscape of internet studies: text mining of social science journal articles 2000–2009. New Media Soc 2012 Nov 26;15(5):644-664. [doi: 10.1177/1461444812462846]
30. Blei D, Ng A, Jordan M. Latent Dirichlet allocation. J Mach Learn Res 2003:1. [doi: 10.1016/b978-0-12-411519-4.00006-9]
31. Agarwal R, Gao G, DesRoches C, Jha AK. The digital transformation of healthcare: current status and the road ahead. Inf Sys Res 2010 Dec;21(4):796-809. [doi: 10.1287/isre.1100.0327]
32. Paton C, Bamidis PD, Eysenbach G, Hansen M, Cabrer M. Experience in the use of social media in medical and health education. Contribution of the IMIA Social Media Working Group. Yearb Med Inform 2011;6:21-29. [Medline: 21938320]
33. Cao B, Liu C, Durvasula M, Tang W, Pan S, Saffer AJ, et al. Social media engagement and HIV testing among men who have sex with men in China: a nationwide cross-sectional survey. J Med Internet Res 2017 Jul 19;19(7):e251 [FREE Full text] [doi: 10.2196/jmir.7251] [Medline: 28724510]
34. Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, et al. Tracking suicide risk factors through Twitter in the US. Crisis 2014;35(1):51-59. [doi: 10.1027/0227-5910/a000234] [Medline: 24121153]
35. Chen X, Hao T. Quantifying and visualizing the research status of social media and health research field. Soc Web Heal Res 2019:31-51. [doi: 10.1007/978-3-030-14714-3_3]
36. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
37. Tucker J, Cao B, Li H, Tang S, Tang W, Wong N, et al. Social media interventions to promote HIV testing. Clin Infect Dis 2016 Jul 15;63(2):282-283 [FREE Full text] [doi: 10.1093/cid/ciw297] [Medline: 27161778]
38. Chang T, Chopra V, Zhang C, Woolford S. The role of social media in online weight management: systematic review. J Med Internet Res 2013 Nov 28;15(11):e262 [FREE Full text] [doi: 10.2196/jmir.2852] [Medline: 24287455]
39. Patel R, Chang T, Greysen SR, Chopra V. Social media use in chronic disease: a systematic review and novel taxonomy. Am J Med 2015 Dec;128(12):1335-1350. [doi: 10.1016/j.amjmed.2015.06.015] [Medline: 26159633]
40. Tang W, Wei C, Cao B, Wu D, Li K, Lu H, et al. Crowdsourcing to expand HIV testing among men who have sex with men in China: a closed cohort stepped wedge cluster randomized controlled trial. PLoS Med 2018 Aug;15(8):e1002645 [FREE Full text] [doi: 10.1371/journal.pmed.1002645] [Medline: 30153265]
41. Moorhead S, Hazlett D, Harrison L, Carroll J, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
42. Dowshen N, Lee S, Matty LB, Castillo M, Mollen C. IknowUshould2: feasibility of a youth-driven social media campaign to promote STI and HIV testing among adolescents in Philadelphia. AIDS Behav 2015 Jun;19 Suppl 2:106-111 [FREE Full text] [doi: 10.1007/s10461-014-0991-9] [Medline: 25563502]
43. Vijaykumar S, Nowak G, Himelboim I, Jin Y. Virtual Zika transmission after the first U.S. case: who said what and how it spread on Twitter. Am J Infect Control 2018 Jan 04:1. [doi: 10.1016/j.ajic.2017.10.015] [Medline: 29306490]
44. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EHY, Olsen JM, et al. Using social media for actionable disease surveillance and outbreak management: a systematic literature review. PLoS One 2015;10(10):e0139701 [FREE Full text] [doi: 10.1371/journal.pone.0139701] [Medline: 26437454]
45. Young SD, Zhang Q. Using search engine big data for predicting new HIV diagnoses. PLoS One 2018;13(7):e0199527 [FREE Full text] [doi: 10.1371/journal.pone.0199527] [Medline: 30001360]
46. Lee K, Noh M, Koo D. Lonely people are no longer lonely on social networking sites: the mediating role of self-disclosure and social support. Cyberpsychol Behav Soc Netw 2013 Jun;16(6):413-418. [doi: 10.1089/cyber.2012.0553] [Medline: 23621716]
47. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci 2017 Dec;18:43-49. [doi: 10.1016/j.cobeha.2017.07.005]

48. Peng T, Liang H, Zhu JJ. Introducing computational social science for Asia-Pacific communication research. Asian J Comm 2019 Apr 16;29(3):205-216. [doi: 10.1080/01292986.2019.1602911]
49. Walters K, Christakis DA, Wright DR. Are Mechanical Turk worker samples representative of health status and health behaviors in the U.S.? PLoS One 2018;13(6):e0198835 [FREE Full text] [doi: 10.1371/journal.pone.0198835] [Medline: 29879207]
50. van Atteveldt W, Peng T. When communication meets computation: opportunities, challenges, and pitfalls in computational communication science. Commun Methods Meas 2018;12(2-3):81-92. [doi: 10.1080/19312458.2018.1458084]
51. Golder SA, Macy MW. Digital footprints: opportunities and challenges for online social research. Annu Rev Sociol 2014 Jul 30;40(1):129-152. [doi: 10.1146/annurev-soc-071913-043145]
52. Dreisbach C, Koleck TA, Bourne PE, Bakken S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. Int J Med Inform 2019 May;125:37-46 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.02.008] [Medline: 30914179]
53. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. Demographic-based content analysis of web-based health-related social media. J Med Internet Res 2016 Jun 13;18(6):e148 [FREE Full text] [doi: 10.2196/jmir.5327] [Medline: 27296242]
54. Amankwah-Amoah J. Emerging economies, emerging challenges: mobilising and capturing value from big data. Technol Forecast Soc Change 2016 Sep;110:167-174. [doi: 10.1016/j.techfore.2015.10.022]
55. Huang D, Wang J. Monitoring hand, foot and mouth disease by combining search engine query data and meteorological factors. Sci Total Environ 2018 Jan 15;612:1293-1299. [doi: 10.1016/j.scitotenv.2017.09.017] [Medline: 28898935]
56. Ling R, Lee J. Disease monitoring and health campaign evaluation using Google search activities for HIV and AIDS, stroke, colorectal cancer, and marijuana use in Canada: a retrospective observational study. JMIR Public Health Surveill 2016 Oct 12;2(2):e156 [FREE Full text] [doi: 10.2196/publichealth.6504] [Medline: 27733330]
57. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
58. Lazer D, Kennedy R, King G, Vespignani A. Google flu trends still appears sick: an evaluation of the 2013-2014 flu season. SSRN Journal 2014:1. [doi: 10.2139/ssrn.2408560]
59. Anastasi A. Evolving concepts of test validation. Ann Rev Psychol 1986 Jan;37(1):1-16. [doi: 10.1146/annurev.ps.37.020186.000245]
60. Driouchi AA. ICTs for Health, Education and Socioeconomics: Regional Cases. Hershey: IGI GLobal; 2013.
61. Lorig K, Ritter PL, Villa FJ, Armas J. Community-based peer-led diabetes self-management: a randomized trial. Diabetes Educ 2009;35(4):641-651. [doi: 10.1177/0145721709335006] [Medline: 19407333]
62. Wenxiu P. Analysis of new media communication based on Lasswell's 5W model. J Educ Soc Res 2015 Sep 1:1. [doi: 10.5901/jesr.2015.v5n3p245]
63. Kim J, Park S, Yoo S, Shen H. Mapping health communication scholarship: breadth, depth, and agenda of published research in Health Communication. Health Commun 2010 Sep;25(6-7):487-503. [doi: 10.1080/10410236.2010.507160] [Medline: 20845126]
64. Carroll JK, Tobin JN, Luque A, Farah S, Sanders M, Cassells A, et al. "Get Ready and Empowered About Treatment" (GREAT) study: a pragmatic randomized controlled trial of activation in persons living with HIV. J Gen Intern Med 2019 Sep;34(9):1782-1789. [doi: 10.1007/s11606-019-05102-7] [Medline: 31240605]
65. Narang B, Park S, Norrmén-Smith IO, Lange M, Ocampo AJ, Gany FM, et al. The use of a mobile application to increase access to interpreters for cancer patients with limited English proficiency: a pilot study. Med Care 2019 Jun;57 Suppl 6 Suppl 2:S184-S189. [doi: 10.1097/MLR.0000000000001035] [Medline: 31095059]

Abbreviations

ICT: information and communication technology
LDA: latent Dirichlet allocation
MERS: Middle East respiratory syndrome

Edited by G Eysenbach; submitted 22.12.19; peer-reviewed by S Ali, R Ho, A Lee, T Steeb; comments to author 23.01.20; revised version received 12.05.20; accepted 25.07.20; published 13.08.20.

Please cite as:
Zhang Y, Cao B, Wang Y, Peng TQ, Wang X
When Public Health Research Meets Social Media: Knowledge Mapping From 2000 to 2018
J Med Internet Res 2020;22(8):e17582
URL: http://www.jmir.org/2020/8/e17582/
doi:10.2196/17582
PMID:32788156

Original Paper

Topic Modeling of Social Networking Service Data on Occupational Accidents in Korea: Latent Dirichlet Allocation Analysis

Kyoung-Bok Min[1], PhD; Sung-Hee Song[1], BA; Jin-Young Min[2], PhD

[1]College of Medicine, Seoul National University, Seoul, Republic of Korea

[2]Institute of Health and Environment, Seoul National University School of Public Health, Seoul, Republic of Korea

Corresponding Author:
Jin-Young Min, PhD
Institute of Health and Environment
Seoul National University School of Public Health
Gwanak-ro 1, Gwanak-gu
Seoul, 08826
Republic of Korea
Phone: 82 8802759
Email: yaemin00@snu.ac.kr

Abstract

Background: In most industrialized societies, regulations, inspections, insurance, and legal options are established to support workers who suffer injury, disease, or death in relation to their work; in practice, these resources are imperfect or even unavailable due to workplace or employer obstruction. Thus, limitations exist to identify unmet needs in occupational safety and health information.

Objective: The aim of this study was to explore hidden issues related to occupational accidents by examining social network services (SNS) data using topic modeling.

Methods: Based on the results of a Google search for the phrases occupational accident, industrial accident and occupational diseases, a total of 145 websites were selected. From among these websites, we collected 15,244 documents on queries related to occupational accidents between 2002 and 2018. To transform unstructured text into structure data, natural language processing of the Korean language was conducted. We performed the latent Dirichlet allocation (LDA) as a topic model using a Python library. A time-series linear regression analysis was also conducted to identify yearly trends for the given documents.

Results: The results of the LDA model showed 14 topics with 3 themes: workers' compensation benefits (Theme 1), illicit agreements with the employer (Theme 2), and fatal and non-fatal injuries and vulnerable workers (Theme 3). Theme 1 represented the largest cluster (52.2%) of the collected documents and included keywords related to workers' compensation (ie, company, occupational injury, insurance, accident, approval, and compensation) and keywords describing specific compensation benefits such as medical expense benefits, temporary incapacity benefits, and disability benefits. In the yearly trend, Theme 1 gradually decreased; however, other themes showed an overall increasing pattern. Certain queries (ie, musculoskeletal system, critical care, and foreign workers) showed no significant variation in the number of queries.

Conclusions: We conducted LDA analysis of SNS data of occupational accident–related queries and discovered that the primary concerns of workers posting about occupational injuries and diseases were workers' compensation benefits, fatal and non-fatal injuries, vulnerable workers, and illicit agreements with employers. While traditional systems focus mainly on quantitative monitoring of occupational accidents, qualitative aspects formulated by topic modeling from unstructured SNS queries may be valuable to address inequalities and improve occupational health and safety.

(J Med Internet Res 2020;22(8):e19222) doi:10.2196/19222

KEYWORDS

topic modeling; occupational accident; social media; knowledge; workplace; accident; model; analysis; safety

Introduction

Occupational health and safety are fundamental components of a good work environment [1]. They are relevant not only to increased productivity but also to the moral and legal responsibilities of both employees and employers [1]. Although many strategies and programs promote occupational safety and health, fatal and nonfatal accidents in the work environment remain a global problem [1,2]. Occupational accidents include any injury, disease, or death arising through the course of employment [2]. The International Labor Organization estimates that 2.78 million workers die each year from occupational injuries, and 374 million workers suffer nonfatal work-related injuries and illnesses [2].

Occupational accident statistics vary from country to country due to differences in coverage, definitions, and classifications; data divided into fatal and nonfatal rates of occupational accidents provide similar perspectives on the risks related to occupational safety and health [2]. The paradox of low levels of nonfatal injuries but high rates of fatal accidents is observed in many countries, including South Korea [3,4]. In 2017, the rate of fatal occupational injuries reached 1.12 per 10,000 Korean workers; however, the rate of nonfatal occupational injuries was 0.54 per 100 workers [5]. This paradox may be rooted in underreporting or covering up of nonfatal occupational accidents to avoid reprisal or stigma at work or due to a perception that an injury is minor or simply part of the job [3,6]. Thus, conventional record keeping of occupational accidents may simply fail to provide the actual occupational injury rates. Furthermore, there are limitations to identifying unmet needs in occupational safety and health information, resources, and interventions for injured workers.

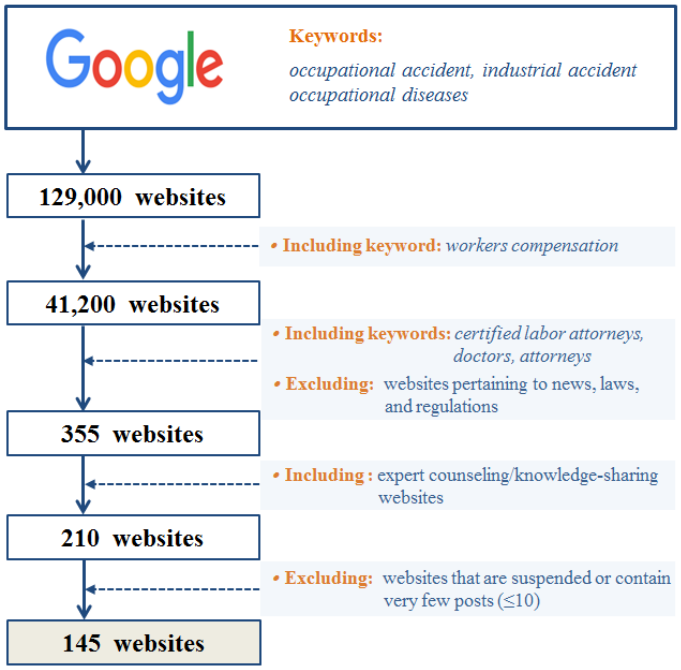
Due to the rapid spread of smart devices and mobile internet services, social network services (SNSs) are changing the ways in which people interact with other people who have similar interests, thoughts, services, or desires [7-9]. SNS users can also engage in knowledge sharing about specific topics (ie, academic, medical, and legal issues) [7-9]. People with work-related injuries or illnesses differ from other SNS users in how they seek professional advice and information over the internet. Employers sometimes present obstacles to injured workers or minimize their benefits, and the process of approving workers' compensation benefits is time-consuming and frustrating. Therefore, it is inevitable that injured workers will seek legal and technical information and discuss their challenges through SNSs.

We believe that SNS data do not show the entire picture of occupational accidents but are useful for discovering hidden aspects of these accidents that are not captured by conventional surveillance. In this study, by applying topic modeling techniques to SNS data, we explored issues regarding occupational accidents that are not shared or discussed publicly but are nonetheless important.

Methods

The data upon which our study is based were obtained using Google's Korean-language search engine (Figure 1). Using the results of a Google search for the phrases *occupational accident*, *industrial accident*, and *occupational diseases*, a total of 129,000 webpages were initially selected. Next, 87,800 webpages were excluded from the initial sites by applying the keyword *workers compensation*. We further filtered the keywords *certified labor attorneys*, *doctors*, and *other attorneys*, and we excluded webpages corresponding to news, laws, and regulations. Ultimately, 355 websites were identified after applying the filter. We subsequently narrowed down our selection to 210 expert counseling and knowledge-sharing websites on which injured workers and their families could discuss difficulties related to treatment and compensation after occupational accidents and could seek answers and advice from experts. Finally, 145 websites were deemed eligible for use in the current study, excluding those that were suspended or contained fewer than 10 posts. We used web crawlers to gather website data from between 2002 and 2018; a total of 23,076 documents were collected from 145 websites. These documents were composed of full-text documents (articles or posts), comments, and blogs. From among these websites, 7832 duplicate documents were excluded. Finally, 15,244 documents were subjected to further analysis.

Figure 1. Flowchart of the identification of 145 websites (as of May 2019).



Prior to topic modeling analysis, the various documents were transformed into a structured form by text preprocessing. Natural language processing (NLP) was performed using KoNLPy, which is an open-source morphological analyzer for Korean processes based on the Python package [10]. Lexical analysis was performed by sentence splitting followed by tokenizing. We removed unnecessary components such as unnecessary white spaces, punctuation, special characters, and stop words such as “a,” “the,” and “it” from the unstructured data. To convert the sentences from the documents into words, tokenization was conducted using MeCab [11], an open-source morphological analyzer and part-of-speech tagger modified for the Korean language.

With keywords created through NLP, a term-document matrix was created to evaluate the importance of a word in a document. We used the term frequency–inverse document frequency (TF-IDF) technique [12], in which the term frequency measures how often a word appears in a given document divided by the total number of words while the inverse document frequency measures how frequently a word appears in all documents divided by the number of documents that contain the word. Thus, a word with a high TF-IDF score is distinctly frequent in a given document compared with other documents in the set. In this way, a term-document matrix of 35,315 words was generated from 15,244 documents. We created a co-occurrence network for high-frequency words in the given documents.

Topic modeling is a machine learning technique that is used to determine the abstract topics discussed in a given text [13]. Topic modeling is recognized as a standard methodology with high performance and convenience; it has been suggested as an alternative method of solving problems of rareness, synonyms, multiplicity, and semantic hierarchy that occur in existing word frequency analyses [13]. Latent Dirichlet allocation (LDA), a generative probabilistic model of a corpus, is a commonly used topic model. LDA assumes sparse Dirichlet prior distributions, encoding the intuition that the probability distribution of words in a topic is skewed so that only a small set of words have high probability [14,15]. In a collection of documents (D) including a word (W) and a preselected number of topics (K), LDA calculates two probabilities: the probability of words in document d assigned to topic t, P(t|d), and the probability of topic t in all the documents in a set for word w, P(w|t). The prior distributions of P(t|d) and P(w|t) are defined by the hyperparameters α and β. Gibbs sampling is used to assess distribution over topics and distribution over words for each document [13,16]. We used LDA in the Gensim library in Python for topic modeling. We set λ as 0.6 and ran the LDA with 3000 Gibbs sampling iterations. The optimal number of topics was based on perplexity and coherence (Supplemental Figure 1 in Multimedia Appendix 1). Perplexity is a method of evaluating how well a probability distribution can predict a held-out sample [17]. The smaller the change in the perplexity value, the better the probabilistic model. Topic coherence measures the score of a single topic by calculating the degree of similarity between the top N words of a topic [18]. The higher the score, the easier it is to choose the appropriate number of topics. Because there are variations in the interpretation of the quality of topics among perplexity, coherence, and human judgment [15], we further evaluated the manually formulated topics by varying the number of topics of the LDA model. Finally, we determined that 14 was a reasonable number of topics to discover hidden structures in the text body. Each topic was represented by a set of several keywords and was named by the authors of this report, who included a physician specializing in occupational medicine. These topics were categorized into three themes based on the LDA plot and the authors’ interpretation. For each topic, we performed time-series linear regression analysis with SAS statistical software (SAS Institute) by using the AUTOREG procedure to identify trends by year. The years were taken as independent variables, and the dependent variables were the average weight values for each

topic by year. We classified the topics as "hot" if the regression coefficient was positive or as "cold" if the coefficient was negative, taking a 5% significance level.

Results

Table 1 shows the numbers of documents related to occupational accidents collected on the internet by year. From the 145 websites, a total of 15,244 documents were identified during the study period (2002 to 2018). The number of documents fluctuated over time. In the first three years, from 2002 to 2005, about 100 documents (0.6% of the total documents) were identified. Occupational accident queries were very frequent in 2011 but then declined briefly. After 2012, the number of queries surged again, resulting in an average of over 1500 queries per year.

Figure 2 shows the co-occurrence network of the high frequency keywords in the set of documents. Using keyword analysis, we investigated the degrees of connection between major keywords. The green lines indicate the connections between keywords; the darker the shade of green, the more connected the words. Among the top 100 keywords derived from TF-IDF, the keywords related to workers' compensation showed the highest frequencies. These keywords were *company*, *process*, *salary*, *insurance*, *request*, *treatment*, and *occupational accidents*. Keywords related to occupational injuries, such as *hospitals*, *surgical operation*, *rates*, *finger*, and *hospitalization*, were also highly frequent. These high-frequency words linked to workers' compensation and occupational injuries were interconnected with many other words.

Figure 3 presents the results of the LDA model; 14 topics were formulated using LDA. This set was plotted in a 2D plan along the transverse (PC1) and longitudinal (PC2) axes. Each topic was displayed as a circle, and the overall prevalence was calculated as the areas of the circles. The centers of each topic were determined by computing the distance between topics. The 14 topics were manually classified into 3 themes.

Table 1. Numbers of documents related to occupational accidents by year (N=15,244).

Year	Documents, n (%)
2002	9 (0.06)
2003	19 (0.12)
2004	34 (0.22)
2005	32 (0.21)
2006	104 (0.68)
2007	307 (2.01)
2008	192 (1.26)
2009	1148 (7.53)
2010	699 (4.59)
2011	480 (3.15)
2012	1602 (10.51)
2013	1977 (12.97)
2014	1774 (11.64)
2015	1754 (11.51)
2016	1508 (9.89)
2017	1934 (12.69)
2018	1671 (10.96)

Figure 2. Co-occurrence network of 100 high-frequency keywords in documents related to occupational accidents.

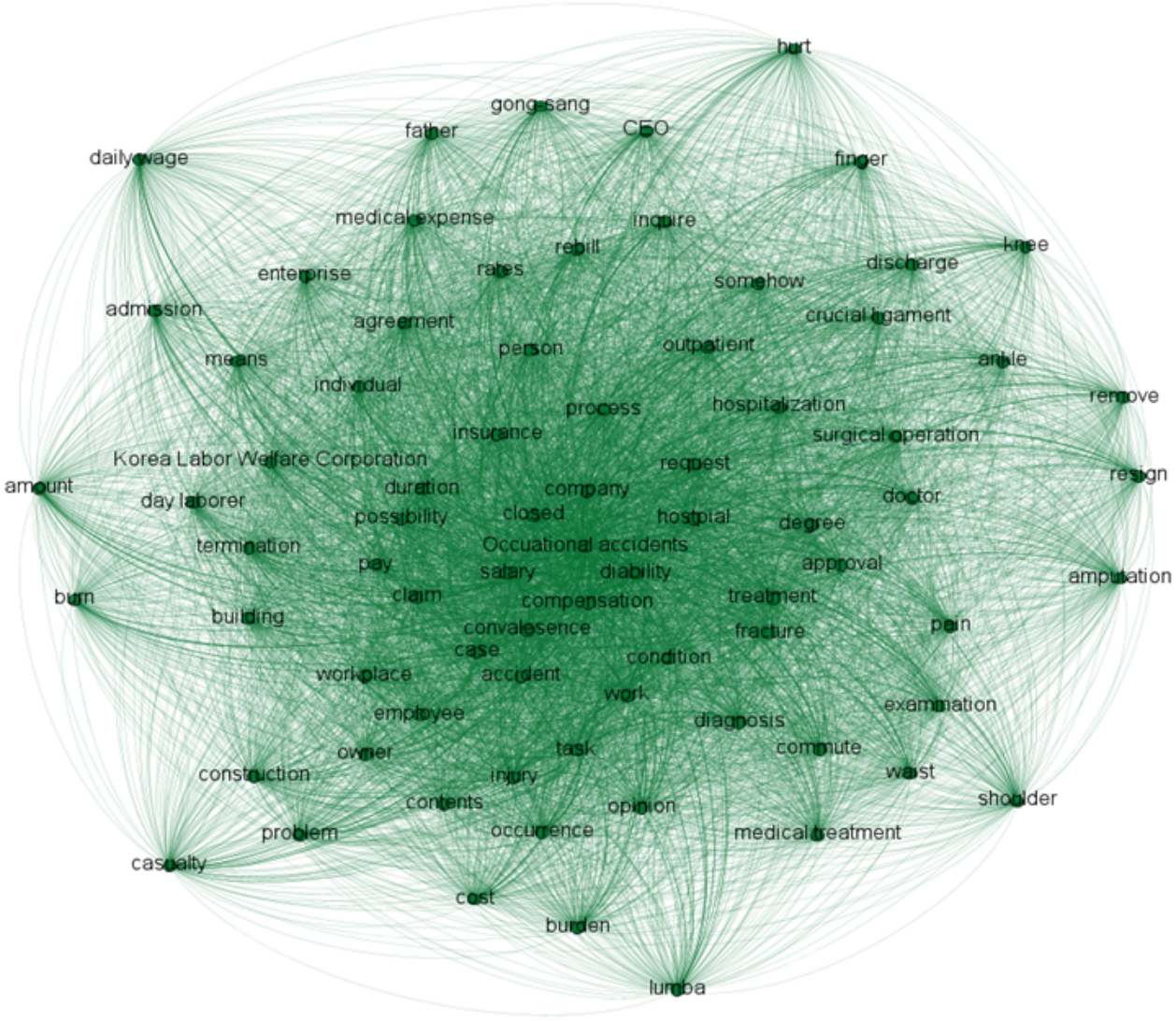


Figure 3. Overview of the formulation of 14 topics and 3 themes with latent Dirichlet allocation topic modeling.

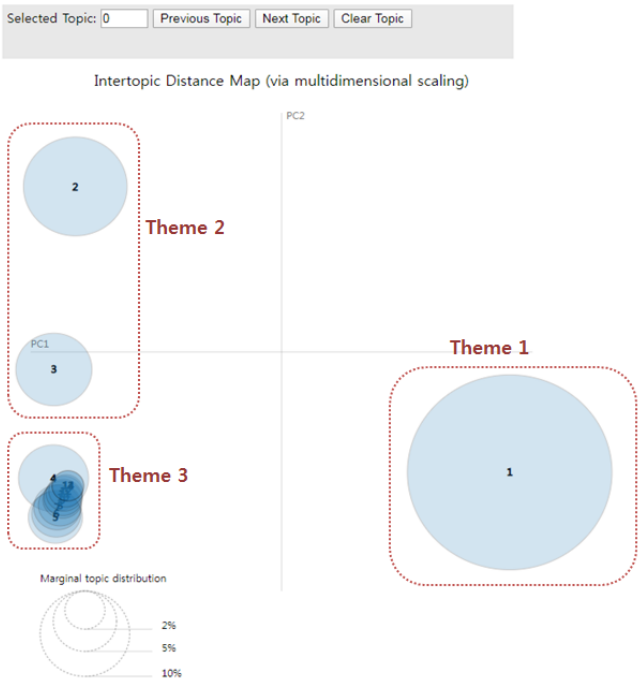


Table 2 shows the formulated topics and keywords from the LDA models, including the themes, topics, subjects, keywords, and topic proportions (ie, percentages). A total of 14 topics were divided into 3 themes. The topic proportion was defined as the proportion of words in a document that belong to a topic; this measurement indicates the main topics in a document. Herein, Topic 1 appeared the most frequently, at 52.2%, and was designated Theme 1 (workers' compensation benefits). Many of the keywords for Topic 1 were associated with workers' compensation (ie, company, process, occupational injury, insurance, and compensation), medical expense benefits (ie, hospital, surgical operation, medical treatment, hospitalization, and convalescence), temporary incapacity benefits (ie, salary, request, approval, workplace, and shutdown), and disability benefits (ie, possibility, rates, disability, fracture, and impairment). Topics 2 and 3, which totaled approximately 20.6% of all topics, were categorized as Theme 2, illicit agreement (called gong-sang in Korean) with the employer. These two topics contained keywords related to gong-sang: specifically, related to the musculoskeletal system of Topic 2 (ie, gong-sang, disk, rupture, MRI, lumbar, traffic accident, backbone, orthopedics, surgical procedure, and X-ray) and physical trauma of Topic 3 (ie, gong-sang, thumb, metal pin, reattachment, suture, tendon, bruise, scar, stiches, and infection). The remaining Topics 4-14, approximately 27.2% of the total, included keywords describing fatal and nonfatal injuries and vulnerable workers, Theme 3. Specifically, Topics 4, 5, 6, 9, and 10 implied keywords related to fatal and nonfatal injuries, such as critical care (Topic 4), fatal accident (Topic 5), lower extremity injury (Topic 6), fracture (Topic 9), and labor-management conflict (Topic 10). The remaining topics included keywords describing vulnerable workers, such as restaurant workers (Topic 7), construction workers (Topic 8), vulnerable jobs (Topic 11), student workers (Topic 12), and foreign workers (Topic 13). In addition, Topic 14 included words such as hearing, hearing loss, hepatocirrhosis, elderly, soft tissue, sudden, and garbage man.

Table 3 displays the regression coefficients and Figure 4 displays the heat map for the yearly changes in interest in the 14 topics from the LDA models. The largest topic was Topic 1, relating to workers' compensation benefits; interest in this topic significantly decreased over time. In contrast, interest in Topic 2 (musculoskeletal system), Topic 4 (critical care), and Topic 13 (foreign workers) continued over time without significant changes. Most of the topics (Topic 3, Topics 5 to 12, and Topic 14) showed significant increases over time (P<.05).

Table 2. Themes, topics, and keywords formulated by latent Dirichlet allocation topic modeling.

Theme and topics	Keywords	%
Theme 1: Workers' compensation benefits		52.2
Topic 1a: Workers' compensation	*company*, *process*, *occupational injury*, *insurance*, *compensation*	
Topic 1b: Medical expense benefits	*hospital*, *surgical operation*, *medical treatment*, *hospitalization*, *convalescence*	
Topic 1c: Temporary incapacity benefits	*salary*, *request*, *approval*, *workplace*, *shutdown*	
Topic 1d: Disability benefits	*possibility*, *rates*, *disability*, *fracture*, *impairment*	
Theme 2: Illicit agreement (*gong-sang*) with employer		
Topic 2: Musculoskeletal system	*gong-sang*, *disk*, *rupture*, *MRI*[a], *lumbar*, *traffic accident*, *backbone*, *orthopedics*, *surgical procedure*, *X-ray*	13.4
Topic 3: Physical trauma	*gong-sang*, *thumb*, *metal pin*, *reattachment*, *suture*, *tendon*, *bruise*, *scar*, *stiches*, *infection*	7.2
Theme 3: Fatal and nonfatal injuries and vulnerable workers		
Topic 4: Critical care	*bleeding*, *intensive care unit*, *nothing*, *whole life*, *hospital room*, *tear*, *infection*	6.1
Topic 5: Fatal accident	*die*, *bereaved*, *police*, *statement*, *rolled*, *autopsy*, *law*	3.7
Topic 6: Lower extremity injury	*cast*, *myoelectric*, *numbness*, *stabbed*, *foot*, *planta pedis*, *bumped*	3.0
Topic 7: Restaurant workers	*service*, *kitchen*, *owner*, *sick*, *piece of meat*, *hotel*, *McDonald*	2.8
Topic 8: Construction workers	*fall*, *construction site*, *carelessness*, *Achilles tendon*, *builders*, *license*, *cement*	2.3
Topic 9: Fracture	*open fracture*, *tumble*, *rib*, *incisura*, *cervical vertebrae*, *muscular pain*, *ventral root*	1.9
Topic 10: Labor-management conflict	*penalty*, *headquarters*, *Labor office*, *disgusted*, *civil and criminal*, *credit*, *tape-recording*	1.8
Topic 11: Vulnerable jobs	*construction workers*, *facilities*, *line*, *factory*, *guard*, *service industry*, *fast food*	1.7
Topic 12: Student workers	*disadvantage*, *part-timers*, *manager*, *professor*, *college student*, *class*, *attendance*	1.4
Topic 13: Foreign workers	*identity*, *visa*, *sojourn*, *local*, *Vietnam*, *Korea*, *reentry*	1.3
Topic 14: Others	*hearing*, *hearing loss*, *hepatocirrhosis*, *elderly*, *soft tissue*, *sudden*, *garbage man*	1.2

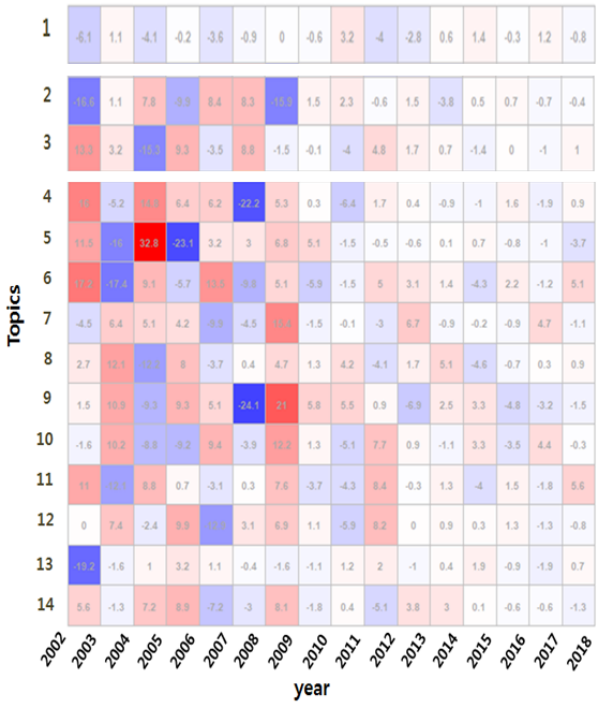
[a]MRI: magnetic resonance imaging.

Table 3. Regression coefficients of the yearly changes in interest in the 14 topics.

Theme and topic	Estimate	*P* value	Cold/Hot
Theme 1			
Topic 1: Workers' compensation benefits	–0.2885	<.001	Cold
Theme 2			
Topic 2: Musculoskeletal system	–0.0665	.10	—[a]
Topic 3: Physical trauma	0.1194	.01	Hot
Theme 3			
Topic 4: Critical care	–0.0042	.90	—
Topic 5: Fatal accident	0.0569	.01	Hot
Topic 6: Lower extremity injury	0.0207	.08	Hot
Topic 7: Restaurant workers	0.0465	<.001	Hot
Topic 8: Construction workers	0.0324	<.001	Hot
Topic 9: Fracture	0.0178	.004	Hot
Topic 10: Labor-management conflict	0.0261	<.001	Hot
Topic 11: Vulnerable jobs	0.0139	.002	Hot
Topic 12: Student workers	0.0208	<.001	Hot
Topic 13: Foreign workers	–0.0040	.64	—
Topic 14: Others	0.0086	.002	Hot

[a]—: no significant variation.

Figure 4. Heat map of the yearly changes in interest in the 14 topics.



Discussion

Principal Findings

Occupational accidents are a major public health challenge [1]. Although surveillance systems have been constructed to monitor the mortality and morbidity of occupational accidents, the surveillance data may be inaccurate due to underreporting or covering up of injuries and illnesses [6,19]. This impedes proper evaluation of the magnitude of health and safety problems in the workplace, which reduces protection of workers from workplace hazards, identification of risks, and implementation of needed interventions. Our study investigated hidden issues in occupational accidents observed on the internet. We collected

data from SNSs on occupational accident–related queries and analyzed them using the LDA topic model. The LDA analysis extracted a total of 14 topics, which we clustered into 3 themes: workers' compensation benefits (Theme 1), illicit agreements with the employer (Theme 2), and fatal and nonfatal injuries and vulnerable workers (Theme 3).

The largest share of 52.2% of the collected documents concerned workers' compensation benefits; this was classified as Theme 1. In general, if a worker is injured while working, they do not or cannot work for a certain period of time. This not only leads to the burden of medical expenses but also considerably impacts the livelihood of the worker and their family due to the suspension of household income. To ensure social security as a collective measure against injuries, disease, and death, the Korean government implemented the Industrial Accident Compensation Insurance Act 2015, in which all workers are required to participate. The Industrial Accident Compensation Insurance Act pays compensation (hereafter referred to as workers' compensation benefits) such as medical care benefits, temporary layoff benefits, and disability benefits to workers who are injured or disabled at work [20,21]. Although Korean workers who require more than three days of treatment are covered by the Industrial Accident Compensation Insurance Act, it has been found that the approval process for workers' compensation benefits can be protracted and difficult because insurers and employers sometimes attempt to limit injured workers' benefits [20]. Therefore, the primary concern of workers suffering from occupational injuries or diseases may focus on obtaining compensation benefits through legal procedures rather than how the occupational injury or disease affects them. That is, affected workers need advice from legal experts (such as certified labor attorneys and other attorneys) regarding the type of compensation they are eligible for. Due to these affected workers' needs, Theme 1 is the most frequently mentioned subject throughout the entire industrial accident query. As shown in Table 2, Theme 1 is composed of Topic 1, which includes keywords related to workers' compensation such as *company*, *process*, *request*, *benefits*, *occupational injury*, *insurance*, *accident*, *approval*, and *compensation*. In addition, it includes keywords describing the specific benefits of industrial accident compensation, namely medical expense benefits (ie, *hospital*, *surgical operation*, *medical treatment*, *hospitalization*, and *convalescence*); temporary incapacity benefits (ie, *salary*, *request*, *approval*, *workplace*, and *shutdown*); and disability benefits (ie, *possibility*, *rates*, *disability*, *fracture*, and *impairment*).

The second-largest cluster was Theme 3, fatal and nonfatal injuries and vulnerable workers, with 28.2 percent of documents. When we considered occupational accidents and industrial accident compensation, we found that the major concern of injured or disabled workers during work was whether they could be beneficiaries of industrial accident compensation. Although numerous injuries and illnesses are known to result from occupational causes [1,22], establishing a causal relationship between work factors and injuries is a prerequisite for the approval of industrial accident compensation. Therefore, the affected workers appear to have described their occupationally related injuries on SNSs, and on the same platforms, they asked whether their accidents or illnesses would constitute legitimate industrial accidents. In Theme 3, topics 4, 5, 6, 9, 10, and 14 encompassed keywords describing fatal and nonfatal occupation-related injuries (ie, *bleeding*, *die*, *cast*, *fracture*, *incisura*, and *hearing loss*). Another issue in Theme 3 is the employment status of affected workers. Workers' employment status affects health and causes safety disparities, which is exacerbated by unregulated and unsafe workplaces [23-25]. International reviews have found that occupational injury rates for nonstandard and temporary workers are greater than those for permanent workers [26,27]. Nonregular workers such as part-time workers, temporary contract workers, and dispatched workers may be excluded from social insurance and corporate welfare programs, and they may receive no compensation when they are affected by industrial accidents. In fact, many workers on fixed terms are not covered by workers' compensation [28]. Keywords related to vulnerable workers (ie, *service*, *hotel*, *construction workers*, *guard*, *fast food*, and *Vietnam*) were found in topics 3, 7, 8, 11, 12, and 13 in Theme 3.

Finally, Theme 2 accounted for the smallest percentage (20.6%) of the collected documents. The subject "illicit agreement" (*gong-sang* in Korean) in Theme 2 refers to cases where injured employees negotiate settlements separately (ie, beyond the legal purview) with their employers instead of legally declaring their accidents and pursuing workers' compensation benefits [29]. Although injuries to the musculoskeletal system (Topic 2) and physical trauma (Topic 3) were the most common types of occupational injury, they are considered to be mild injuries, and it can be difficult to prove that they are work-related. Therefore, workers' compensation claims for these injuries are rarely approved. Thus, when workers suffer from musculoskeletal disease or physical trauma (ie, *disk*, *rupture*, *bruise*, and *reattachment*), they often enter into illicit agreements with their employers to receive payment from the companies.

We also performed a time-series analysis to investigate the annual changes in themes and topics between 2002 and 2018. Typical queries regarding the compensation for occupational accidents (seen in Theme 1) were found to decrease over time, whereas most queries related to themes 2 and 3 increased. Notably, certain queries regarding the musculoskeletal system (Topic 2 in Theme 2), critical care (Topic 4 in Theme 3), and foreign workers (Topic 13 in Theme 3) showed no significant variation in the number of queries during the study period (2002 to 2018). This appears to indicate that those topics surface consistently every year in discussions surrounding occupational accidents.

LDA topic modeling is a method that is used to identify the underlying topics contained in unstructured text data. This method is used widely in topic detection in medicine, marketing research, political science, and linguistics [16,30,31]. Our study is the first to use LDA techniques to discover latent issues related to occupational accidents. Based on web-based and unstructured web-based documents, we obtained novel insights into the unfulfilled needs of industrial accident workers and the information they sought via expert consultation as well as the occupations that are vulnerable to industrial accidents.

Limitations

We should mention several important limitations regarding the methods used in this study. The main critique here relates to the instability of topic distribution and interpretation. Because topic modeling is sensitive to input data and analysis, changes (such as adding new documents and implementing text mining algorithms such as tokenization and stemming) can generate completely different topics. Therefore, the topics are often an amalgam, and it is difficult to assign truth to the interpretation and validation in the given corpus [32]. In addition, we focused on SNS data from experts (ie, certified labor attorneys, other attorneys, and physicians) on counseling websites between 2002 and 2018 with the aim of exploring latent issues in occupational accidents. However, depending on the SNS data source and time period collected, different topics and themes can be derived from our results. Therefore, our results may not genuinely represent the overall view of occupational accidents in our country. Moreover, applying our results to other countries with distinct industrial accident laws and regulations would require considerable attention.

Conclusion

We conducted LDA analysis with SNS data related to occupational accidents and discovered that the primary concerns of workers posting about occupational injuries and diseases were workers' compensation benefits, fatal and nonfatal injuries and vulnerable workers, and illicit agreements with employers. While traditional systems focus mainly on quantitative monitoring of occupational accidents, qualitative aspects formulated by topic modeling from unstructured SNS queries may be valuable for providing practical knowledge and information to the affected workers. This approach may be useful to address inequality among workers and improve occupational health and safety.

Acknowledgments

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (No. NRF-2017R1E1A1A01078235).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary table and figure.
[DOCX File , 72 KB - jmir_v22i8e19222_app1.docx]

References

1. Alli BO. Fundamental Principles of Occupational Health and Safety, Second Edition. Geneva: International Labour Office; 2008.
2. Global Estimates of Occupational Accidents and Work-Related Illnesses 2017. Singapore: Workplace Safety and Health Institute; Sep 2017.
3. Mendeloff J, Burns R. States with low non-fatal injury rates have high fatality rates and vice-versa. Am J Ind Med 2013 May;56(5):509-519. [doi: 10.1002/ajim.22047] [Medline: 22473632]
4. Rhee KY, Choe SW, Kim YS, Koo KH. The trend of occupational injuries in Korea from 2001 to 2010. Saf Health Work 2013 Mar;4(1):63-70 [FREE Full text] [doi: 10.5491/SHAW.2013.4.1.63] [Medline: 23515324]
5. Statistics on industrial accident in 2017. Korean Statistical Information Service. South Korea; 2017. URL: http://www.index.go.kr [accessed 2020-06-10]
6. Moore JT, Cigularov KP, Sampson JM, Rosecrance JC, Chen PY. Construction workers' reasons for not reporting work-related injuries: an exploratory study. Int J Occup Saf Ergon 2013;19(1):97-105. [doi: 10.1080/10803548.2013.11076969] [Medline: 23498712]
7. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and Privacy Implications of Using the Internet and Social Media to Recruit Participants for Health Research: A Privacy-by-Design Framework for Online Recruitment. J Med Internet Res 2017 Apr 06;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]
8. Chan WS, Leung AY. Use of Social Network Sites for Communication Among Health Professionals: Systematic Review. J Med Internet Res 2018 Mar 28;20(3):e117. [doi: 10.2196/jmir.8382] [Medline: 29592845]
9. Dagan N, Beskin D, Brezis M, Reis BY. Effects of Social Network Exposure on Nutritional Learning: Development of an Online Educational Platform. JMIR Serious Games 2015;3(2):e7 [FREE Full text] [doi: 10.2196/games.4002] [Medline: 26441466]
10. Park E, Cho S. KoNLPy: Korean natural language processing in Python. 2014 Oct 10 Presented at: Paper presented at the 26th Annual Conference on Human & Cognitive Language Technology; Oct 2014; Chuncheon, Korea.
11. MeCab: Yet Another Part-of-Speech and Morphological Analyzer. Webpage in Japanese. GitHub. URL: http://taku910.github.io/mecab/ [accessed 2020-07-26]

12. Jeffrey D, Ullman J, Jur L. Mining of Massive Datasets. Stanford University Infolab. URL: http://infolab.stanford.edu/~ullman/mmds/ch9.pdf [accessed 2020-06-10]
13. Blei D, Ng A, Jordan M. Latent Dirichllocation. J Mach Learn Res 2003;3:993-1022.
14. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr;55(4):77-84. [doi: 10.1145/2133806.2133826]
15. Chang J, Boyd-Graber J, Wang C, Gerrish S, Blei DM. Reading tea leaves: How humans interpret topic models. Neural Information Processing Systems. 2009. URL: https://papers.nips.cc/paper/3700-reading-tea-leaves-how-humans-interpret-topic-models.pdf [accessed 2020-06-10]
16. Huang M, ElTayeby O, Zolnoori M, Yao L. Public Opinions Toward Diseases: Infodemiological Study on News Media Data. J Med Internet Res 2018 May 08;20(5):e10047 [FREE Full text] [doi: 10.2196/10047] [Medline: 29739741]
17. Jacobi C, van Atteveldt W, Welbers K. Quantitative analysis of large amounts of journalistic texts using topic modelling. Digital Journalism 2015 Oct 13;4(1):89-106. [doi: 10.1080/21670811.2015.1093271]
18. Newman D, Jey HL, Karl G, Timothy B. Automatic Evaluation of Topic Coherence. 2010 Jun 2 Presented at: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics; June 2-4, 2010; Los Angeles, CA URL: https://www.aclweb.org/anthology/N10-1012.pdf
19. Lim S, Yoon J, Rhie J, Bae SW, Kim J, Won J. The Relationship between Free Press and Under-Reporting of Non-Fatal Occupational Injuries with Data from Representative National Indicators, 2015: Focusing on the Lethality Rate of Occupational Injuries among 39 Countries. Int J Environ Res Public Health 2018 Dec 14;15(12):2856 [FREE Full text] [doi: 10.3390/ijerph15122856] [Medline: 30558139]
20. Kang S, Kwon O. Occupational Injury Statistics in Korea. Saf Health Work 2011 Mar;2(1):52-56. [doi: 10.5491/shaw.2011.2.1.52]
21. Kim JG, Duk YJ. Development and Operation of Workers' Compensation Insurance Scheme in S. Korea. Ministry of Employment and Labor, Republic of Korea. 2012. URL: https://www.moel.go.kr/english/download_eng.jsp?type=&file=12664_2.pdf [accessed 2020-06-10]
22. LaDou J. International occupational health. Int J Hyg Environ Health 2003 Aug;206(4-5):303-313. [doi: 10.1078/1438-4639-00226] [Medline: 12971685]
23. Min J, Park S, Hwang SH, Min K. Disparities in precarious workers' health care access in South Korea. Am J Ind Med 2016 Dec;59(12):1136-1144. [doi: 10.1002/ajim.22658] [Medline: 27699816]
24. Min KB, Park SG, Song JS, Yi KH, Jang TW, Min JY. Subcontractors and increased risk for work-related diseases and absenteeism. Am J Ind Med 2013 Nov;56(11):1296-1306. [doi: 10.1002/ajim.22219] [Medline: 23794385]
25. Siqueira CE, Gaydos M, Monforton C, Slatin C, Borkowski L, Dooley P, et al. Effects of social, economic, and labor policies on occupational health disparities. Am J Ind Med 2014 May;57(5):557-572 [FREE Full text] [doi: 10.1002/ajim.22186] [Medline: 23606055]
26. Nicholson VJ, Bunn TL, Costich JF. Disparities in work-related injuries associated with worker compensation coverage status. Am J Ind Med 2008 Jun;51(6):393-398. [doi: 10.1002/ajim.20565] [Medline: 18381597]
27. Quinlan M. The effects of non-standard forms of employment on worker health and safety. International Labour Organization. Geneva: International Labour Office; 2015. URL: https://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---travail/documents/publication/wcms_443266.pdf [accessed 2020-08-09]
28. Employment arrangements: Improved outreach could help ensure proper worker classification. United States Government Accountability Office. 2006. URL: https://www.gao.gov/new.items/d06656.pdf [accessed 2020-08-09]
29. Min J, Song S, Kim H, Min K. Mining Hidden Knowledge About Illegal Compensation for Occupational Injury: Topic Model Approach. JMIR Med Inform 2019 Sep 26;7(3):e14763 [FREE Full text] [doi: 10.2196/14763] [Medline: 31573948]
30. Hao H, Zhang K. The Voice of Chinese Health Consumers: A Text Mining Approach to Web-Based Physician Reviews. J Med Internet Res 2016 May 10;18(5):e108 [FREE Full text] [doi: 10.2196/jmir.4430] [Medline: 27165558]
31. Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimed Tools Appl 2018 Nov 28;78(11):15169-15211. [doi: 10.1007/s11042-018-6894-4]
32. Zhao W, Zou W, Chen JJ. Topic modeling for cluster analysis of large biological and medical datasets. BMC Bioinformatics 2014;15 Suppl 11:S11 [FREE Full text] [doi: 10.1186/1471-2105-15-S11-S11] [Medline: 25350106]

Abbreviations

LDA: Latent Dirichlet allocation
NLP: natural language processing
SNS: social network service
TF-IDF: term frequency–inverse document frequency

Edited by G Eysenbach; submitted 08.04.20; peer-reviewed by R Poluru, M Lamba; comments to author 30.04.20; revised version received 15.06.20; accepted 07.07.20; published 13.08.20.

Please cite as:
Min KB, Song SH, Min JY
Topic Modeling of Social Networking Service Data on Occupational Accidents in Korea: Latent Dirichlet Allocation Analysis
J Med Internet Res 2020;22(8):e19222
URL: http://www.jmir.org/2020/8/e19222/
doi:10.2196/19222
PMID:32663156

Original Paper

Social Media Text Mining Framework for Drug Abuse: Development and Validation Study With an Opioid Crisis Case Analysis

Tareq Nasralah[1*], PhD; Omar El-Gayar[2*], PhD; Yong Wang[3*], PhD

[1]Supply Chain and Information Management Group, D'Amore-McKim School of Business, Northeastern University, Boston, MA, United States

[2]College of Business and Information Systems, Dakota State University, Madiosn, SD, United States

[3]The Beacom College of Computer and Cyber Sciences, Dakota State University, Madiosn, SD, United States

[*]all authors contributed equally

Corresponding Author:
Tareq Nasralah, PhD
Supply Chain and Information Management Group
D'Amore-McKim School of Business
Northeastern University
360 Huntington Ave
Boston, MA, 02115
United States
Phone: 1 617 373 4460
Email: t.nasralah@northeastern.edu

Abstract

Background: Social media are considered promising and viable sources of data for gaining insights into various disease conditions and patients' attitudes, behaviors, and medications. They can be used to recognize communication and behavioral themes of problematic use of prescription drugs. However, mining and analyzing social media data have challenges and limitations related to topic deduction and data quality. As a result, we need a structured approach to analyze social media content related to drug abuse in a manner that can mitigate the challenges and limitations surrounding the use of such data.

Objective: This study aimed to develop and evaluate a framework for mining and analyzing social media content related to drug abuse. The framework is designed to mitigate challenges and limitations related to topic deduction and data quality in social media data analytics for drug abuse.

Methods: The proposed framework started with defining different terms related to the keywords, categories, and characteristics of the topic of interest. We then used the Crimson Hexagon platform to collect data based on a search query informed by a drug abuse ontology developed using the identified terms. We subsequently preprocessed the data and examined the quality using an evaluation matrix. Finally, a suitable data analysis approach could be used to analyze the collected data.

Results: The framework was evaluated using the opioid epidemic as a drug abuse case analysis. We demonstrated the applicability of the proposed framework to identify public concerns toward the opioid epidemic and the most discussed topics on social media related to opioids. The results from the case analysis showed that the framework could improve the discovery and identification of topics in social media domains characterized by a plethora of highly diverse terms and lack of a commonly available dictionary or language by the community, such as in the case of opioid and drug abuse.

Conclusions: The proposed framework addressed the challenges related to topic detection and data quality. We demonstrated the applicability of the proposed framework to identify the common concerns toward the opioid epidemic and the most discussed topics on social media related to opioids.

(J Med Internet Res 2020;22(8):e18350) doi:10.2196/18350

KEYWORDS

drug abuse; social media; infodemiology; infoveillance; text mining; opioid crisis

Introduction

Background

Social media are used by patients to exchange information and discuss different health-related topics [1]. Popular social media platforms, such as Twitter, provide efficient methods of information access for health surveillance and social intelligence [2] and could be used to recognize communication and behavioral themes of problematic use of prescription drugs [3]. Social media have been used in several studies as a resource for monitoring prescription medication abuse [3-7]. The literature shows that clear signals of medication abuse can be drawn from social media posts [5]. Furthermore, the literature used text mining to examine and compare discussion topics to discover the thematic similarity, difference, and membership in online mental health communities [8], provide timely information for epidemiologic surveillance [9], and analyze the public's reactions to the opioid crisis [10].

Prior Work

Social media users' posts are used to better understand providers' attitudes toward using recovery drugs, such as "naloxone," to treat opioid addiction [11]. Indeed, social media, such as Twitter, can serve as data sources for approaches that automatically detect opioid addicts and support a better practice of opioid addiction, prevention, and treatment [12]. Several studies have used social media as sources of input data to identify individuals amenable to drug recovery interventions [13] and used text mining to examine and compare discussion topics on social media communities to discover the thematic similarity, difference, and membership in online mental health communities [8].

Kalyanam et al [4] developed a strategy in the field of digital epidemiology to better identify, analyze, and understand trends in the nonmedical use of prescribed medications and drugs through social media by utilizing unsupervised machine learning methods. The results showed that social media data mining could provide insights regarding knowledge from daily life that could support a better practice of opioid addiction prevention and treatment. Cherian et al [6] characterized representations of codeine misuse through analysis of public posts on Instagram using content analysis to identify common themes arising in images. The results showed that codeine misuse was commonly represented with the ingestion of alcohol, cannabis, and benzodiazepines.

Further, Lu et al [7] analyzed Reddit data to gain insight into drug use/misuse by classifying user posts using a binary classifier that predicts transitions from casual drug discussion forums to drug recovery forums. Analysis and results showed that the proposed approach "delineates drugs that are associated with higher rates of transitions from recreational drug discussion to support/recovery discussion, offers insights into modern drug culture, and provides tools with potential applications in combating the opioid crisis" [7]. Jelodar et al [14] examined online discussions to discover knowledge and evaluate patients' behaviors based on their opinions and discussions about alcohol, using a semantic framework based on a topic model (latent Dirichlet allocation [LDA]) and random forest. The results showed that social media data could be helpful in detecting relevant safety problems in a patient's daily life.

To demonstrate that the use of machine learning and linguistic rules separately is not enough to achieve better results for information extraction from social media, Jenhania et al [15] proposed a hybrid system combining dictionaries, linguistic patterns, and machine learning to extract structured and salient drug abuse information from health-related tweets. The results showed that the use of a linguistic method based on a dictionary with no dictionary updates is a failed solution. Combining linguistic rules, machine learning, and domain achieved good performance compared with other approaches.

Goal of This Study

Despite recent advances, there are several limitations that exist when social media data are used for studying drug abuse. First, limitations exist in terms of data relevance and the ability to capture relevant data [6,16]. Second, challenges exist with social media data in terms of completeness and inconsistencies, especially with data collected from multiple resources [17]. Third, user-generated content often includes users' personal opinions and thoughts, making the task of extracting high-quality information from such data increasingly important [18]. Finally, obtaining high-quality data is a key to avoid any issues in the data preparation step. Several studies reported issues with informal language used on social media [17-19], which could lead to low data quality.

Based on the aforementioned challenges and issues, there is a need to develop a framework that identifies important and relevant quality data on social media to study drug abuse. Furthermore, research that systematically analyzes social media content to study drug abuse in a manner that mitigates challenges and limitations related to topic deduction and data quality is needed.

This research proposed a social media text mining framework for drug abuse that provides a systematic approach to analyze social media data and addresses challenges related to topic deduction and data quality. We demonstrated the applicability of the proposed framework using the opioid epidemic as a drug abuse case analysis by analyzing Twitter data. Twitter was selected because it is an instant day-to-day micro-blogging platform [20] and is widely considered to have an advantage during crises [21]. From a theoretical perspective, this research highlights the importance of developing and adapting text mining techniques for social media data analytics in the context of drug abuse. From a practical perspective, automatically analyzing social media user–generated content can help understand the public themes and topics regarding drug abuse that exist in social media networks.

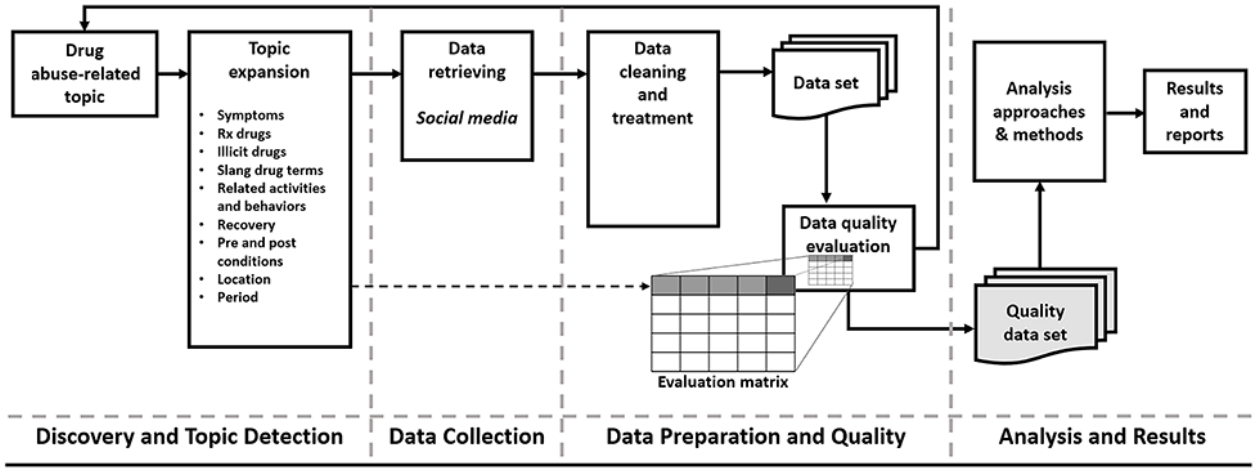
Methods

Social Media Text Mining Framework for Drug Abuse

Figure 1 shows the systematic framework to study drug abuse–related topics using social media data. The framework addresses topic detection and data quality challenges. The framework consists of four phases, namely, discovery and topic

detection, data collection, data preparation and quality evaluation, and finally, analysis and results.

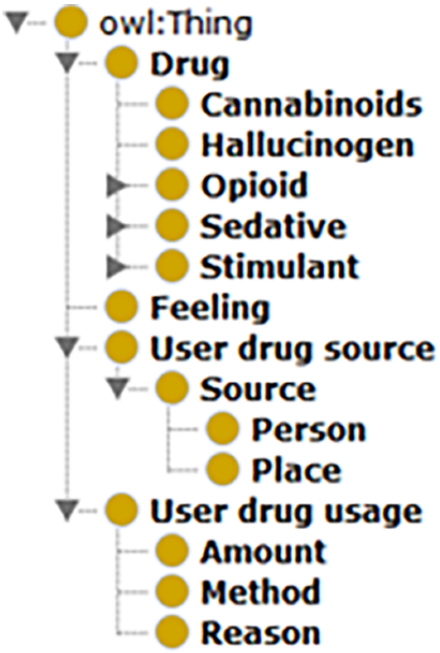
Figure 1. Social media text mining framework for drug abuse. Rx: prescription.



Phase I: Discovery and Topic Detection

According to the literature, challenges and limitations in discovery and topic detection exist [16,17,19]. The interdisciplinary nature of social media data and difficulties in determining the topic that social media posts represent are the most common challenges in social media analysis [17]. To address these challenges, our framework includes a topic expansion step. This step identifies drug abuse–related topics that address the research domain and objectives. To formalize the identification process, we created an ontology for drug abuse based on the literature and expanded the drug abuse ontology proposed by Cameron et al [22] by including related concepts and instances in prescription drug classes (Figure 2) that relate to symptoms, prescription drugs, illicit drugs, slang terms, related activities and behaviors, recovery, conditions, location, and period. The new ontology builds on the ontology by Cameron et al [22] by grouping concepts and instances into themes, reorganizing the concepts hierarchy, and including additional slang keywords and terminologies based on collected data that are related to opioid abuse.

Figure 2. Drug abuse ontology for drug main terms and classes.



Phase II: Data Collection

Data collection is determined by the date range of interest, social media data sources, such as Twitter, relevant keywords to search for posts, and restrictions to impose (language: English, geographic location: United States, etc). The selection of relevant search keywords is based on the proposed ontology in the topic expansion step.

Phase III: Data Preparation and Quality

The veracity of data leads to issues in data preparation [17]. Therefore, we need to preprocess the collected data and clean it from stop words, punctuations, URLs, etc. To extract quality data for the analysis phase, we evaluated the quality of the data with respect to the terms from the topic expansion step, using an evaluation matrix (Figure 3). The evaluation matrix examines each user's post in the data set to ensure it includes related terms from the ontology in the topic expansion step.

We automatically generated and populated the evaluation matrix using natural language processing (Python NLTK package) to examine the relevant user tweet. Thereafter, each user's post (represented as a row) was evaluated against different terms (represented as columns) from the topic expansion step. If the term is present in the post, the value of the term will be one; otherwise, it will be zero.

Evaluation is performed by assigning a data quality score for each post. The value of the quality score was calculated based on the summation of all the term values. The quality score of each user's post was used as a metric for filtering out low-quality irrelevant posts. Specifically, posts with quality scores from 2 to 10 were retained as relevant posts and those with scores less than 2 or greater than 10 were considered not relevant.

Thresholds were selected based on manual analysis of the collected data. The choice of 2 as the minimum quality score for a post to be relevant was based on the presence of two keywords from the ontology, which increases the possibility of making the post relevant to the topic. The presence of another feature in the post increases the chance of making the context of the post relevant to the study topic. On the other hand, the choice of 10 as the maximum quality score for a post to be relevant was based on manual analysis, where the presence of many words (>10) in a post makes the subject matter and the context of the post too scattered and inaccurate to be relevant to the topic. The evaluation matrix performance was validated against a ground truth. The ground truth represents manually labeled posts.

Figure 3. Evaluation matrix for users' postquality assessment.

Post ID	Word 1	Word 2	Word 3	Word 4	.....	Word n-1	Word n	Score	Auto_Lable
Post 1					.....				
Post 2					.....				
Post 3					.....				
Post 4					.....				
⋮									
Post m-1					.....				
Post m					.....				

Phase IV: Analysis Approaches and Methods

Once quality data are ascertained, the researcher can choose the suitable data analysis approach based on research questions and objectives. Such approaches could be unsupervised machine learning approaches like topic modeling [23] and supervised machine learning approaches like classification [24].

Evaluation of the Framework

We instantiated the proposed framework using opioid drug abuse as a case study. To demonstrate and evaluate the proposed topic expansion step in the framework, we instantiated an opioid drug abuse ontology from the proposed drug abuse ontology in Figure 2. To demonstrate the applicability of the opioid drug ontology, we used a sample data set of 10,000 tweets belonging to self-identified opioid users on Twitter. We studied the distribution of the ontology terms and their occurrence over the collected samples. We relied on data collected from Twitter using the Crimson Hexagon platform from June 29, 2018, to April 11, 2019.

To evaluate the performance of the evaluation matrix as the first step in the process, we randomly selected 1000 tweets from the results of the search query. Two independent researchers reviewed the 1000 tweets to determine if each tweet was relevant. We measured the level of agreement using the Cohen kappa interrater reliability metric [25]. Considering the search query as our "base" classifier (classifier 1), where all 1000 tweets were predicted as relevant, we evaluated its performance against the "ground truth" obtained from manually evaluating the relevance of the 1000 randomly selected tweets, using standard data mining and machine learning performance metrics [26]. Using the same 1000 tweets, we applied the evaluation matrix (referred to as classifier 2) to classify the tweets as relevant or nonrelevant based on their quality score. Using the ground truth data obtained earlier, we evaluated the performance of classifier 2 and compared its performance metrics against those obtained for classifier 1.

Since our interest was to identify the different topics that exist in Twitter data about the opioid epidemic, we applied unsupervised text modeling using LDA [23] to extract the different topics that Twitter users have in their tweets. In our case analysis of studying the opioid epidemic, we followed the best practices suggested by Arun et al [27] and computed the term frequency-inverse document frequency (TF-IDF) and perplexity of a held-out test set to evaluate LDA models using a different number of topics. We trained several LDA models with a different number of topics (k) and evaluated the perplexity of a held-out test set. We held out 20% of the data for test purposes and trained the models on the remaining 80%.

Results

The results demonstrate the instantiation of the four phases of the proposed framework. Regarding phase I, Figure 4 depicts the results of an instantiation of the ontology in Figure 2 for opioid drug abuse, while Figure 5 shows the opioid drug abuse ontology classes using a tree representation. The same post can belong to more than one class in the tree. The number next to each branch in the tree represents the number of tweets based on the node terms. For example, searching relevant tweets using opioid drug–related terms, such as opioid, opioids, opiate, and opiates, yielded 6000 related tweets out of the 10,000 tweets. However, searching tweets using more specific and focused terms that belong to all subclasses of the "opioid drug" class yielded 9886 related tweets. A sample of the defined terms in the ontology can be found in Multimedia Appendix 1.

Figure 4. Opioid drug abuse ontology that includes opioid-related terms and concepts.

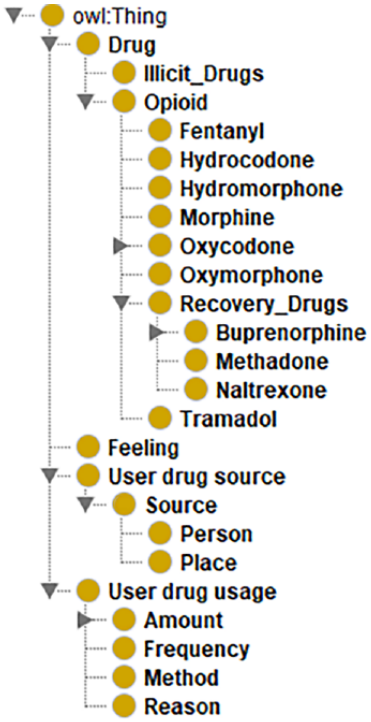
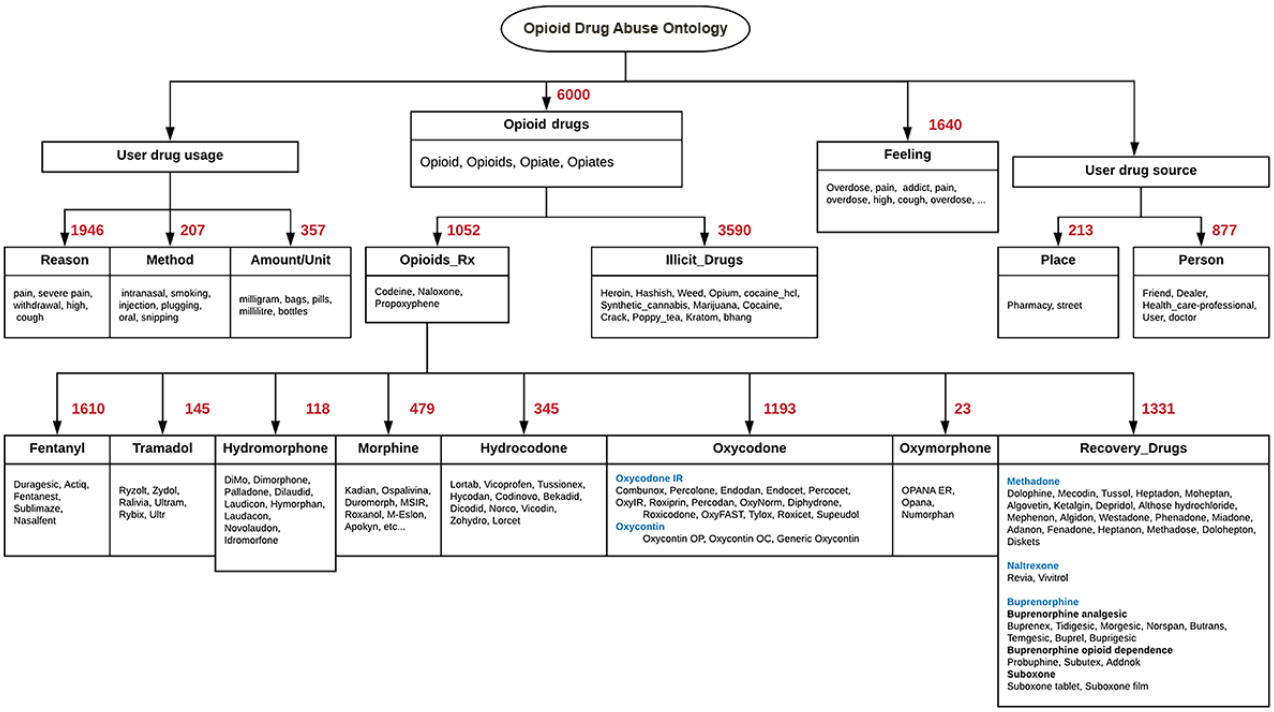


Figure 5. Opioid drug abuse ontology tree hierarchy that reflects the distribution of sample tweets over the ontology concepts and terms. Rx: prescription.



In phase II, using the Crimson Hexagon platform, a social media analytics tool for data collection and analysis, we created a search query (Figure 6) using terms extracted from the opioid drug abuse ontology to retrieve data with no retweets or URLs. For the time period from June 29, 2018, to April 11, 2019, we collected Twitter data related to the opioid epidemic from social media users living in the United States, including practitioners, leaders, patients, journalists, etc, who tweeted about opioids. Overall, we were able to collect 502,830 English-language tweets. Figure 7 shows a sample of the collected tweets.

Reviewing the 1000 randomly selected tweets as part of phase III resulted in 764 relevant tweets and 236 nonrelevant tweets representing the ground truth of the evaluation process. The resultant level of agreement using the Cohen kappa interrater reliability metric was 0.70, which represents moderate agreement [25].

The "base" classifier (classifier 1) (reflecting the search query without the application of the evaluation matrix) defaults to all 1000 tweets predicted as relevant, and this effectively indicates that 764 tweets were categorized as true positive and the remaining 236 tweets were categorized as false positive. The resultant performance metrics for our "base" classifier are summarized using the evaluation metrics under classifier 1 in Table 1.

Using the evaluation matrix (referred to as classifier 2) to classify the 764 relevant tweets representing the ground truth, classifier 2 resulted in 738 tweets classified as relevant (true positives) and 26 tweets classified as nonrelevant (false negatives). From the 236 nonrelevant tweets, classifier 2 classified 190 tweets as nonrelevant (true negative) and 46 tweets as relevant (false positive). The performance metrics using the evaluation matrix are summarized under classifier 2 in Table 1.

The results from Table 1 demonstrate that the proposed evaluation matrix outperforms the manual process. Such results are considered sufficient to adopt the evaluation matrix for evaluating the quality of the collected tweets.

We demonstrate the use of the evaluation matrix to automatically evaluate the relevance of the opioid-related tweets. Using the opioid drug abuse ontology, we ended up with more than 250 related terms. To obtain good data quality, we used a variety of opioid drug abuse terms in the evaluation matrix as features. The terms were adapted from the opioid ontology. Figure 8 shows a sample of the evaluation matrix results.

Based on the evaluation matrix score that represents the summation of the occurrence of ontology terms in a tweet, the matrix labeled relevant tweets with 1 if the tweet score was $\in$ [2, 10] and nonrelevant tweets with zero if the tweet score was $\notin$ [2, 10].

According to the evaluation matrix, 366,736 tweets out of the 502,830 collected tweets were deemed relevant (good quality) based on their scores. Multimedia Appendix 1 includes tables showing samples of good quality and excluded tweets. After we obtained the good quality tweet data set, we applied several preprocessing steps to prepare the data for the analysis phase, including the removal of emojis, lemmatization, and tokenization.

Figure 6. A search query based on the terms and concepts from the opioid drug abuse ontology to collect opioid-related users' posts.

```
(
Opioid OR Opioids OR Opiate OR Opiates OR Codeine OR Naloxone OR Propoxyphene
OR Hydrocodone OR Vicodin OR Oxycodone OR OxyContin OR Oxy OR Oxys OR Percocet
OR Oxymorphone OR Opana OR Morphine OR Hydromorphone OR Tramadol OR Fentanyl
OR Duragesic OR Actiq OR Subsys OR Recovery_Drugs OR Methadone OR Dolophine
OR Methadose OR Diskets OR Naltrexone OR Revia OR Vivitrol OR Buprenorphine
OR Probuphine OR Subutex  OR Suboxone
)

AND -
(http OR https OR RT)
```

Figure 7. Sample of the collected tweets.

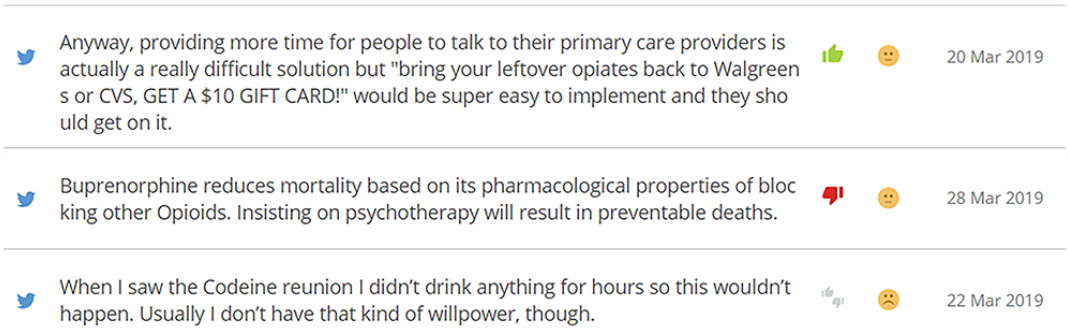


Table 1. Metrics comparing the performance of manual analysis (classifier 1) and the evaluation matrix (classifier 2) [26].

Variable	Without evaluation matrix (classifier 1)	With evaluation matrix (classifier 2)
Precision	0.764	0.941
Recall	1.000	0.966
F-measure	0.866	0.953
Accuracy	0.764	0.928

Figure 8. Sample outcomes from the evaluation matrix, where tweets autolabeled 0 are irrelevant and tweets autolabeled 1 are relevant.

Tweets	boi	die	dies	dr	...	tobacco	tram	tylenol	vivitrol	watsons	weed	whitepowder	Score	Auto_Label
Dat codeine gimme chills 🤠 💙 so cold I give e...	0	0	1	0	...	0	0	0	0	0	0	0	1	0
@BobCorlewTN This is just hysterical bullshit ...	0	0	0	0	...	0	0	1	0	0	0	0	2	1
That #LivePD bust reminded me of 1 of my favor...	0	0	0	1	...	0	0	0	1	0	0	0	5	1
'codeine crazy' is why TM88 probably gets a ca...	0	0	0	0	...	0	0	0	1	0	0	0	1	0

Once we were able to finalize the set of relevant tweets, in phase IV, we applied the LDA topic mining algorithm as noted in the methodology. In that regard, the results showed that the perplexity decreased with an increase in the number of topics but tended to converge at a specific point. This occurred at around 50 topics; hence, we set the number of topics to 50. We examined the LDA model results and manually labelled and grouped 18 topics from among the 50 topics of the public opioid tweets. The topics were labelled by two researchers independently and then reviewed iteratively. Figure 9 shows the word clouds for the top nine topics, with the size of the word representing the unigram TF-IDF score.

For each tweet, the LDA algorithm calculated the probability that tweet *x* belongs to topic *y*. Thereafter, we computed the topics' weights by determining how many tweets belonged to a specific topic. Table 2 shows the distribution of public opioid tweets over the topics. The most prevalent topics were related to the opioid crisis. Many posts were related to topics, such as chronic pain medications, the opioid crisis and how the US government deals with it, opioids drugs coming across the US border, deaths because of overdose due to opioid fentanyl and heroin, opioid treatments, opioid crisis as a real problem, taking opioid medications for health problems such cough, opioid overdose deaths, opioid crisis impact on American communities, and patients suffering from opioid addiction.

Figure 9. Top nine topic word clouds.

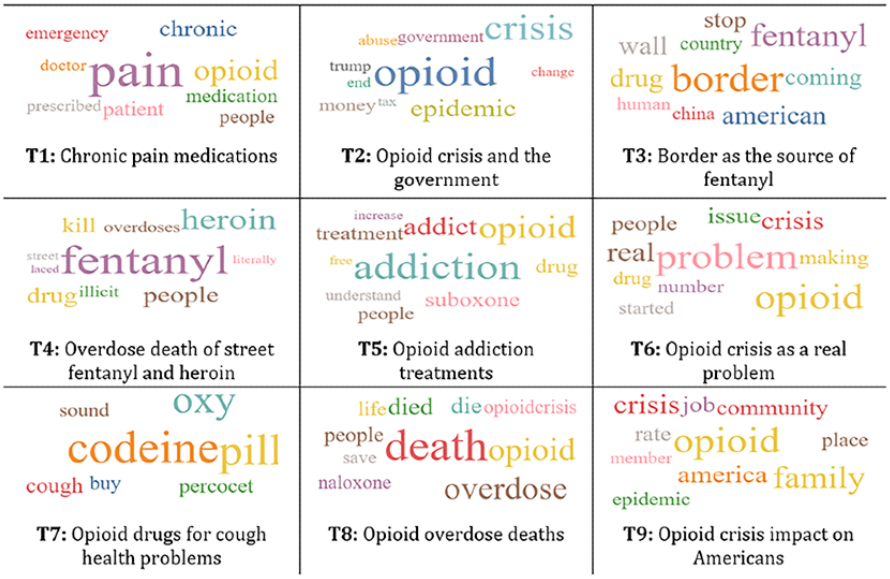


Table 2. Public opioid topic weights.

Topic	Description	Top 10 topic words	Topic weight[a] (N=264,522)
1	Chronic pain medications	Pain, opioid, chronic, med, patient, medication, people, emergency, doctor, and prescribed	40,437 (15.29%)
2	Opioid crisis and government	Opioid, crisis, epidemic, money, government, abuse, trump, end, tax, and change	28,210 (10.66%)
3	Border as the source of fentanyl	Border, fentanyl, drug, wall, American, coming, stop, country, human, and China	25,380 (9.59%)
4	Overdose death from street fentanyl and heroin	Fentanyl, heroin, people, drug, kill, illicit, overdoses, street, literally, and laced	25,226 (9.54%)
5	Opioid addiction treatments	Addiction, opioid, addict, suboxone, drug, treatment, people, understand, free, and increase	18,160 (6.87%)
6	Opioid crisis as a real problem	Problem, opioid, real, crisis, issue, people, making, number, drug, and started	13,966 (5.28%)
7	Opioid drugs for cough and other health problems	Codeine, pill, oxy, shit, cough, percocet, sex, yall, buy, and sound	13,950 (5.27%)
8	Opioid overdose deaths	Death, overdose, opioid, died, die, people, life, naloxone, opioid crisis, and save	13,179 (4.98%)
9	Opioid crisis impact on Americans	Opioid, family, crisis, America, job, rate, community, place, epidemic, and member	11,214 (4.24%)
10	Patient suffering from opioid prescriptions	Patient, doctor, opioid, cancer, prescribing, control, doc, suffering, opioids, and suicide	10,909 (4.12%)
11	Taking opioids after surgery or hospitalization	Day, morphine, feel, surgery, hospital, gave, time, home, needed, and sick	10,326 (3.90%)
12	Illegal market for getting prescription drugs	Drug, prescription, illegal, street, market, dealer, law, opioid, supply, and sell	9948 (3.76%)
13	Legalizing medical marijuana and cannabis	Medical, marijuana, opioid, cannabis, legal, research, pot, study, state, and cannabidiol	9129 (3.45%)
14	People dying from opioids	People, opioid, white, dying, news, crime, crack, folk, black, and house	9012 (3.41%)
15	Public health and substance	Care, health, opioid, substance, public, worse, world, guy, mental, and vote	8065 (3.05%)
16	Opioid addiction and withdrawal	Addicted, opiate, percocet, week, people, opioid, withdrawal, hooked, thinking, and common	7662 (2.90%)
17	Methadone clinic solutions for addiction	High, methadone, solution, fix, clinic, level, crazy, gone, heroine, and wait	4912 (1.86%)
18	School health care education programs	Today, program, school, access, healthcare, policy, jail, recovery, act, and education	4837 (1.83%)

[a]The number represents the number of tweets in each topic, and the percentage represents the proportion of tweets with respect to all tweets.

Discussion

Principal Findings

As shown in Table 1, use of the data quality evaluation matrix resulted in a distinct improvement, as depicted by the improvements in accuracy, precision, and F-measure. The slight decline in recall is an artifact of the metric, where all posts generated using the search query were predicted as relevant. The F-measure, which captures precision and recall, depicted a 10% increase in the classification relevance when the evaluation matrix was used.

With respect to the case study, the topics identified using the LDA topic modeling algorithm match the exiting topics in the literature, such as efforts by the former president of the United States to address the opioid epidemic, promotion and legalization of marijuana as an effective alternative for managing pain, marijuana as an alternative to opioids, roles of foreign countries in the epidemic and production of synthetic opioids, advertisements promoting opioid recovery programs [10], and opioids as “medications” to alleviate pain [28].

With respect to medication and pain management, topics 1, 7, 10, and 11 captured the public’s concerns about the prescriptions of chronic pain opioid medications for emergency health conditions and surgery-related chronic pain management. The strategy to address such issues is to change the policies for prescribing chronic pain medications. Example policies and strategies include promoting the responsible use of opioids, reducing the supply of opioids, implementing drug take-back

programs [29], tracking and monitoring prescription drug abuse, and reporting electronic prescriptions [30].

With respect to prescription drug abuse, topics 4, 8, 12, and 14 dominated discussions about overdose deaths from taking opioids and illicit drugs, such as heroin, offered by street dealers and other illegal venues. Tightening monitoring and control over such venues (particularly online) can play a major role in the availability of such drugs and can ultimately reduce fatal cases of drug overdose.

With respect to the opioid drug crisis in the United States, topics 2, 3, 6, and 9 reflected discussions focusing on the opioid use crisis, how opioid drugs coming across the US border exacerbate the crisis, and the US government's actions toward this crisis. Interventions to solve such a situation can involve supporting existing agencies such as customs and law enforcement units and drug interdiction agencies. The results are consistent with findings from the study by Glowacki et al [10], where many discussion topics were related to efforts from the former president of the United States to address the opioid epidemic, warnings from the Food and Drug Administration about mixing opioids with sedatives, and attempts from opioid makers to stop the legalization of marijuana.

With respect to compulsive drug seeking and use treatments, topics 5, 17, and 18 relate to opioid addiction treatments and resources that provide such treatments. Providing individuals and clinics with information about opioid treatment programs and increasing the number of providers of such programs can help in mitigating opioid addiction and overdose problems. Furthermore, providing schools with health care education programs about opioid addiction and recovery programs can create awareness in the community.

With respect to drug legalization for medical or recreational use, topic 13 involved discussions to legalize marijuana for medical and recreational use instead of using opioid prescriptions. Such discussions are in agreement with findings and recommendations regarding the promotion and legalization of marijuana as an effective alternative for managing pain [10,28], as well as advertisements promoting opioid recovery programs [10,31].

Our analysis identified specific discussions that were not identified by prior research. These discussions were mainly about the public's awareness of problems with opioid overdose, its causes, and its consequences; the benefits of rehabilitation clinics as solutions for opioid addiction and overdose; and the need for health care educational programs at schools.

Overall, the most discussed topics in the analysis can help in understanding the different concerns that the public has around the opioid crisis in the United States. This can serve as a key input for defining and implementing innovative solutions and strategies to address the opioid epidemic.

Conclusions

Online social media are rich sources of data on an individual's daily activities and lifestyle. Applying text mining techniques can help in understanding the concerns of online social communities. This study aimed to formulate a systematic analysis approach to obtain good quality social media data sets of drug abuse. We developed a social media text mining framework for drug abuse. We addressed how the framework can help in solving associated challenges related to topic detection and data quality. Further, we demonstrated the applicability of our proposed framework to identify the common concerns toward the opioid epidemic, and we addressed the most discussed topics on social media related to opioids. The insights from the daily posts of public and opioid-addicted social media network users can help provide better opioid prevention, treatment, and recovery strategies. From an information systems perspective, the framework and associated processes can be applied to other domains where there are challenges associated with topic identification and data quality. This research strengthens public health data reporting and collection through social media. With regard to the broader impact of the research results, we expect better insights into drug abuse epidemics.

Theoretical and Practical Implications

From a theoretical perspective, this research highlights the importance of developing and adapting text mining techniques to social media data analytics for drug abuse. A particular significance is the emphasis on developing methods for improving the discovery and identification of topics in social media domains characterized by a plethora of highly diverse terms and lack of a commonly available dictionary or language by the community, such as in the case of opioid and drug abuse. The framework addresses problems associated with data quality in such contexts and can be applied to other domains where there are challenges associated with topic identification and data quality.

From a practical perspective, automatically analyzing social media users' posts can help decision-makers to understand the public themes and topics that exist in online communities. Addressing the most discussed topics on social media related to drug abuse, such as the opioid epidemic, can help understand the problem dimensions and create proper strategies. Moreover, classifying the online social activities of people who are addicted or have been addicted to opioids can help understand the nature of their issues of misusing or overdosing opioid prescriptions, as well as understand user experience. This can help in identifying their concerns and their common issues. Furthermore, it can help in understanding different themes, such as the ways that lead individuals to be addicted, the illicit ways that they obtain opioids, the management of their addiction (if they do manage it), the kinds of medications they use to recover, the other drugs they use or are addicted to, and the types of opioids they are addicted to and their percentages.

Limitations and Future Research

Additional refinement of the data quality evaluation matrix is needed. For the study case analysis, additional refinement of the defined categories can further support the applicability of the framework. Finally, there is a need to supplement the collected data with surveys of opioid users to better understand their specific concerns and experiences.

Future research aims to explore the proposed framework using different social media platforms to discover the relations

between “opioid” online communities and other online health communities, such as “chronic pain,” “posttraumatic stress disorder,” and “anxiety.” Such communities could have a strong relation with people addicted to opioids and could further improve the effectiveness of detecting drug abuse topics from users’ posts.

Acknowledgments

The authors would like to thank Dakota State University for generously supporting this research by providing access to the Crimson Hexagon platform.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Samples of the opioid drug abuse terms and tweets.
[DOC File , 76 KB - jmir_v22i8e18350_app1.doc]

References

1. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What Patients Can Tell Us: Topic Analysis for Social Media on Breast Cancer. JMIR Med Inform 2017 Jul 31;5(3):e23 [FREE Full text] [doi: 10.2196/medinform.7779] [Medline: 28760725]
2. Wang F, Carley KM, Zeng D, Mao W. Social Computing: From Social Informatics to Social Intelligence. IEEE Intell. Syst 2007 Mar;22(2):79-83. [doi: 10.1109/MIS.2007.41]
3. Kim SJ, Marsch LA, Hancock JT, Das AK. Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data. J Med Internet Res 2017 Oct 31;19(10):e353 [FREE Full text] [doi: 10.2196/jmir.6426] [Medline: 29089287]
4. Kalyanam J, Katsuki T, R G Lanckriet G, Mackey TK. Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning. Addict Behav 2017 Feb;65:289-295. [doi: 10.1016/j.addbeh.2016.08.019] [Medline: 27568339]
5. Sarker A, O'Connor K, Ginn R, Scotch M, Smith K, Malone D, et al. Social Media Mining for Toxicovigilance: Automatic Monitoring of Prescription Medication Abuse from Twitter. Drug Saf 2016 Mar;39(3):231-240 [FREE Full text] [doi: 10.1007/s40264-015-0379-4] [Medline: 26748505]
6. Cherian R, Westbrook M, Ramo D, Sarkar U. Representations of Codeine Misuse on Instagram: Content Analysis. JMIR Public Health Surveill 2018 Mar 20;4(1):e22 [FREE Full text] [doi: 10.2196/publichealth.8144] [Medline: 29559422]
7. Lu J, Sridhar S, Pandey R, Hasan M, Mohler G. Redditors in Recovery: Text Mining Reddit to Investigate Transitions into Drug Addiction. 2018 Presented at: 2018 IEEE International Conference on Big Data; December 10-13, 2018; Seattle, WA, USA. [doi: 10.1109/bigdata.2018.8622552]
8. Park A, Conway M, Chen AT. Examining Thematic Similarity, Difference, and Membership in Three Online Mental Health Communities from Reddit: A Text Mining and Visualization Approach. Comput Human Behav 2018 Jan;78:98-112 [FREE Full text] [doi: 10.1016/j.chb.2017.09.001] [Medline: 29456286]
9. Lokala U, Lamy FR, Daniulaityte R, Sheth A, Nahhas RW, Roden JI, et al. Global trends, local harms: availability of fentanyl-type drugs on the dark web and accidental overdoses in Ohio. Comput Math Organ Theory 2019 Mar;25(1):48-59 [FREE Full text] [doi: 10.1007/s10588-018-09283-0] [Medline: 32577089]
10. Glowacki EM, Glowacki JB, Wilcox GB. A text-mining analysis of the public's reactions to the opioid crisis. Subst Abus 2018;39(2):129-133. [doi: 10.1080/08897077.2017.1356795] [Medline: 28723265]
11. Haug NA, Bielenberg J, Linder SH, Lembke A. Assessment of provider attitudes toward #naloxone on Twitter. Subst Abus 2016;37(1):35-41. [doi: 10.1080/08897077.2015.1129390] [Medline: 26860229]
12. Fan Y, Zhang Y, Ye Y, li X, Zheng W. Social Media for Opioid Addiction Epidemiology: Automatic Detection of Opioid Addicts from Twitter and Case Studies. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017 Presented at: 2017 ACM on Conference on Information and Knowledge Management; November 2017; Singapore p. 1259-1267. [doi: 10.1145/3132847.3132857]
13. Eshleman R, Jha D, Singh R. Identifying individuals amenable to drug recovery interventions through computational analysis of addiction content in social media. In: IEEE. 2017 Presented at: 2017 IEEE International Conference on Bioinformatics and Biomedicine; November 13-16, 2017; Kansas City, MO, USA. [doi: 10.1109/bibm.2017.8217766]
14. Jelodar H, Wang Y, Rabbani M, Xiao G, Zhao R. A Collaborative Framework Based for Semantic Patients-Behavior Analysis and Highlight Topics Discovery of Alcoholic Beverages in Online Healthcare Forums. J Med Syst 2020 Apr 07;44(5):101. [doi: 10.1007/s10916-020-01547-0] [Medline: 32266484]
15. Jenhani F, Gouider MS, Said LB. Hybrid System for Information Extraction from Social Media Text: Drug Abuse Case Study. Procedia Computer Science 2019;159:688-697. [doi: 10.1016/j.procs.2019.09.224]

16. Kazemi D, Borsari B, Levine M, Dooley B. Systematic review of surveillance by social media platforms for illicit drug use. J Public Health (Oxf) 2017 Dec 01;39(4):763-776 [FREE Full text] [doi: 10.1093/pubmed/fdx020] [Medline: 28334848]
17. Stieglitz S, Mirbabaie M, Ross B, Neuberger C. Social media analytics – Challenges in topic discovery, data collection, and data preparation. International Journal of Information Management 2018 Apr;39:156-168. [doi: 10.1016/j.ijinfomgt.2017.12.002]
18. Ghani NA, Hamid S, Targio Hashem IA, Ahmed E. Social media big data analytics: A survey. Computers in Human Behavior 2019 Dec;101:417-428. [doi: 10.1016/j.chb.2018.08.039]
19. Tricco AC, Zarin W, Lillie E, Jeblee S, Warren R, Khan PA, et al. Utility of social media and crowd-intelligence data for pharmacovigilance: a scoping review. BMC Med Inform Decis Mak 2018 Jun 14;18(1):38 [FREE Full text] [doi: 10.1186/s12911-018-0621-y] [Medline: 29898743]
20. El-Gayar O, Nasralah T, El Noshokaty A. IT for diabetes self-management - What are the patients? expectations? 2018 Presented at: 24th Americas Conference on Information Systems (AMCIS 2018); August 16-18, 2018; New Orleans, LA.
21. Austin L, Jin Y. Social Media and Crisis Communication. New York, NY: Taylor & Francis; 2017.
22. Cameron D, Smith GA, Daniulaityte R, Sheth AP, Dave D, Chen L, et al. PREDOSE: a semantic web platform for drug abuse epidemiology using social media. J Biomed Inform 2013 Dec;46(6):985-997 [FREE Full text] [doi: 10.1016/j.jbi.2013.07.007] [Medline: 23892295]
23. Blei D, Ng A, Jordan M. Latent Dirichlet Allocation. Journal of Machine Learning Research 2003;3(Jan):993-1022.
24. Hopkins D, King G. A Method of Automated Nonparametric Content Analysis for Social Science. American Journal of Political Science 2010;54(1):229-247. [doi: 10.1111/j.1540-5907.2009.00428.x]
25. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012;22(3):276-282. [doi: 10.11613/bm.2012.031]
26. Han J, Pei J, Kamber M. Data Mining: Concepts and Techniques. New York: Elsevier; 2011.
27. Arun R, Suresh V, Veni Madhavan CE, Narasimha Murthy MN. On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In: Zaki MJ, Yu JX, Ravindran B, Pudi V, editors. Advances in Knowledge Discovery and Data Mining. PAKDD 2010. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2010:391-402.
28. Lossio-Ventura JA, Bian J. An inside look at the opioid crisis over Twitter. 2018 Presented at: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); December 3-6, 2018; Madrid.
29. Penm J, MacKinnon NJ, Boone JM, Ciaccia A, McNamee C, Winstanley EL. Strategies and policies to address the opioid epidemic: A case study of Ohio. J Am Pharm Assoc (2003) 2017;57(2S):S148-S153 [FREE Full text] [doi: 10.1016/j.japh.2017.01.001] [Medline: 28189539]
30. Phillips J. Prescription drug abuse: problem, policies, and implications. Nurs Outlook 2013;61(2):78-84. [doi: 10.1016/j.outlook.2012.06.009] [Medline: 23245611]
31. Russell D, Spence NJ, Thames KM. 'It's so scary how common this is now:' frames in media coverage of the opioid epidemic by Ohio newspapers and themes in Facebook user reactions. Information, Communication & Society 2019 Jan 16;22(5):702-708. [doi: 10.1080/1369118x.2019.1566393]

Abbreviations

LDA: latent Dirichlet allocation
TF-IDF: term frequency-inverse document frequency

Edited by G Eysenbach; submitted 21.02.20; peer-reviewed by A Wahbeh, A Sarker; comments to author 06.04.20; revised version received 12.05.20; accepted 04.06.20; published 13.08.20.

Please cite as:
Nasralah T, El-Gayar O, Wang Y
Social Media Text Mining Framework for Drug Abuse: Development and Validation Study With an Opioid Crisis Case Analysis
J Med Internet Res 2020;22(8):e18350
URL: https://www.jmir.org/2020/8/e18350
doi:10.2196/18350
PMID:32788147

Short Paper

Characteristics of Twitter Use by State Medicaid Programs in the United States: Machine Learning Approach

Jane M Zhu[1], MD, MPP, MSHP; Abeed Sarker[2], PhD; Sarah Gollust[3], PhD; Raina Merchant[4], MD, MSc; David Grande[5], MD, MPA

[1]Division of General Internal Medicine, Oregon Health and Science University, Portland, OR, United States

[2]Department of Biomedical Informatics, Emory University, Atlanta, GA, United States

[3]Division of Health Policy and Management, University of Minnesota, Minneapolis, MN, United States

[4]Center for Digital Health, University of Pennsylvania, Philadelphia, PA, United States

[5]Division of General Internal Medicine, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Jane M Zhu, MD, MPP, MSHP
Division of General Internal Medicine
Oregon Health and Science University
3181 SW Sam Jackson Park Road
Portland, OR, 97239
United States
Phone: 1 5034945040
Email: zhujan@ohsu.edu

Abstract

Background: Twitter is a potentially valuable tool for public health officials and state Medicaid programs in the United States, which provide public health insurance to 72 million Americans.

Objective: We aim to characterize how Medicaid agencies and managed care organization (MCO) health plans are using Twitter to communicate with the public.

Methods: Using Twitter's public application programming interface, we collected 158,714 public posts ("tweets") from active Twitter profiles of state Medicaid agencies and MCOs, spanning March 2014 through June 2019. Manual content analyses identified 5 broad categories of content, and these coded tweets were used to train supervised machine learning algorithms to classify all collected posts.

Results: We identified 15 state Medicaid agencies and 81 Medicaid MCOs on Twitter. The mean number of followers was 1784, the mean number of those followed was 542, and the mean number of posts was 2476. Approximately 39% of tweets came from just 10 accounts. Of all posts, 39.8% (63,168/158,714) were classified as general public health education and outreach; 23.5% (n=37,298) were about specific Medicaid policies, programs, services, or events; 18.4% (n=29,203) were organizational promotion of staff and activities; and 11.6% (n=18,411) contained general news and news links. Only 4.5% (n=7142) of posts were responses to specific questions, concerns, or complaints from the public.

Conclusions: Twitter has the potential to enhance community building, beneficiary engagement, and public health outreach, but appears to be underutilized by the Medicaid program.

(J Med Internet Res 2020;22(8):e18401) doi:10.2196/18401

KEYWORDS

medicaid; public health; health communication; community engagement; social media

Introduction

Approximately 20% of online adults use Twitter [1], a social media platform that allows users to share short messages ("tweets"). With over 500 million tweets per day on topics including health-related issues, Twitter is a potentially valuable tool for public health officials to engage the public. Prior studies suggest that overall social media use is independent of educational attainment, race/ethnicity, education, income, and health care access [2]. Twitter use is also significantly higher in young adults compared with older age groups, which maps well to the Medicaid population, of which 93% are aged <65

years and 80% are aged <45 years [3]. However, evidence suggests that these platforms are currently underutilized. A minority of local public health departments and federal health agencies in the United States use Twitter [4], and these accounts typically have low public engagement [5].

Twitter may be particularly useful for state Medicaid programs, which together constitute the single largest source of public health insurance in the United States, covering 72 million children, older adults, people with disabilities, and low-income populations. Medicaid is currently undergoing a variety of program and policy changes, including eligibility expansion to low-income childless adults; establishment of work requirements for some enrollees; delivery system reforms like accountable care organizations and value-based payment models; and addressing emerging public health priorities like substance use disorder treatment. Given this context, there are significant opportunities for state agencies to better understand, respond to, and engage with the diverse needs of Medicaid-eligible and Medicaid-enrolled populations [6] and to communicate with the general public as a key stakeholder. Despite this potential, to our knowledge no prior study has examined the extent to which Medicaid agencies and health plans are using Twitter to communicate with the public.

Methods

We identified active Twitter profiles of state Medicaid agencies and Medicaid managed care organizations (MCOs), which are health insurance plans that contract with states to provide Medicaid health benefits and related services. We included the latter group because more than 70% of Medicaid beneficiaries receive their care through comprehensive managed care plans. To identify Medicaid MCOs, we used the Kaiser Family Foundation's publicly available Medicaid Managed Care Market Tracker [7], which lists all Medicaid MCOs and their parent firms across all 50 states and the District of Columbia. For each state, we searched Twitter, agency websites, and Google to identify and verify relevant Medicaid agency and MCO accounts. Our focus was on state-level Medicaid programs and plans, so we excluded accounts of higher-level governmental agencies (eg, Department of Health and Human Services), parent firms of managed care plans operating across multiple states, and professional accounts of individual Medicaid administrators and leadership.

We used the Twitter public application programming interface [8] to collect a total of 158,714 public posts by the identified accounts, spanning March 2014 through June 2019. To conduct content analysis on these tweets, we manually coded 800 tweets and developed a coding scheme with 5 broad categories of tweet content (Table 1). Three coders independently coded identical samples of tweets, and resolved disagreements and updated the coding guidelines via discussion. Two coders then reviewed a new set of 5338 tweets (excluding retweets), yielding an intercoder agreement of 0.78 using Cohen over 998 overlapping tweets. Tweets that could not be categorized as belonging to one of the 5 categories were labeled "Other" (n=198).

We used the coded tweets to train and evaluate supervised machine learning algorithms, experimenting with four approaches: naive Bayes, support vector machines, random forests, and an ensemble of these three classifiers, which combined the predictions of the independently trained classifiers using predetermined rules [9]. The ensemble classifier obtained the best performance when evaluated using the manually annotated labels, with an accuracy of 74.1%, and therefore we used this classifier to automatically classify all the posts in our data set.

Results

We identified 96 Twitter accounts, including 15 state Medicaid agencies and 81 Medicaid MCOs (out of 51 total state Medicaid agencies and 323 Medicaid MCOs). Among these accounts, the mean number of followers was 1784 (range 2-38,352, median 1434), the mean number of those followed was 542, and the mean number of posts was 2476. Twitter accounts had been active for a mean of 79 months, and approximately 39% of tweets came from just 10 accounts.

The most active accounts in terms of number of tweets were those of health plans, including Fidelis (New York, @FidelisCare, 14,748), HAP Midwest Health Plan Inc (Michigan, @hapmichigan, 12,598), Hawaii Medical Service Association (Hawaii, @AskHMSA, 12,568), Blue Cross Blue Shield of Texas (Texas, @BCBSTX, 12,157), and UPMC Health Plan (Pennsylvania, @UPMCHealthPlan, 8302). Top accounts in terms of follower count were also health plans, including Fidelis (38,352), UPMCHealthPlan (14,845), Blue Cross Blue Shield of Texas (9355), Seton Health Plan (Texas, @setonfamily, 8326), and Anthem Blue Cross Partnership Plan (California, @AnthemBC_News, 7173). In comparison, top Medicaid agencies in terms of follower count included Ohio (@OhioMedicaid, 3165), Massachusetts (@MassHealth, 2223), Washington (@WA_Health_Care, 1858), Colorado (@CHCPF, 1794), and Georgia (@GADCH, 1596).

Of all posts (N=158,714), 39.8% (n=63,168) were classified as general public health education and outreach; 23.5% (n=37,298) constituted outreach about specific Medicaid policies, programs, services, or events; 18.4% (n=29,203) contained relationship-building content including organizational promotion of staff and activities; and 11.6% (n=18,411) contained general news and news links (Table 1). Only 4.5% (n=7142) of the tweets were customer service responses to specific questions, concerns, or complaints from other users. Additionally, 2.2% (n=3492) of the tweets did not fit within these categories and were classified as "Other."

The majority of posts received little or no engagement from the public, with only 1% of tweets having 9 or more likes, or 6 or more retweets. Tweets with more active engagement were more likely to contain words that conveyed actions directed at the public, like "email," "assist," "send," and "contact" (see sample tweets in Table 1).

Table 1. Frequency and types of posts from Medicaid agencies and managed care organizations on Twitter (N=158,714)[a].

Category	Definition	Example	Posts, n (%)
Public health education	Public health announcements about prevention, disease conditions, vaccinations, resources, and provision of general health-related information or health-seeking behaviors.	Learn the symptoms for each type of gynecologic cancer. If you have some of the symptoms, talk to your provider. https://t.co/UPxaBrxY5R	63,168 (39.8)
Outreach about specific Medicaid policies and programs	Information to raise awareness of specific services, programs, and events offered by Medicaid, directed toward consumers	• All SoonerCare members are required to have a valid Oklahoma mailing address on file Don't have your Oklahoma address listed on your account? Update your information today at https://t.co/AEkZDU7W4E. • Unsure if your insurance is still active or if you need to reapply? Call #HUSKYHealth at 1.800.859.9889 for help.	37,298 (23.5)
Relationship building or promotional	Posts directed at a general audience that intends to elicit consumer engagement (responses, replies, likes, feedback), including promotion of agency staff and accomplishments, and greetings.	• What would you want to improve about your Medicaid program? • Like us on Facebook. Get access to health tips, recipes, and events happening around the state. Visit: https://t.co/0XI6PEfhcq	29,203 (18.4)
General news and announcements	Press releases, links to news articles, general announcements	New post (Three regional agencies team up to support MassHealth consumers with disabilities, complex medical needs) has been published on Advocate News Online - https://t.co/JMfIU0sdhH	18,411 (11.6)
Customer service and responses to constituents	Responses from the plan or agency directed toward a specific question or complaint raised by a consumer	@[redacted] Renewals are done yearly. You will be notified when it's time to renew your coverage.	7142 (4.5)

[a]Note that 2.2% (n=3492) of the tweets did not fit within these categories and were classified as "Other," which is not shown in this table.

Discussion

Principal Findings

Social media is one way for Medicaid programs to reach constituents and to launch low-cost health promotion and public engagement campaigns in the face of resource constraints. Our results suggest that despite Twitter's reach among the general public, it is underutilized in the Medicaid program, compared to other public health organizations [10,11].

Twitter has the potential to enhance community building and engagement, which is increasingly a priority for Medicaid programs. Prior work found that two-way communication on Twitter between public health entities and constituents led to an increase in action and awareness that, in turn, resulted in an improvement in community health [12]. However, many public health entities use social media simply to broadcast information without engaging audiences [4,10]. While we found that some Medicaid programs are using Twitter, the majority have relatively few followers and overall low engagement with the public. A number of studies point to methods to increase public engagement on Twitter. A study of US children's hospitals using social media found improvements in engagement when posts included pictures and content featuring patient narratives and community partnerships [13]. Other studies in the health care sector have found increased engagement with posts that feature hashtags, URLs, and user mentions [4].

Our study has limitations. Although we used multiple search strategies to verify accounts, we may be missing some active Medicaid accounts. We also excluded a number of umbrella state agencies on Twitter whose activities may include some degree of Medicaid program oversight and outreach, which may lead to underestimation of Medicaid agency presence and activity. Additionally, we used machine learning algorithms to automate content analysis of a large set of tweets, and there may be some misclassification of text across coding categories. Finally, our study focused on Twitter and is not generalizable to other social media platforms that Medicaid agencies and health plans may be using to engage with the public.

Conclusions

Changes to the Medicaid program have accelerated a number of efforts to increase consumer engagement around disease and care management, cost-sharing, and health care behavior change. As state Medicaid programs confront important public health challenges and expand to serve new populations, social media represents an important and underutilized tool for engaging both Medicaid-eligible individuals and the general public. Future research is needed to understand how social media platforms like Twitter can help these programs improve community engagement around public health programming and interventions.

Acknowledgments

The authors would like to thank Alexis Upshur and Vivek Kumar for their assistance with annotating Twitter post content, and Haitao Cai for assistance with data collection. This study was supported by the Robert Wood Johnson Foundation (#76242). The study sponsor had no role in study design; collection, analysis, and interpretation of data; writing the report; or the decision to submit the report for publication.

Conflicts of Interest

None declared.

References

1. Perrin A, Anderson M. Share of US adults using social media, including Facebook, is mostly unchanged since 2018. FactTank.: Pew Research Center; 2019 Apr 10. URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2020-04-22]
2. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
3. Medicaid and CHIP Beneficiaries at a Glance. Health Care Quality Measures.: Centers for Medicare and Medicaid Services; 2020 Feb. URL: https://www.medicaid.gov/medicaid/quality-of-care/downloads/beneficiary-ataglance.pdf [accessed 2020-03-16]
4. Bhattacharya S, Srinivasan P, Polgreen P. Engagement with health agencies on Twitter. PLoS One 2014 Nov 7;9(11):e112235 [FREE Full text] [doi: 10.1371/journal.pone.0112235] [Medline: 25379727]
5. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH. Use of Twitter among local health departments: an analysis of information sharing, engagement, and action. J Med Internet Res 2013 Aug 19;15(8):e177 [FREE Full text] [doi: 10.2196/jmir.2775] [Medline: 23958635]
6. Harris JK, Choucair B, Maier RC, Jolani N, Bernhardt JM. Are public health organizations tweeting to the choir? Understanding local health department Twitter followership. J Med Internet Res 2014 Feb 26;16(2):e31 [FREE Full text] [doi: 10.2196/jmir.2972] [Medline: 24571914]
7. Fichtenberg CM, Alley DE, Mistry KB. Improving Social Needs Intervention Research: Key Questions for Advancing the Field. Am J Prev Med 2019 Dec;57(6 Suppl 1):S47-S54 [FREE Full text] [doi: 10.1016/j.amepre.2019.07.018] [Medline: 31753279]
8. Medicaid Managed Care Market Tracker. State Health Facts.: Henry J Kaiser Family Foundation; 2019. URL: https://www.kff.org/data-collection/medicaid-managed-care-market-tracker/ [accessed 2020-02-19]
9. Yoon S, Elhadad N, Bakken S. A practical approach for content mining of Tweets. Am J Prev Med 2013 Jul;45(1):122-129 [FREE Full text] [doi: 10.1016/j.amepre.2013.02.025] [Medline: 23790998]
10. Dietterich T. Ensemble Methods in Machine Learning. In: Multiple Classifier Systems. Berlin, Heidelberg: Springer; 2000:1-15.
11. Thackeray R, Neiger BL, Smith AK, Van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012 Mar 26;12(1):242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]
12. Thackeray R, Neiger BL, Burton SH, Thackeray CR. Analysis of the purpose of state health departments' tweets: information sharing, engagement, and action. J Med Internet Res 2013 Nov 11;15(11):e255 [FREE Full text] [doi: 10.2196/jmir.3002] [Medline: 24217361]
13. Wong CA, Ostapovich G, Kramer-Golinkoff E, Griffis H, Asch DA, Merchant RM. How U.S. children's hospitals use social media: A mixed methods study. Healthc (Amst) 2016 Mar;4(1):15-21. [doi: 10.1016/j.hjdsi.2015.12.004] [Medline: 27001094]

Abbreviations

MCO: managed care organization

Edited by G Eysenbach; submitted 24.02.20; peer-reviewed by MG Kim, C Mather; comments to author 20.04.20; revised version received 11.05.20; accepted 03.06.20; published 17.08.20.

Please cite as:
Zhu JM, Sarker A, Gollust S, Merchant R, Grande D
Characteristics of Twitter Use by State Medicaid Programs in the United States: Machine Learning Approach
J Med Internet Res 2020;22(8):e18401
URL: http://www.jmir.org/2020/8/e18401/
doi:10.2196/18401
PMID:32804085

Original Paper

Viewing Trends and Users' Perceptions of the Effect of Sleep-Aiding Music on YouTube: Quantification and Thematic Content Analysis

Ransome Eke[1*], MD, PhD; Tong Li[2*], BSc; Kiersten Bond[1*], BA, MA; Arlene Ho[2*], BSc; Lisa Graves[2*], MD

[1]Department of Health Science, University of Alabama, Tuscaloosa, AL, United States

[2]Western Michigan University School of Medicine, Kalamazoo, MI, United States

[*]all authors contributed equally

Corresponding Author:
Ransome Eke, MD, PhD
Department of Health Science
University of Alabama
105 Russell Building
Tuscaloosa, AL, 35487
United States
Phone: 1 205 348 2553
Email: reke@ua.edu

Abstract

Background: Sleep plays an essential role in the psychological and physiological functioning of humans. A report from the Centers for Disease Control and Prevention (CDC) found that sleep duration was significantly reduced among US adults in 2012 compared to 1985. Studies have described a significant association between listening to soothing music and an improvement in sleep quality and sleep duration. YouTube is a platform where users can access sleep-aiding music videos. No literature exists pertaining to the use of sleep-aiding music on YouTube.

Objective: This study aimed to examine the patterns of viewing sleep-aiding music videos on YouTube. We also performed a content analysis of the comments left on sleep-aiding music video posts, to describe the perception of users regarding the effects of these music videos on their sleep quality.

Methods: We searched for sleep-aiding music videos published on YouTube between January 1, 2012, and December 31, 2017. We sorted videos by view number (highest to lowest) and used a targeted sampling approach to select eligible videos for qualitative content analysis. To perform the content analysis, we imported comments into a mixed-method analytical software. We summarized variables including total views, likes, dislikes, play duration, and age of published music videos. All descriptive statistics were completed with SAS statistical software.

Results: We found a total of 238 sleep-aiding music videos on YouTube that met the inclusion criteria. The total view count was 1,467,747,018 and the total playtime was 84,252 minutes. The median play length was 186 minutes (IQR 122 to 480 minutes) and the like to dislike ratio was approximately 9 to 1. In total, 135 (56.7%) videos had over 1 million views, and 124 (52.1%) of the published sleep-aiding music videos had stayed active for 1 to 2 years. Overall, 4023 comments were extracted from 20 selected sleep-aiding music videos. Five overarching themes emerged in the reviewed comments, including viewers experiencing a sleep problem, perspective on the positive impact of the sleep-aiding music videos, no effect of the sleep-aiding music videos, time to initiation of sleep or sleep duration, and location of viewers. The overall κ statistic for the codes was 0.87 (range 0.85-0.96).

Conclusions: This is the first study to examine the patterns of viewing sleep-aiding music videos on YouTube. We observed a substantial increase in the number of people using sleep-aiding music videos, with a wide variation in viewer location. This study supports the hypothesis that listening to soothing music has a positive impact on sleep habits.

(J Med Internet Res 2020;22(8):e15697) doi:10.2196/15697

KEYWORDS

insomnia; sleep deprivation; YouTube; utilization; pattern; perception; content analysis

Introduction

Sleep plays an essential role in the psychological and physiological functioning of humans. Adults need 7 or more hours of sleep per night to be physically and mentally healthy [1]. The Centers for Disease and Control and Prevention (CDC) reports that sleep duration in adults substantially decreased from 1985 to 2012 [2]; in total, 35% of adults in the United States have a short sleep duration, which is defined as less than 7 hours per night [1]. Low sleep duration, or sleep deprivation, is associated with cognitive issues (including decreased task performance) and health issues such as diabetes, depression, obesity, cardiovascular disease, decreased cognitive performance, and decreased immune function [3-9].

Chronic sleep deprivation could be due to sleep disorders, such as insomnia. Insomnia includes difficulty falling asleep, trouble staying asleep, waking up too early, and poor quality of sleep [10], which directly affects an individual's sleep duration. Insomniacs (persons diagnosed with insomnia) have higher rates of work absenteeism, decreased quality of life, and increased health care utilization when compared to good sleepers [11-13]. Insomnia can be treated with pharmacological, nonpharmacological, or both treatment methods, depending on the factors that trigger insomnia for that individual [14].

Nonpharmacological therapies are useful for increasing total sleep time or decreasing sleep onset latency and could help cut down on health care utilization costs and also offer fewer side effects when compared to pharmacological methods [14]. One nonpharmacological approach that appears to be effective in improving sleep quality is listening to music. Previous studies have found a significant association between an improvement in sleep quality and sleep duration and listening to soothing music [15,16]. Further, studies show that listening to soothing music for 45 minutes before bedtime may facilitate relaxation of the body and decrease serum cortisol by reducing stress, leading to improved quality of sleep in adults experiencing insomnia [17,18].

YouTube is a popular video-sharing platform, and the third most visited social media site in the world. Users can upload their own videos to YouTube, including sleep-aiding music videos. This platform allows users to share comments such as life experiences and relevant health information [19-22]. Published commentary about YouTube indicates that general usage increased about tenfold in 2017 compared to 2012; furthermore, it is estimated that over 1 billion hours of videos are watched daily, with about 400 hours of video uploaded each minute [23]. This medium combines fundamental technical features with a community formation function, allowing content creators to upload their videos to YouTube, while the company enables the delivery of this content to millions of viewers.

To date, no data exists describing the viewership of sleep-aiding music videos posted on YouTube and their impact on the sleep quality of users. Therefore, the primary aim of this study is to examine user viewing patterns of sleep-aiding music videos posted on YouTube. Furthermore, this study aims to describe, through content analysis, the perceptions of users regarding the effect that listening to sleep-aiding music videos has on sleep quality.

Methods

Data Source

We obtained data for this study from videos and comments posted on YouTube from January 1, 2012, to December 31, 2017. For ethical research purposes, this study is considered a nonhuman subject and therefore was exempt from institutional review board review.

Search Approach

We searched YouTube video titles and descriptions for sleep-aiding music videos. The following keywords and phrases were used to search the related music videos: sleep, sleeping, music, soothing, and relaxing. These keywords were searched using Boolean logic "AND" and "OR" connectors. First, we searched for the keywords in the YouTube video title. Second, we searched with "the exact phrase." Lastly, we searched with "all of the words" without a specific order.

The most common phrases used to reflect sleep-aiding music videos in YouTube post titles and descriptions included "sleep music," "sleeping music," "soothing music," "relaxing music," and "music for insomnia."

Inclusion Criteria

We selected videos posted in the English language from January 1, 2012, to December 31, 2017, by a YouTuber (an unofficial term used to describe people who create content and upload a video to YouTube). We only included videos that allowed active commenting at that time.

Exclusion Criteria

We excluded YouTube videos posted by marketing agents or videos that promoted commercial content. Other exclusion criteria are duplicate videos, live streaming videos, languages other than English, videos with inappropriate or offensive materials, and videos that had disabled comments or ratings on the post. Additionally, we excluded similar comments that appeared multiple times across different videos.

Data Extraction

We extracted data for descriptive statistics with the free, publicly available YouTube Comment Scraper project created by Philip Klostermann. This web client, licensed under the Internet Software Consortium (ISC), is written in Node.js (an open-source development platform for executing JavaScript code server-side) and uses the YouTube comment application programming interface (API) module to gain access to the comments. Given a YouTube video URL, the client will request all comments for that video from the API. Details about the YouTube Comment Scrapper project coding for all personal and local use is provided elsewhere [24].

We added available videos that met the study's inclusion criteria to a private playlist. We clicked through each video and downloaded the parameters in comma-separated values (CSV) format. Parameters include music video title, view count, date

of publication, comment text, number of likes, and dislikes. We generated a macro and used a macro recorder function in Microsoft Excel (Microsoft Corp) to automate the repetitive data extraction process. The age of the music video was estimated using the publication date.

To obtain data for content analysis, we selected videos using targeted sampling. Targeted sampling is an iterative procedure that assesses video features at several points, allowing adjustment to obtain a final sample similar to that of the hypothesized target population [25]. We defined the target sample for this study as the most viewed sleep-aiding YouTube music videos posted between January 1, 2012, and December 31, 2017. First, we sorted music videos by view number (highest to lowest) using the sort option on YouTube and relevance to our study. Next, using a random number generator and targeted sampling approach, we selected the 20 most viewed sleep-aiding music videos from the available videos.

As observed in a previous study, the criterion of most views is appropriate because an evaluation of these videos is likely to have the most significant impact (eg, a large number of viewers and comments); thus, this is a reasonable approach to the first evaluation of the media source [26]. Finally, we imported the Excel file into NVivo for Windows (Version 12 Plus; QSR International), a mixed-methods data analysis software [27]. We used the Word Cloud feature in NVivo 12 Plus to observe the most commonly used terms in the music video titles as well as the comments to get a picture of emerging concepts and ideas (Multimedia Appendices 1 and 2).

Data Analysis

Descriptive statistics were generated based on the total number of videos posted, year of publication, video length, length of time since a given video was posted, number of views, and number of likes or dislikes. We summarized categorical variables as frequency and percentages, while numerical variables were summarized as means and standard deviation. SAS (Version 9.4; SAS Institute) was used to conduct all descriptive analyses.

Based on the word cloud of frequently used words in comments, two trained coders (authors TL and RE) applied a method described in a previous publication by Burla and colleagues [28] to code the extracted text. The coders excluded replies to individual comments due to their sheer number and structural challenges to storing these comments. Next, the coders sorted comments by character length before reading through relevant comments. Coders focused only on comments that described a user's personal experience when listening to the sleep-aiding music video. The coders used NVivo to organize study data by creating nodes, sorting and reordering nodes within the same level of hierarchy and ratings, and merging nodes into other existing or new nodes. This enabled us to rate the music videos independently and identify recurring themes and patterns. The coders reviewed the video comments independently and documented codes in an Excel spreadsheet. We used the Microsoft Excel macro function to merge and match the comments under each code. The codes were developed using an inductive reasoning approach and discussed among the research team. Further, we computed the total number of relevant comments for each code. In addition, the coders examined any disparities of judgment to reach a consensus resolution. Lastly, the codes were merged into broader categories, including the positive and negative connotations of the comments, such as perceived pleasing and displeasing experiences.

We compared coding generated between coders and performed intercoder reliability assessment. For reference purposes only, we compared our themes with the automated items generated by NVivo.

Results

Descriptive Statistics

After applying the filter option with the dates January 1, 2012, and December 31, 2017, we identified 238 YouTube sleep-aiding music videos that met the inclusion criteria for our study. Of the eligible music videos, there were a total of 1,467,747,018 view counts and a total playtime of 84,252 minutes. The median play length was 186 minutes (IQR 122 to 480 minutes), the like to dislike ratio was approximately 9 to 1 (Figure 1), 56.7% (135/238) had over 1 million views, and 52.1% (124/238) of the published sleep-aiding music videos had stayed active for 1 to 2 years (Table 1).

The proportion of total eligible music videos increased from 4.2% (10/238) in 2012 to 59.7% (142/238) in 2017. The average play length was 120 minutes and 25.6% (61/238) of the videos had been active for over 2 years.

Figure 1. Comparison of the number of video likes versus dislikes by year of viewing.

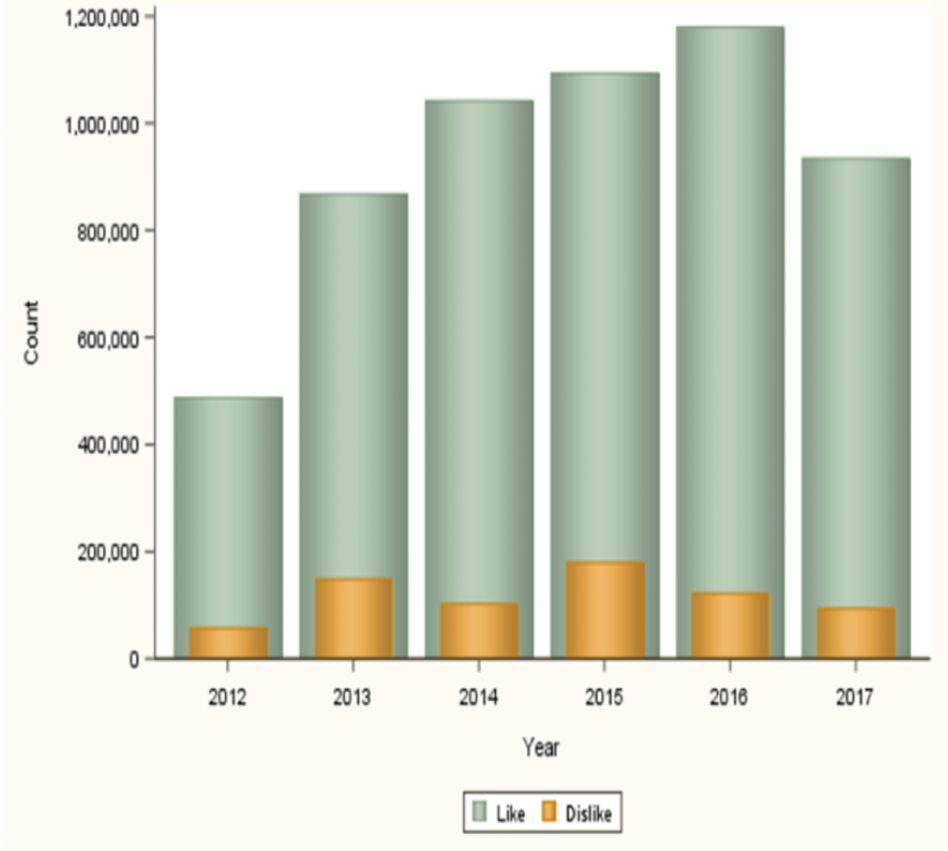


Table 1. Summary of sleep-aiding YouTube music videos included in the study analysis (N=238).

Variable	Values, n (%)
Length of music (minutes)	
<60	22 (9.2)
60-120	32 (13.5)
120	184 (77.3)
Duration of post (years)	
<1	43 (18.1)
1-2	124 (52.1)
2	71 (29.8)
Number of video views (million)	
<0.5	71 (29.8)
0.5-1	32 (13.5)
1	135 (56.7)
Year posted	
2012	10 (3.4)
2013	13 (4.4)
2014	21 (7.1)
2015	19 (6.5)
2016	33 (11.2)
2017	142 (48.3)

Content Analysis

We considered only comments from 20 sleep-aiding music videos selected via the targeted sampling approach (Multimedia Appendix 3). The chosen videos had at least 100 comments relevant to our study. Five overarching themes emerged in the reviewed comments (Table 2): experiencing sleep problem, perspective on the positive impact of the sleep-aiding music videos, the perspective of no effect of the sleep-aiding music videos, opinion of time to initiation of sleep or sleep duration, and location of viewers. Table 3 describes the coding frequency for coded themes and interrater agreement rates. The overall κ statistic for the codes was 0.87 (range 0.85-0.96).

Table 2. Classification of themes from YouTube sleep-aiding music video comments.

Theme	Definition
Experiencing a sleeping problem	This theme is an indication of the viewers' perception of whether they are experiencing sleep deprivation or insomnia
Perspective on the positive impact of the sleep-aiding music videos	Content that serves as an indication that listening to the sleep-aiding music video on YouTube helped the viewer to sleep
The perspective of no impact of the sleep-aiding music videos	Content that serves as an indication that listening to a sleep-aiding music video on YouTube did not help the viewer to sleep or did not have any effect on sleep
The opinion of time to initiation of sleep or sleep duration	Content that serves as an indication of how long it took the viewer to sleep or how long the viewer slept after listening to the sleep-aiding music video (ie, viewer slept within minutes or had a longer sleep duration while listening to the music video)
Variation in the location of viewers	Content that serves as an indication of where the viewer is located while listening to the sleep-aiding music video

Table 3. Comments coding frequency, agreement rate, and percentage of nonassignment by coders[a].

Code	Coding description	Coding frequency, n	Agreement rate (%)	Nonassignment caused by coder 1 (%)	Nonassignment caused by coder 2 (%)
001	Indication of whether the viewer expressed having a problem with sleeping (deprivation or insomnia)	670	69	57	43
002	Positive experience with viewing sleep-aiding music video on YouTube (listening to the music helped the viewer sleep)	2805	86	61	39
003	Negative experience with viewing sleep-aiding music video on YouTube (did not help the viewer sleep, did not have any effect on sleep)	437	74	48	52
004	Indication of how long it took the viewer to sleep or duration of sleep while listening to the sleep-aiding music video (ie, viewer slept within minutes or had a longer sleep duration)	786	93	50	50
005	An indication of whether the viewer mentioned his or her location in the comment	1236	100	65	35

[a]The data encompasses 20 selected music videos and 4023 comments. Overall code agreement rate: 84%. Intercoder reliability for all 5 codes: κ=0.87.

Experiencing Sleeping Problem

Within this theme, viewers' comments expressed whether they were experiencing any difficulty with sleeping, such as insomnia, or an inability to sleep or maintain an adequate duration of sleep. Over 16.6% (670/4024) of the comments reviewed contained messages expressing trouble sleeping. Additionally, some comments (362/4024, 9.0%) described the severity of the sleep problem. Some viewers' comments (604/4024, 15.0%) indicated the use of music videos for babies or children experiencing sleep difficulty. A sample of coded comments is presented here:

Thank you for sharing this! I suffer from insomnia and anxiety. My doctor prescribed meds to help me sleep. This has a calming effect and helps me sleep. Thank you so much! [Ref 8]

I literally cannot fall asleep unless i listen to this. works every time [Ref 1]

I am a sufferer of insomnia for more than two years and I thought the problem cannot be resolved. This sleep plan …. was suggested to me by a cognitive behavioral therapist. It totally changed the way I think about rest. I`m now sleeping comfortably again every night. My bed is now my friend again. [Ref 5]

I have major depressive disorder and generalized anxiety disorder which has led to crippling insomnia-your videos are the only thing that help me sleep! ….. Thank you for these videos! Love and peace to you! [Ref 14]

I was diagnosed with GAD (General Anxiety Disorder) one of the problems I had with my anxiety was restless nights. This channel has helped me INCREDIBLY. Lack of sleep is killer, and it felt like I was going insane.……back to my normal sleep cycle. I thank this channel so much. I'm truly happy again:) [Ref 20]

Hi. I'm an insomniac mother with a newborn and almost 1 year old. We've all used this video. If they are crying and can't sleep I put this on.! [Ref 14]

Perspective on the Positive Impact of the Sleep-Aiding Music Videos

The effectiveness of listening to sleep-aiding music videos on YouTube was a dominant theme in our analysis. Most of the comments from viewers (2805/4023, 69.7%) suggested listening to the YouTube sleep-aiding music videos had a positive impact on their sleeping problem and sleep quality. Examples of comments in this theme include the following:

I discovered this music after being in the hospital for 4 days with viral meningitis. I got home extremely tired and couldn't fall asleep. After half an hour of no sleep, I tried searching up relaxing music….. I remember instantly falling asleep to this beautiful music. 9 months later I was struggling to get to sleep. I searched up relaxing music ….. I've put it on ever since then and I can't sleep without it…. Good night from Canada [Ref 2]

turned this on after 4 days of insomnia and working my tail off to wear myself out to no avail played less the five minutes of it and I feel completely drained and relaxed I think I may actually be able to sleep now…. Simply amazing how quickly it worked on me. Thank you [Ref 6]

Omg you have no idea how good this works, every night I turn this on and fall asleep in minutes. FULLY RECOMMEND [Ref 20]

The Perspective of No Impact of the Sleep-Aiding Music Videos

On the contrary, not all viewers had a pleasant experience listening to YouTube sleep-aiding music videos. Some comments (437/4023, 10.9%) indicate that the music videos had no impact on their sleep problems or sleep quality.

I have a problem with sleeping, and this song doesn't work to make me sleep even when i am doing nothing but it can make me feel better,,, ….i have my own way to make me feel asleep although not always successful,…..because the people's problem is so heterogeneous and sometimes it isn't solved immediately. [Ref 1]

Wide awake from CA. Guess soothing music doesn't work on insomnia, but worth a try, eh? [Ref 12]

if you're reading this, then you are awake, which means this track is not working for you... on to the next one!!! [Ref 4]

I still can't go to sleep with this music on [Ref 18]

Opinion About Time to Initiation of Sleep or Sleep Duration

Another important observation from the content analysis was the time it took viewers to sleep or how long they slept while listening to the sleep-aiding music videos. Information from analyzed comments (785/4023, 19.5%) indicated that time to initiation of sleep was between 5 and 30 minutes, and some viewers reported having 10 or more hours of sleep while listening to the music videos:

I fell asleep the first 8 freaking minutes [Ref 5]

I actually fell asleep with this playing about 10 or so minutes into the video [Ref 18]

I swear normally it takes me like one hour before i can get a shut eye.... so i came up with this and used it within 20 minutes i was in REM sleep... [Ref 5]

As a person with insomnia, it's hard to fall asleep. Once I found this, I fell asleep within five minutes!!! Now that I have this, I went from only having 3-6 hours of sleep to 10-15 hours of sleep. The best part is I feel so refreshed in the morning!!! [Ref 1]

Variation of the Location of Viewers

Viewers frequently reported their place of viewing in some comments. YouTube is a widely accessible social media platform used worldwide; however, we were unable to get coordinate information to verify the location of users. Nevertheless, a word cloud of the comments showed that this was a significant theme, and 30.7% (1236/4023) of the comments expressed the area of viewers. The wide variation in the places mentioned in comments suggests that sleep-aiding music videos on YouTube are extensively used. For instance, the following comments describe users' locations:

Goodnight from France [Ref 2]

From US! Thank you for this amazing video! I've tried several different sounds to fall asleep as well as different methods.….. [Ref 8]

I have anxiety, and I always go to sleep around 3 in the morning …..Normally I only get 4 or less hours of sleep …..Goodnight from Canada :) [Ref 9]

Good morning from Italy! My family listens to it before to go to bed, and me, before arriving to work….Thank you! [Ref 17]

Good nite from Philippines [Ref 19]

Goodnight from Africa, thank you! [Ref 12]

Discussion

Principal Findings

Our study shows that YouTube is a widely used social medium where viewers with sleep issues access sleep-aiding music videos to improve their sleep quality and habits. Dominant themes in the content analysis suggest that most of the users have a problem with sleeping such as insomnia; listening to the sleep-aiding music videos helps users to sleep within a short duration; and users have a more extended sleep period while listening to the sleep-aiding music videos. Further, the content analysis of comments revealed a wide variation in the location of viewers of sleep-aiding music videos published on YouTube.

This study provides evidence that suggests a growing number of people are accessing sleep-aiding music on YouTube. A body of literature has reported ever-increasing numbers of people suffering from insufficient sleep, as well as the effect of sleep deprivation on quality of life and productivity [29-31]. In the United States, a study found that there are between 50 and 70 million people with perceived chronic sleep or wakefulness issues. These numbers have increased over time, and over 35% of adults report having insufficient sleep [29,30]. Additionally, research on the effect of problems (both physical and psychosocial) related to sleeping difficulty and the treatment of sleep disorders has continued to gain attention in the past decade [32]. Findings from this study suggest that listening to sleep-aiding music videos on YouTube could be one of the approaches employed by people having some form of sleeping issues, both in the United States and around the world. A possible explanation for the increasing numbers of viewers using sleeping music videos could be that the rising cost of obtaining health care services and increasing difficulty in accessing health care is driving many individuals to less costly and more accessible alternatives to help with their sleep-related issues [33-35].

Recent studies have focused on health information acquisition from the internet and social media platforms, and aggregated data from these sources can provide useful public health information. For example, analyzing Twitter data was one of the approaches employed by the CDC to generate surveillance data during an influenza outbreak in the United States [32-34]. Results from this study suggest that over the 6-year study period, there was a substantial increase in the use of YouTube sleep-aiding music, and this could reflect a rising number of people who may have a sleeping disorder. Furthermore, our observations indicate that YouTube is a social support system where individuals with similar health or life experiences share information on important health topics impacting their health quality of life. Sleep quality is an essential component in the measures of quality of life.

There is a dearth of research examining the effect of frequent use of YouTube on sleep patterns. Therefore, we were unable to compare our findings with any similar research. Our study focused on the perceptions of viewers regarding the effectiveness of using YouTube sleep-aiding music videos to improve their sleep pattern. The results from our study show that most viewers perceive a positive effect of YouTube sleep videos on their sleep quality and duration. Nevertheless, it has been shown by several studies that the frequent use of social media adversely affects sleep quality and pattern among users [36-39]. A study conducted among Canadian youths showed the use of social media for at least 1 hour per day was associated with higher odds of short sleep duration in a dose-response manner [40]. A study by Levenson and colleagues [41] examined the independent association of social media use 30 minutes before bed and disturbed sleep. Their study found a significant linear trend in the odds ratios between the frequency of checking social media in the 30 minutes before bed and increased sleep disturbance. Even though our study result is not directly comparable with these reports, the comments obtained for our content analysis established that individuals with sleep disturbances use sleep-aiding music videos on YouTube to enhance their sleep habits and maintain adequate and prolonged sleep. This topic requires further exploration to understand how frequent use of YouTube could impact the sleep quality of users.

The capability of using social media to monitor public health issues by location is well documented [42-45]. For instance, Twitter provides location information for some tweets, aiding geographical biosurveillance of emerging health topics. During the recent Zika outbreak, communities worldwide discussed the disease and critical issues associated with it on Twitter. Data collected from tweets reflected the spread of interest in Zika from its original hotspot in South America to North America and then across the globe [43]. Unlike Twitter, YouTube does not provide the location coordinates of users. Still, users of YouTube could freely offer information about their location in comments. Data from our content analysis of comments include the varied geographic location of users of sleep-aiding music videos on YouTube. Observations from analyzed comments indicate that the users were located both in developed and developing countries. This finding also suggests that difficulty in sleeping is a global issue and supports the notion that social networking platforms can provide an understanding of the burden of health problems through spatial information from comments and the platforms' geocoding systems.

This study provided valuable information on the pattern of use of sleep-aiding music videos on YouTube; the findings could indicate a significant health burden exists worldwide. However, this study had some limitations. First, the comments used for the content analysis were self-reported and we are unable to verify claims of sleeping problems such as insomnia. Second, our units of analysis for this study were the number of views and comments posted, and we were unable to assess individual-level data such as sociodemographic variables to compare differences between groups. In addition, YouTube users could have multiple accounts with privileges such as posting the same videos with different titles and commenting under different usernames. To minimize this issue in our analysis, we sorted and excluded videos with similar content, and any comments with the exact same wording that appeared multiple times across videos. Third, due to the volume of sleep-aiding music videos on YouTube and the large amount of comments posted, we used the targeted sampling approach to select samples from the most viewed videos. This approach could have eliminated some vital music videos, possibly

resulting in our missing other essential themes. This limitation presents an opportunity for further research on this topic using a similar data source. Finally, the contents of YouTube videos are accessible to the public, and no formal approval is required to access the contents of the postings. Notwithstanding, users must register to upload videos or post comments on the site. Although the usernames in the comment section are publicly available, users are free to create multiple accounts and hold numerous pseudonyms. Therefore, we cannot guarantee that all the comments we extracted were genuine with no astroturfing.

Conclusions

This study provides information regarding the use of YouTube sleep-aiding music videos to ameliorate sleep problems and improve sleep quality. We observed a substantial increase in the number of people using sleep-aiding music videos, and a wide variation in their location. This study also supports the positive impact of listening to soothing music on sleep habits. Finally, this study demonstrates that YouTube is an essential social medium for acquiring and analyzing crucial public health information.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Word cloud of the most frequent terms used in the music video titles.
[PNG File , 181 KB - jmir_v22i8e15697_app1.png]

Multimedia Appendix 2
Word cloud of the most frequent words used in comments. The larger words represent more frequently used terms, while the smaller words were less frequently used.
[PNG File , 293 KB - jmir_v22i8e15697_app2.png]

Multimedia Appendix 3
Description of videos randomly selected for thematic content analysis.
[DOCX File , 16 KB - jmir_v22i8e15697_app3.docx]

References

1. Consensus Conference Panel, Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, et al. Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society on the Recommended Amount of Sleep for a Healthy Adult: Methodology and Discussion. J Clin Sleep Med 2015 Aug 15;11(8):931-952 [FREE Full text] [doi: 10.5664/jcsm.4950] [Medline: 26235159]
2. Ford E, Cunningham T, Croft J. Trends in Self-Reported Sleep Duration among US Adults from 1985 to 2012. Sleep 2015 May 01;38(5):829-832 [FREE Full text] [doi: 10.5665/sleep.4684] [Medline: 25669182]
3. Goel N, Rao H, Durmer J, Dinges D. Neurocognitive consequences of sleep deprivation. Semin Neurol 2009 Sep 9;29(4):320-339 [FREE Full text] [doi: 10.1055/s-0029-1237117] [Medline: 19742409]
4. Morin CM, LeBlanc M, Daley M, Gregoire JP, Mérette C. Epidemiology of insomnia: prevalence, self-help treatments, consultations, and determinants of help-seeking behaviors. Sleep Med 2006 Mar 01;7(2):123-130 [FREE Full text] [doi: 10.1016/j.sleep.2005.08.008] [Medline: 16459140]
5. Doghramji K. The epidemiology and diagnosis of insomnia. Am J Manag Care 2006 May;12(8 Suppl):S214-S220 [FREE Full text] [Medline: 16686591]
6. Rosekind M, Gregory K. Insomnia risks and costs: health, safety, and quality of life. Am J Manag Care 2010 Aug;16(8):617-626 [FREE Full text] [Medline: 20712395]
7. Canivet C, Nilsson PM, Lindeberg SI, Karasek R, Östergren P. Insomnia increases risk for cardiovascular events in women and in men with low socioeconomic status: A longitudinal, register-based study. Journal of Psychosomatic Research 2014 Apr 01;76(4):292-299. [doi: 10.1016/j.jpsychores.2014.02.001]
8. Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord 2011 Dec 01;135(1-3):10-19. [doi: 10.1016/j.jad.2011.01.011] [Medline: 21300408]
9. Alkadhi K, Zagaar M, Alhaider I, Salim S, Aleisa A. Neurobiological Consequences of Sleep Deprivation. CN 2013 Apr 01;11(3):231-249. [doi: 10.2174/1570159x11311030001]
10. Bonnet M, Arand D. Chronic insomnia. Sleep Disorders: Diagnosis and Therapeutics 2008 Jul 23:100-109. [doi: 10.3109/9780203091715-9]
11. Wu M, Lin H, Weng S, Ho C, Wang J, Hsu Y. Insomnia Subtypes and the Subsequent Risks of Stroke. Stroke 2014 May;45(5):1349-1354. [doi: 10.1161/strokeaha.113.003675]

12. Pigeon W. Insomnia as a predictor of depression: do insomnia subtypes matter? Sleep 2010 Dec 01;33(12):1585-1586 [FREE Full text] [doi: 10.1093/sleep/33.12.1585] [Medline: 21120120]
13. Morin C. Insomnia: Prevalence, Burden, and Consequences. Insomnia Rounds 2012;01(01):01-06. [doi: 10.1037/e325732004-011]
14. Ringdahl EN, Pereira SL, Delzell JE. Treatment of primary insomnia. J Am Board Fam Pract 2004 May 01;17(3):212-219 [FREE Full text] [doi: 10.3122/jabfm.17.3.212] [Medline: 15226287]
15. Wang C, Sun Y, Zang H. Music therapy improves sleep quality in acute and chronic sleep disorders: a meta-analysis of 10 randomized studies. Int J Nurs Stud 2014 Jan 01;51(1):51-62. [doi: 10.1016/j.ijnurstu.2013.03.008] [Medline: 23582682]
16. Chang E, Lai H, Chen P, Hsieh Y, Lee L. The effects of music on the sleep quality of adults with chronic insomnia using evidence from polysomnographic and self-reported analysis: a randomized control trial. Int J Nurs Stud 2012 Aug 01;49(8):921-930. [doi: 10.1016/j.ijnurstu.2012.02.019] [Medline: 22494532]
17. Wang Q, Chair SY, Wong EML, Li X. The Effects of Music Intervention on Sleep Quality in Community-Dwelling Elderly. J Altern Complement Med 2016 Jul 01;22(7):576-584. [doi: 10.1089/acm.2015.0304] [Medline: 27223689]
18. Shum A, Taylor BJ, Thayala J, Chan MF. The effects of sedative music on sleep quality of older community-dwelling adults in Singapore. Complement Ther Med 2014 Feb 01;22(1):49-56. [doi: 10.1016/j.ctim.2013.11.003] [Medline: 24559816]
19. Khan ML. Social media engagement: What motivates user participation and consumption on YouTube? Computers in Human Behavior 2017 Jan 01;66:236-247. [doi: 10.1016/j.chb.2016.09.024]
20. Siersdorfer S, Chelaru S, Nejdl W, San PJ. How useful are your comments? Analyzing and predicting YouTube comments and comment ratings. 2010 Apr 26 Presented at: Proceedings of the 19th International Conference on World Wide Web; 2010; Raleigh, NC, USA p. 891-900. [doi: 10.1145/1772690.1772781]
21. Yasmina D, Hajar M, Hassan AM. Using YouTube Comments for Text-based Emotion Recognition. Procedia Computer Science 2016;83:292-299. [doi: 10.1016/j.procs.2016.04.128]
22. Dynel M. Participation framework underlying YouTube interaction. Journal of Pragmatics 2014 Nov;73:37-52. [doi: 10.1016/j.pragma.2014.04.001]
23. Loechner J. YouTube Serves Up A Billion Hours Of Video A Day. Research Brief. 2017 May 11. URL: https://www.mediapost.com/publications/article/300622/youtube-serves-up-a-billion-hours-of-video-a-day.html [accessed 2017-05-11]
24. Klostermann P. YouTube Comment Scraper project. URL: https://github.com/philbot9/youtube-comment-scraper [accessed 2017-03-08]
25. Watters JK, Biernacki P. Targeted Sampling: Options for the Study of Hidden Populations. Social Problems 1989 Oct;36(4):416-430. [doi: 10.1525/sp.1989.36.4.03a00070]
26. Hussin M, Frazier S, Thompson JK. Fat stigmatization on YouTube: a content analysis. Body Image 2011 Jan;8(1):90-92. [doi: 10.1016/j.bodyim.2010.10.003] [Medline: 21126932]
27. NVivo version 12 plus for Windows. QSR International. URL: https://www.qsrinternational.com/nvivo/nvivo-products/nvivo-12-plus [accessed 2020-01-05]
28. Burla L, Knierim B, Barth J, Liewald K, Duetz M, Abel T. From Text to Codings. Nursing Research 2008 Feb 29;57(2):113-117. [doi: 10.1097/01.nnr.0000313482.33917.7d]
29. Alami YZ, Ghanim BT, Zyoud SH. Epworth sleepiness scale in medical residents: quality of sleep and its relationship to quality of life. J Occup Med Toxicol 2018 Jul 13;13(1):21-00 [FREE Full text] [doi: 10.1186/s12995-018-0203-z] [Medline: 30008792]
30. Luik A, Emsley R, Kyle S, Gordon C, Drake C, Marsden A. The impact of treating insomnia upon functional health, psychological wellbeing and sleep-related quality of life: a randomised controlled trial with mediation analysis. J Sleep Res 2018;27:2-3.
31. Groeger JA, Zijlstra FRH, Dijk D. Sleep quantity, sleep difficulties and their perceived consequences in a representative sample of some 2000 British adults. J Sleep Res 2004 Dec;13(4):359-371 [FREE Full text] [doi: 10.1111/j.1365-2869.2004.00418.x] [Medline: 15560771]
32. Zee P, Pagel JN. Comorbid Conditions Caused by Sleeping Disorders. Med Roundtable Gen Med Ed. 2019 Aug 01. URL: https://themedicalroundtable.com/article/comorbid-conditions-caused-sleeping-disorders [accessed 2020-08-17]
33. Sommers BD, Gawande AA, Baicker K. Health Insurance Coverage and Health - What the Recent Evidence Tells Us. N Engl J Med 2017 Aug 10;377(6):586-593. [doi: 10.1056/nejmsb1706645]
34. Reynolds SA, Ebben MR. The Cost of Insomnia and the Benefit of Increased Access to Evidence-Based Treatment: Cognitive Behavioral Therapy for Insomnia. Sleep Med Clin 2017 Mar;12(1):39-46. [doi: 10.1016/j.jsmc.2016.10.011] [Medline: 28159096]
35. Reynolds SA, Ebben MR. The Cost of Insomnia and the Benefit of Increased Access to Evidence-Based Treatment: Cognitive Behavioral Therapy for Insomnia. Sleep Med Clin 2017 Mar;12(1):39-46. [doi: 10.1016/j.jsmc.2016.10.011] [Medline: 28159096]
36. Monacis L, de Palo V, Griffiths MD, Sinatra M. Social networking addiction, attachment style, and validation of the Italian version of the Bergen Social Media Addiction Scale. Journal of Behavioral Addictions 2017 Jun;6(2):178-186. [doi: 10.1556/2006.6.2017.023]

37. Sahin C. Social Media Addiction Scale-Student Form: The Reliability and Validity Study. Tojet - Turkish Online J Educ Technol 2018 Jan 01;17(01):169-182.
38. Andreassen CS, Billieux J, Griffiths MD, Kuss DJ, Demetrovics Z, Mazzoni E, et al. The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: A large-scale cross-sectional study. Psychology of Addictive Behaviors 2016 Mar;30(2):252-262. [doi: 10.1037/adb0000160]
39. Aksoy M. A Qualitative Study on the Reasons for Social Media Addiction. European J Ed Res 2018 Oct 15;7(4):861-865. [doi: 10.12973/eu-jer.7.4.861]
40. Sampasa-Kanyinga H, Hamilton HA, Chaput J. Use of social media is associated with short sleep duration in a dose-response manner in students aged 11 to 20 years. Acta Paediatr 2018 Apr 24;107(4):694-700. [doi: 10.1111/apa.14210] [Medline: 29363166]
41. Levenson J, Shensa A, Sidani J, Colditz J, Primack B. Social Media Use Before Bed and Sleep Disturbance Among Young Adults in the United States: A Nationally Representative Study. Sleep 2017 Sep 01;40(9):zsx113. [doi: 10.1093/sleep/zsx113] [Medline: 28934521]
42. Villanti AC, Johnson AL, Ilakkuvan V, Jacobs MA, Graham AL, Rath JM. Social Media Use and Access to Digital Technology in US Young Adults in 2016. J Med Internet Res 2017 Jun 07;19(6):e196 [FREE Full text] [doi: 10.2196/jmir.7303] [Medline: 28592394]
43. Stefanidis A, Vraga E, Lamprianidis G, Radzikowski J, Delamater PL, Jacobsen KH, et al. Zika in Twitter: Temporal Variations of Locations, Actors, and Concepts. JMIR Public Health Surveill 2017 Apr 20;3(2):e22 [FREE Full text] [doi: 10.2196/publichealth.6925] [Medline: 28428164]
44. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec 17;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]
45. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. Vaccine 2017 May 25;35(23):3033-3040 [FREE Full text] [doi: 10.1016/j.vaccine.2017.04.060] [Medline: 28461067]

Abbreviations

API: application programming interface
CDC: Centers for Disease Control and Prevention

Edited by G Eysenbach; submitted 30.07.19; peer-reviewed by M Wick, C Ulmer, K McCausland; comments to author 04.11.19; revised version received 30.12.19; accepted 13.05.20; published 24.08.20.

Please cite as:
Eke R, Li T, Bond K, Ho A, Graves L
Viewing Trends and Users' Perceptions of the Effect of Sleep-Aiding Music on YouTube: Quantification and Thematic Content Analysis
J Med Internet Res 2020;22(8):e15697
URL: http://www.jmir.org/2020/8/e15697/
doi:10.2196/15697
PMID:32831182

Original Paper

Associations of Internet Addiction Severity With Psychopathology, Serious Mental Illness, and Suicidality: Large-Sample Cross-Sectional Study

Wanjun Guo[1*], MD, PhD; Yujie Tao[1*], MSc; Xiaojing Li[1], PhD; Xia Lin[2,3], MSc; Yajing Meng[1], PhD; Xia Yang[1], BSc; Huiyao Wang[1], MSc; Yamin Zhang[1], PhD; Wanjie Tang[4], PhD; Qiang Wang[1], MD, PhD; Wei Deng[1], MD, PhD; Liansheng Zhao[1], BSc; Xiaohong Ma[1], MD, PhD; Mingli Li[1], PhD; Ting Chen[1], MSc; Jiajun Xu[1], PhD; Jing Li[1], MD, PhD; Wei Hao[5], PhD; Sing Lee[6], FRCS; Jeremy W Coid[1], PhD; Andrew J Greenshaw[7], PhD, FRSA; Tao Li[1], MD, PhD

[1]Mental Health Center and Psychiatric Laboratory, State Key Laboratory of Biotherapy, West China Hospital of Sichuan University, Chengdu, Sichuan, China

[2]Institute of Emergency Management and Post-Disaster Reconstruction, Sichuan University, Chengdu, China

[3]North Sichuan Medical College (NSMC), Affiliated Hospital of North Sichuan Medical College, Nanchong City, China

[4]Centre for Educational and Health Psychology, Sichuan University, Chengdu, China

[5]Mental Health Institute, The Second Xiangya Hospital, Central South University, Hunan, Changsha, China

[6]Department of Psychiatry, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[7]Department of Psychiatry, University of Alberta, Edmonton, AB, Canada

[*]these authors contributed equally

Corresponding Author:
Tao Li, MD, PhD
Mental Health Center and Psychiatric Laboratory
State Key Laboratory of Biotherapy
West China Hospital of Sichuan University
28 Dianxin South Street
Chengdu, Sichuan
China
Phone: 86 18980601817
Email: xuntao26@hotmail.com

Abstract

Background: Internet addiction has become a major global concern and a burden on mental health. However, there is a lack of consensus on its link to mental health outcomes.

Objective: The aim of this study was to investigate the associations between internet addiction severity and adverse mental health outcomes.

Methods: First-year undergraduates enrolled at Sichuan University during September 2015, 2016, 2017, and 2018 were invited to participate in the current study survey, 85.13% (31,659/37,187) of whom fully responded. Young's 20-item Internet Addiction Test, Patient Health Questionnaire-15, Patient Health Questionnaire-9, Symptom Checklist 90, Six-Item Kessler Psychological Distress Scale, and Suicidal Behaviors Questionnaire-Revised were used to evaluate internet addiction, four psychopathologies (high somatic symptom severity, clinically significant depression, psychoticism, and paranoia), serious mental illness, and lifetime suicidality.

Results: The prevalence of students with mild, moderate, and severe internet addiction was 37.93% (12,009/31,659), 6.33% (2003/31,659), and 0.20% (63/31,659), respectively. The prevalence rates of high somatic symptom severity, clinically significant depression, psychoticism, paranoid ideation, and serious mental illness were 6.54% (2072/31,659), 4.09% (1294/31,659), 0.51% (160/31,659), 0.52% (165/31,659), and 1.88% (594/31,659), respectively, and the lifetime prevalence rates of suicidal ideation, suicidal plan, and suicidal attempt were 36.31% (11,495/31,659), 5.13% (1624/31,659), and 1.00% (315/31,659), respectively. The prevalence rates and odds ratios (ORs) of the four psychopathologies and their comorbidities, screened serious mental illness, and suicidalities in the group without internet addiction were much lower than the average levels of the surveyed population.

Most of these metrics in the group with mild internet addiction were similar to or slightly higher than the average rates; however, these rates sharply increased in the moderate and severe internet addiction groups. Among the four psychopathologies, clinically significant depression was most strongly associated with internet addiction after adjusting for the confounding effects of demographics and other psychopathologies, and its prevalence increased from 1.01% (178/17,584) in the students with no addiction to 4.85% (582/12,009), 24.81% (497/2,003), and 58.73% (37/63) in the students with mild, moderate, and severe internet addiction, respectively. The proportions of those with any of the four psychopathologies increased from 4.05% (713/17,584) to 11.72% (1408/12,009), 36.89% (739/2003), and 68.25% (43/63); those with lifetime suicidal ideation increased from 24.92% (4382/17,584) to 47.56% (5711/12,009), 67.70% (1356/2003), and 73.02% (46/63); those with a suicidal plan increased from 2.59% (456/17,584) to 6.77% (813/12,009), 16.72% (335/2003), and 31.75% (20/63); and those with a suicidal attempt increased from 0.50% (88/17,584) to 1.23% (148/12,009), 3.54% (71/2003), and 12.70% (8/63), respectively.

Conclusions: Moderate and severe internet addiction were strongly associated with a broad group of adverse mental health outcomes, including somatic symptoms that are the core features of many medical illnesses, although clinically significant depression showed the strongest association. This finding supports the illness validity of moderate and severe internet addiction in contrast to mild internet addiction. These results are important for informing health policymakers and service suppliers from the perspective of resolving the overall human health burden in the current era of "Internet Plus" and artificial intelligence.

(J Med Internet Res 2020;22(8):e17560) doi:10.2196/17560

KEYWORDS

internet; addiction; psychopathology; suicidality; serious mental illness

Introduction

The internet has conceptually transformed the Earth into a high-dimensional information network village, and human experience has benefited from the unprecedented availability and exchange of information. However, the potential adverse effects of internet exposure on human health have emerged as a major global concern [1].

In June 2017, the English addiction therapist Mandy Saligari warned parents that giving a smartphone to their child may be as damaging as giving them "a gram of cocaine" due to its potential addictive features and association with adverse mental consequences [2]. In September 2017, the French government banned students from using mobile phones in the country's primary, junior, and middle schools [3]. In China, the world's largest market for online gaming since 2015 due to its high population size and ubiquitous national internet access [4] (China's 4G network covered 95% of its administrative villages and 99% of its population by the end of April 2018 [5]), overuse of internet services such as gaming and social media, and the association of overuse with adverse mental consequences among youth has received great public attention. News associated with possible internet addiction–associated adverse events (eg, decline in school grades; self-harm, including suicide and accidental death relevant to internet addiction; and "digital detox" boot camps to treat internet addiction in isolated settings) has dramatically increased in recent years, and some Chinese parents and teachers have depicted internet addiction as "electronic opium" or "electronic heroin" [6,7]. Accordingly, in April 2018, the Ministry of Education of China released an urgent notice concerning prevention of primary and middle school students from indulging in the internet [8].

Since Dr. Kimberly Young published a case report in 1996, internet addiction, sometimes called internet dependence or problematic internet use (PIU), has been increasingly conceptualized as "a kind of psychopathological disorder" [9,10]. Studies have found that internet addiction appears to share a phenomenology akin to that of impulse control/addictive disorders [11]. Some studies have even documented possible biomarkers for this condition [12]. The prevalence of internet addiction is high in some young populations and may have increased in recent years, although this prevalence varies regionally around the world [13]. Further research has suggested that internet addiction may be associated with significant functional and psychological impairment, and may be a contributing factor to issues of interparental conflict, family dissatisfaction, recent stressful events, low reward dependence, and low self-esteem, which all may further increase vulnerability to internet addiction [14,15].

After a wide-ranging debate on whether internet addiction primarily reflects the adverse effects of internet contents (such as gaming and social media as opposed to information seeking) or technologies [16-20], some internet addiction–related types have been identified as official mental disorders in recent revisions of international disease classification and diagnostic instruments. The Diagnostic and Statistical Manual of Mental Disorders (DSM)-5 included internet gaming disorder (IGD) in the appendix as a disorder requiring further study [21]. The International Classification of Diseases (ICD)-11 defined and specified gaming disorder as "predominantly online" or "predominantly offline" [22]. Some researchers have argued that classification of IGD or gaming disorder as predominantly online is premature. This is based on several concerns, including low quality/extent of available research, the fact that current operationalization leans too heavily on substance use and may be locked into a confirmatory approach rather than an exploration of the boundaries of normal vs pathological behavior [23], and a lack of consensus on the link of internet addiction or specific internet addiction patterns (including IGD) with other mental health outcomes [19,24]. Some cross-sectional studies found that internet addiction (or PIU) was significantly associated with poor sleep quality [18], attention-deficit/hyperactivity disorder (ADHD) symptoms [25], anxiety [25], depression [25], and even suicidality [26]. In a

longitudinal study, Lau et al [27] found a bidirectional predictive relationship between internet addiction and depression. A higher level of hyperactivity/inattention and self-esteem problems were predictors of IGD, and IGD was a predictor of adolescent emotional distress in the context of further longitudinal work [28]. Some studies suggest that, relative to internet addiction, the association of problematic online gaming might be stronger with hyperactivity/inattention but less strong with other common psychopathologies such as depression [19,20]. However, these association studies suffer from two major limitations: (1) their relatively small sample sizes (from hundreds to approximately 3000 participants) precluded an analysis of the rates and odds ratios (ORs) of linked mental health outcomes based on detailed internet addiction severity rather than dichotomous decisions (ie, either a person has or does not have internet addiction or PIU), and (2) each of these studies investigated only a single or a small number of psychopathologies rather than a series of representative syndromes. This latter limitation has precluded the observation of variance in the association patterns and comorbidity across syndromes.

To overcome these limitations, in the present study, we investigated the prevalence of four representative psychopathologies (high somatic symptom severity, clinically significant depression, psychoticism, and paranoid ideation), which could reflect a broad range of symptoms and different levels of severity of clinical manifestation of mental disorders, as well as serious mental illness and suicidality among groups with different levels of internet addiction severity. The study was conducted using a large Chinese undergraduate sample comprising the largest sample evaluated in internet addiction research to date.

Methods

Study Design and Ethics

The present study, which was approved by the Ethics Committee of West China Hospital, Sichuan University (no. 2016-171), is based on the cross-sectional data from the Online Psychosomatic Health Survey (OPHS) system of Sichuan University. Sichuan University is a national comprehensive university located in southwestern China whose undergraduates come from all of the provincial administrative regions of China. Every respondent was informed of the aim of this investigation prior to the formal survey, and internet-based informed consent was obtained from each participant.

Survey Development

The questions and scales in the OPHS system were designed and selected by a group of researchers from the Mental Health Center, West China Hospital of Sichuan University, and the Centre for Educational and Health Psychology of Sichuan University. This survey was conducted through internet messages with a website link. Students voluntarily logged in with their student ID and were able to quit at any time without penalty. The final survey domains included in the OPHS system were: sociodemographics, childhood adversity, Adolescent Self-Rating Life Events Check List, psychopathologies (including high somatic symptom severity, clinically significant depression, psychoticism, and paranoid ideation), psychological distress, suicidality, and internet usage. In addition to demographics of biological sex and age (since the age for most students that attend college is 18 years old in China, the respondents were categorized into three groups: younger than, equal to, or older than 18 years old), data were collected using the Young 20-item Internet Addiction Test (IAT), Patient Health Questionnaire-15 (PHQ-15), Patient Health Questionnaire-9 (PHQ-9), psychoticism and paranoid ideation subscales of the Symptom Checklist 90 (SCL-90), six-item Kessler psychological distress scale (K6), and Suicidal Behaviors Questionnaire-Revised (SBQ-R) for the specified research aims. The raw data were saved and stored in a manner only accessible by the research administrators.

Participants and Survey Administration

All undergraduate freshmen enrolled in university during September 2015, September 2016, September 2017, and September 2018 were invited to participate. For inclusion in the study, participants were required to finish the entire self-administered questionnaire on the internet by October 31 of their year of enrollment. Altogether, 37,187 freshmen were invited to participate, 34,140 (91.80%) of whom agreed to participate and logged on to the online survey system. Participants in this study were excluded for the following reasons: failure to complete all of the survey questions, taking less than 10.0 minutes to finish the survey (the median time to finish the survey was 30.6 minutes), did not turn in the questionnaire by November, were not between 15 and 23 years old, or provided obvious dishonest information (eg, selected the same severity option for all questions in a separate scale or questionnaire, or a unique ID was logged in by two or more participants). The final sample comprised 31,659 respondents, yielding an overall effective response rate of 85.13%.

Measurements

Young IAT

Young IAT has been internationally validated to assess internet addiction [29,30]. The IAT consists of 20 questions asking about the extent of an individual's involvement with the internet, with each question rated on a scale of 1 to 5, in addition to the rating of "not applicable" as 0. Total scores can range from 0 to 100, with the severity of internet addiction categorized as normal (0-30), mild (31-49), moderate (50-79), and severe (80-100) [9]. Cronbach α for this questionnaire in this study was .90.

PHQ-15

The PHQ-15, based on the diagnostic criteria of depression from DSM-IV, was used to measure high somatic symptom severity of participants in this study. PHQ-15 is a self-administered, internationally validated, and widely used questionnaire for screening somatization and monitoring somatic symptom severity in clinical practice and research. Its test-retest reliability coefficient was found to be 0.75 in a Chinese population [31,32]. Subjects were asked to rate the severity of 13 somatic symptoms as 0 ("not bothered at all"), 1 ("bothered a little"), or 2 ("bothered a lot") during the 4 weeks preceding the study. Two additional physical symptoms—feeling tired or having little energy and trouble sleeping—were coded as 0 ("not at all"), 1 ("several days"), or 2 ("more than half the days" or

"nearly every day"). The range of the scale is 0 to 30 and we used a conventional cutoff of ≥10 to define high somatic symptom severity. Cronbach α for the questionnaire in this study was .79.

PHQ-9

The nine questions of PHQ-9 ask about the participant's experience in the last 2 weeks, with the possible responses ranging from 0 ("not at all") to 3 ("nearly every day") and a score ranging from 0 to 27. A cutoff of ≥10, reported as the optimal cutoff to detect major depressive episodes in the Chinese population, was used to define clinically significant depression in the present study [33]. Cronbach α for the questionnaire in this study was .84.

Psychoticism and Paranoid Ideation Subscales of the SCL-90

Subscales of the SCL-90 were used to assess psychoticism and paranoid ideation in this study. High validity and reliability of SCL-90 has been demonstrated both in the Chinese general population and with university students [34,35]. The psychoticism and paranoid ideation subscales comprise 10 and 6 specific symptoms, respectively. Responses fit on a 5-point scale, from 0 ("not at all") to 4 ("extremely") for both subscales. Each subscale score was calculated as sum score/item numbers. Mean normative scores (1.50, SD 0.51 for psychoticism and 1.63, SD 0.57 for paranoid ideation based on meta-analysis results of studies on Chinese university students) were used to define psychoticism and paranoid ideation [35], yielding subscale score criteria of >2.01 and >2.20 to identify individuals as having psychoticism and paranoid ideation, respectively. Cronbach α of the subscales of psychoticism and paranoid ideation was .83 and .80, respectively, in this study.

K6

The K6 is a self-report 6-item scale that asks respondents to rate how frequently they have felt "nervous," "hopeless," "restless or fidgety," "so depressed that nothing could cheer you up," "that everything was an effort," and "worthless" during the past 30 days. The 5-point response options for each question ranged from 0 (none) to 4 (all of the time) [36]. For the Chinese version, reported validity in both the general population and university students was good for rapidly screening serious mental illness, which has been defined as meeting the criteria for one or more DSM-IV/Composite International Diagnostic Interview (CIDI) mental disorders and resulting in serious impairment [36,37]. The optimal cutoff of 12/13, identified by these studies, was the criterion for serious mental illness in the present study. Cronbach α for the questionnaire in this study was .86.

SBQ-R

In this study, the first SBQ-R question was used to survey lifetime suicidality in participants. This self-report questionnaire consists of 4 questions used to identify suicide risk and has been validated for use with university students. Both single SBQ-R Item 1 and SBQ-R total scores have been recommended for use in clinical and nonclinical settings [38]. Question 1 explores whether the respondent has ever entertained risky thoughts of suicide or attempted to kill himself/herself in his/her lifetime. Response options include "1 (never)," "2 (It was just a brief passing thought)," "3a (I have had a plan at least once to kill myself but did not try to do it)," "3b (I have had a plan at least once to kill myself and really wanted to die)," "4a (I have attempted to kill myself but did not want to die)," and "4b (I have attempted to kill myself, and really hoped to die)." The present study defined scores of ≥2, ≥3, and 4 as suicidal ideation, suicidal plan, and suicidal attempt, respectively.

Statistical Analysis

Data analyses were conducted using SPSS 22.0 software. The mean (95% CI) and median age, number of respondents in each demographic and internet addiction severity group, and prevalence rates (95% CIs) across levels of internet addiction severity, psychopathology, serious mental illness, and suicidality were calculated. The Chi-square test was used to compare prevalence rates among demographic or internet addiction severity groups, and the adjusted ORs and their 95% CIs of other mental health outcomes among the internet addiction severity groups were estimated and compared using binary logistic regression, adjusting for confounding effects of demographics (ie, age, sex, and enrollment year) and the interactions among psychopathologies. A two-tailed alpha level of .05 was used to evaluate statistical significance.

Results

Demographics

The 31,659 respondents included 16,109 (50.88%) men and 15,550 (49.12%) women. The groups aged 15-17 years, 18 years, and 19-23 years included 4821 (15.23%), 19,034 (60.12%), and 7804 (24.65%) respondents, respectively.

Prevalence of Different Severities of Internet Addiction

Overall, the rates of those without internet addiction and those with mild, moderate, and severe internet addiction were 55.5%, 37.9%, 6.3%, and 0.2%, respectively. Internet addiction severity rates differed between sexes, among the age groups, and among the year-of-survey groups. The rate of moderate internet addiction was higher in women than in men. The oldest group (aged 19-23 years) included more respondents without internet addiction and fewer individuals with mild internet addiction than the other two age groups. Among the year-of-survey groups, the rates of all levels of internet addiction severity increased with recency (Table 1).

Table 1. Prevalence (%) and severities of internet addiction in the surveyed population.

Subgroup	Mild internet addiction		Moderate internet addiction		Severe internet addiction		χ^2	df	*P* value
	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI			
Whole sample (N=31,659)	12,009 (37.93)	37.40-38.47	2003 (6.33)	6.06-6.60	63 (0.20)	0.15-0.25			
Sex							36.02	3	<.001
Male (n= 16,109)	5992 (37.20)	36.45-37.94	920 (5.71)	5.35-6.07	29 (0.18)	0.11-0.25			
Female (n= 15,550)	6017 (38.70)	37.93-39.46	1083 (6.96)	6.56-7.36	34 (0.22)	0.15-0.29			
Age (years)							27.51	6	<.001
15-17 (n= 4821)	1877 (38.93)	37.56-40.31	342 (7.10)	6.37-7.82	5 (0.10)	0.01-0.19			
18 (n= 19,034)	7309 (38.40)	37.71-39.09	1198 (6.29)	5.95-6.64	39 (0.20)	0.14-0.27			
19-23 (n= 7804)	2823 (36.17)	35.11-37.24	463 (5.61)	5.09-6.12	19 (0.24)	0.13-0.35			
Year of survey							272.32	9	<.001
2015 (n= 8723)	3167 (36.31)	35.30-37.32	330 (3.78)	3.38-4.18	6 (0.07)	0.01-0.12			
2016 (n= 6501)	2318 (35.66)	34.49-36.82	435 (6.69)	6.08-7.30	16 (0.25)	0.13-0.37			
2017 (n= 8644)	3291 (38.07)	37.05-39.10	593 (6.86)	6.33-7.39	18 (0.21)	0.11-0.30			
2018 (n= 7791)	3233 (41.50)	40.40-42.59	645 (8.28)	7.67-8.89	23 (0.30)	0.17-0.42			

Prevalence of Four Psychopathologies and Their Associations With Internet Addiction Severity

Overall, the prevalence rates of high somatic symptom severity, depression, psychoticism, paranoid ideation, and serious mental illness in the current sample were 6.5%, 4.1%, 0.5%, 0.5%, and 1.9%; lifetime prevalence rates of suicidal ideation, suicidal plan, and suicidal attempt were 36.3%, 5.1%, and 1.0%, respectively. All of these rates significantly increased from the group without internet addiction to the groups with mild, moderate, and severe internet addiction. Compared with the average levels of the surveyed population, these rates in the group without internet addiction were much lower, whereas those in the group with mild internet addiction were similar (psychoticism, paranoid ideation, serious mental illness, and suicidal attempt) or mildly higher (high somatic symptom severity, clinically significant depression, suicidal ideation, and suicidal plan); however, these rates sharply increased in the moderate and severe internet addiction groups (Figure 1). Accordingly, the OR of each psychopathology and serious mental illness adjusted for demographics also significantly increased with internet addiction severity. ORs for each psychopathology in the moderate and severe internet addiction groups (except for that of clinically significant depression, which was most strongly associated with internet addiction) became significantly smaller when further adjusted for other psychopathologies (Table 2). In addition, although the adjusted ORs of suicidality also decreased after further adjusting for the four psychopathologies, they still increased significantly with internet addiction severity, especially for suicidal plans and suicidal attempts (Table 2). These associations were all in the same direction when analyzed for biological sex, although there were differences in the respective ORs (Multimedia Appendix 1 and Multimedia Appendix 2).

Figure 1. Prevalence rates (95% CI) of high somatic symptom severity (HSSS), clinically significant depression (CSD), psychoticism, paranoid ideation, serious mental illness (SMI), suicidal ideation, suicidal plan, and suicidal attempt in the whole sample, and in those without internet addiction (IA) and with mild, moderate, and severe IA.

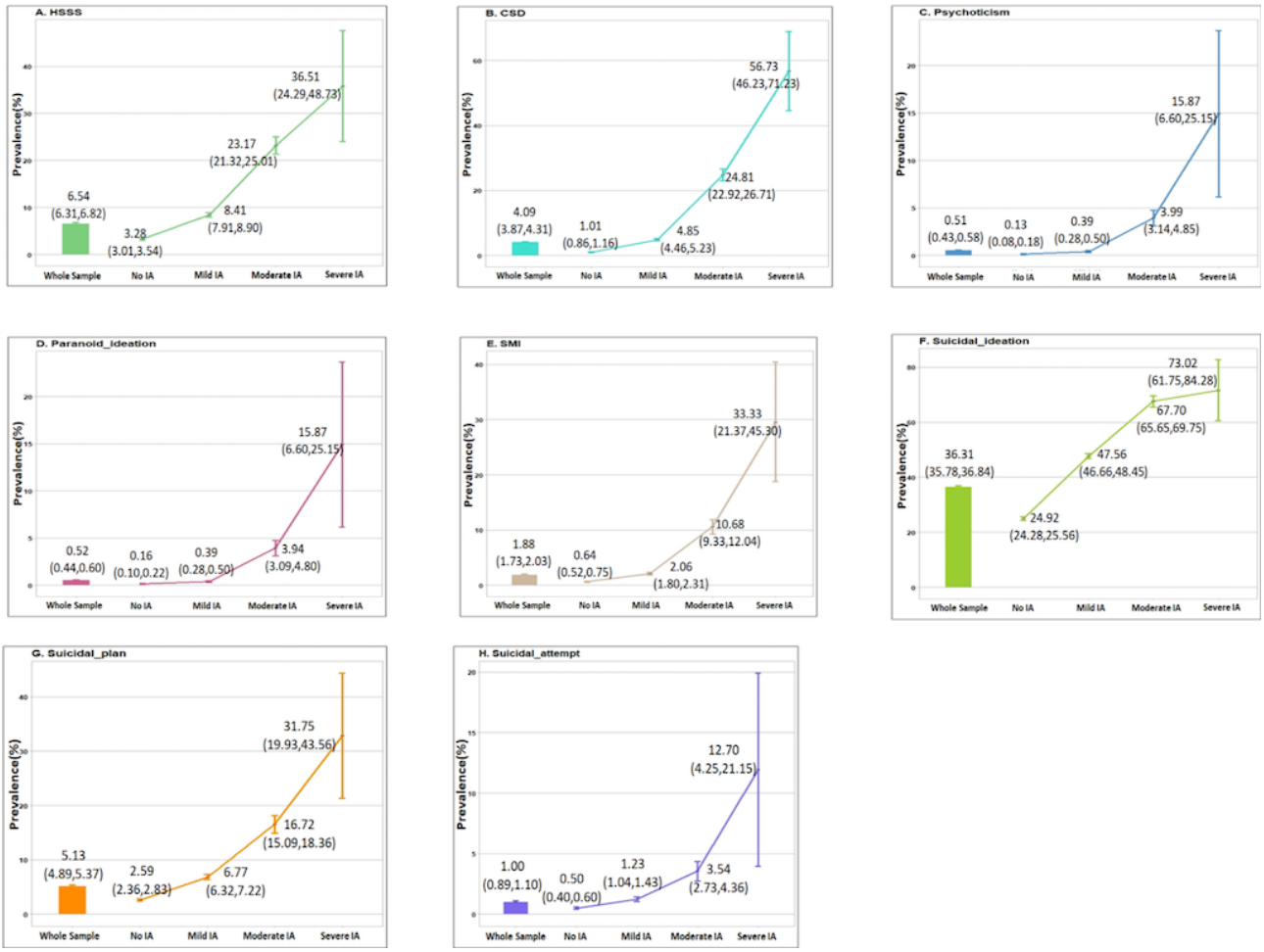


Table 2. Adjusted odds ratios (aORs) of mental health outcomes in the groups with mild, moderate, and severe internet addiction (IA).

Mental health outcome	Model 1[a]		Model 2[b]	
	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
HSSS[c]				
Mild IA	2.61 (2.35-2.90)	<.001	2.30 (2.06-2.56)	<.001
Moderate IA	8.24 (7.20-9.43)	<.001	4.32 (3.71-5.04)	<.001
Severe IA	15.74 (9.27-26.73)	<.001	3.55 (1.90-6.64)	<.001
CSD[d]				
Mild IA	4.86 (4.10-5.76)	<.001	4.17 (3.50-4.96)	<.001
Moderate IA	30.55 (25.51-36.59)	<.001	18.88 (15.59-22.86)	<.001
Severe IA	127.95 (75.57-216.64)	<.001	65.25 (35.72-119.20)	<.001
Psychoticism				
Mild IA	2.91 (1.76-4.79)	<.001	1.92 (1.11-3.36)	.02
Moderate IA	29.41 (18.40-47.00)	<.001	4.73 (2.63-8.50)	<.001
Severe IA	128.71. (58.11-285.08)	<.001	7.12 (2.18-23.30)	.001
Paranoid ideation				
Mild IA	2.29 (1.44-3.65)	<.001	1.46 (0.88-2.42)	.15
Moderate IA	22.71 (14.76-34.96)	<.001	3.39 (1.94-5.95)	<.001
Severe IA	101.76 (47.01-220.28)	<.001	5.16 (1.56-17.11)	.01
SMI[e]				
Mild IA	3.14 (2.51-3.93)	<.001	N/A[f]	N/A
Moderate IA	16.89 (13.34-21.37)	<.001	N/A	N/A
Severe IA	67.97 (38.77-119.16)	<.001	N/A	N/A
Suicidal ideation				
Mild IA	2.69 (2.56-2.83)	<.001	2.51 (2.39-2.64)	<.001
Moderate IA	6.03 (5.45-6.66)	<.001	4.24 (3.81-4.71)	<.001
Severe IA	7.75 (4.43-13.56)	<.001	3.07 (1.66-5.65)	<.001
Suicidal plan				
Mild IA	2.60 (2.31-2.93)	<.001	2.24 (1.99-2.53)	<.001
Moderate IA	6.77 (5.82-7.87)	<.001	3.31 (2.78-3.93)	<.001
Severe IA	15.73 (9.12-27.14)	<.001	3.79 (2.03-7.09)	<.001
Suicidal attempt				
Mild IA	2.35 (1.80-3.06)	<.001	1.98 (1.51-2.60)	<.001
Moderate IA	6.41 (4.66-8.82)	<.001	2.60 (1.79-3.78)	<.001
Severe IA	25.38 (11.66-55.22)	<.001	4.97 (2.07-11.94)	<.001

[a]Binary logistic regression adjusted for age, gender, and year of survey; no internet addiction group as reference (1).

[b]Binary logistic regression adjusted for ag, gender, year of survey, and psychopathologies; no internet addiction group as reference (1).

[e]HSSS: high somatic symptom severity, defined according to the total score of the Patient Health Questionnaire-15 using a cutoff of ≥10.

[d]CSD: clinically significant depression, defined according to the total score of the Patient Health Questionnaire-9 using a cutoff of ≥10.

[e]SMI: serious mental illness, defined according to the total score of the 6-item Kessler psychological distress scale using a cutoff of ≥13.

[f]N/A: not applicable.

Rates of Comorbidity with the Four Psychopathologies and Their Associations with Internet Addiction Severity

Among all respondents, the rates of those who had at least one, two, and three of the four psychopathologies were 9.2%, 1.9%, and 0.4%, respectively. These rates increased significantly from the group without internet addiction to those with mild, moderate, and severe internet addiction (Table 3). Accordingly, the ORs based on the number of comorbid conditions adjusted for demographics also significantly increased with internet addiction severity (Table 3).

Table 3. Adjusted odds ratios (aORs) of the prevalence (%) of comorbidities of four psychopathologies[a] in the groups with mild, moderate, and severe internet addiction (IA).

Number of comorbidities		Prevalence (95%CI)	aOR[b] (95%CI)	*P* value
≥1				
	No IA	(3.76-4.35)	1	
	Mild IA	11.73 (11.15-12.30)	3.05 (2.78-3.35)	<.001
	Moderate IA	36.89 (34.78-39.01)	13.04 (11.57-14.69)	<.001
	Severe IA	68.25 (56.44-80.07)	48.00 (27.89-82.62)	<.001
≥2				
	No IA	0.41 (0.32-0.50)	1	
	Mild IA	2.02 (1.76-2.27)	4.84 (3.72-6.31)	<.001
	Moderate IA	13.98 (12.46-15.50)	36.68 (28.14-47.82)	<.001
	Severe IA	36.51 (24.29-48.73)	127.94 (72.51-225.73)	<.001
≥3				
	No IA	0.09 (0.04-0.13)	1	
	Mild IA	0.26 (0.17-0.35)	2.96 (1.60-5.49)	.001
	Moderate IA	3.74 (2.91-4.58)	42.84 (24.48-74.96)	<.001
	Severe IA	15.87 (6.60-25.15)	202.08 (86.50-472.09)	<.001

[a]The four psychopathologies are: high somatic symptom severity, clinically significant depression, psychoticism, and paranoid ideation.

[b]aOR: Adjusted odds ratio based on binary logistic regression analysis controlling for age, gender, and year of survey groups.

Discussion

Principal Findings and Research Priorities

The IAT, which is the most widely used internet addiction survey instrument in the world, allows researchers to differentiate the severity of internet addiction, as in the present study [10]. However, and likely a consequence of relatively small sample sizes, most previous studies used only the IAT to identify PIU using a total score cutoff at ≥50, which includes moderate to severe internet addiction; this issue clearly applies to prior research on the association of internet addiction with other mental health outcomes. The present study, which to our knowledge included the largest ever sample size in internet addiction research to date, is the first to use a methodology that enabled a comparison of prevalence rates and ORs of a series of representative psychopathologies, serious mental illness, and suicidality among groups without internet addiction, and with mild, moderate, and severe internet addiction. Consequently, this study has three prominent strengths in comparison with previous studies on associations between internet addiction and other mental health outcomes.

First, by comparing the prevalence rates and ORs of mental health outcomes among the differentiated internet addiction severity groups, the present study found a surprising incremental pattern of these conditions from the group without internet addiction through the group with severe internet addiction. That is, the prevalence rates of the four psychopathologies, serious mental illness, and suicidality in the group without internet addiction were much lower than the average levels of the surveyed population, and those in the group with mild internet addiction were similar to (psychoticism, paranoid ideation, serious mental illness, and suicidal attempt) or mildly higher than (high somatic symptom severity, clinically significant depression, suicidal ideation, and suicidal plan) the average levels of the surveyed population; however, the prevalence rates and their OR increments from mild to moderate and severe internet addiction were surprisingly large for most conditions. Previous surveys typically compared prevalence rates and ORs of individual syndromes between the groups without and with internet addiction. Based on this approach in the original studies, Ho et al [25] performed a large meta-analysis including eight studies comprising 1641 internet addiction cases and 11,210 controls, and the authors concluded that the relative ORs of ADHD symptoms, alcohol abuse, anxiety, and depression in internet addiction/PIU cases compared to those without internet addiction were respectively 3.05 (95% CI 2.14-4.37), 2.85 (95% CI 2.15-3.77), 2.77 (95% CI 2.04-3.75), and 2.70 (95% CI 1.46-4.97) [25]. However, due to the limited sample sizes in the studies included in the meta-analysis, there may not have

been a sufficiently large number of severe internet addiction cases included in the analysis. Consequently, these studies failed to reveal either the dramatically increasing pattern of other mental health outcomes associated with internet addiction severity or the surprisingly high prevalence rates of other mental health outcomes in those with severe internet addiction. The prevalence rate of severe internet addiction might be lower, as noted in the present study, but is of paramount importance for an understanding of mental health outcomes associated with severe internet addiction, because most of the adverse events with a social impact related to internet exposure appear to be related mainly to such severe cases. By surveying the rates of a group of mental health outcomes in the mild internet addiction group, the present study also provides the first evidence that the rate of mental health problems is similar to or only mildly higher than the average level of the surveyed population and does not support the validity of mild internet addiction as a mental disorder.

Second, by investigating the largest number (four) of representative psychopathologies in the context of an internet addiction survey to date, the present study enabled an analysis of the overlap between these psychopathologies and their comorbidity, demonstrating a clear relationship with internet addiction severity. Previous surveys have typically reported an association of internet addiction with individual psychopathology because of a focus on the investigation of a single or very small number of psychopathologies. This may explain a possible significant underestimation of the prevalence of psychopathologies in groups with varying internet addiction severities, in addition to a lack of prior understanding of the pattern of overlap and comorbidity of the mental health outcomes linked to internet addiction. For example, the meta-analysis by Ho et al [25] showed that the prevalence of ADHD, alcohol abuse, anxiety, and depression among internet addiction/PIU cases was 21.7%, 13.3%, 23.3%, and 26.3%, respectively. However, the authors failed to report the rates and ORs of having at least one of these psychopathologies or having a comorbidity with multiple psychopathologies. Similar to the present study, the highest rate of common psychopathology among those with internet addiction/PIU in the meta-analysis was 26.3% for depression [25], but this rate was much lower than that of individuals with at least one of the four psychopathologies found in the present study (approximately 2/5 in the respondents with moderate internet addiction and more than 2/3 in those with severe internet addiction). Our approach also allowed us to observe the variability in the pattern of associations among the four psychopathology types. For example, clinically significant depression was less prevalent than high somatic symptom severity in the whole population, although the rate of clinically significant depression in the groups with moderate to severe internet addiction was the highest among the four psychopathologies. The rates of psychoticism and paranoid ideation were very low in the overall population but were substantial in the severe internet addiction cases and resulted in an approximately 100-fold increase in OR relative to the group without internet addiction after adjusting for demographics. When investigating the independent associations of internet addiction with other mental health outcomes by controlling for confounding effects of demographics and the interactions among psychopathologies, the present study provides empirical evidence for the previous presumption that depression might be the most strongly internet addiction–associated psychopathology [25], although other psychopathologies also remain independently associated with internet addiction. Another important finding from the present analyses that considered the confounding effect with psychopathologies was that the types of suicidality were also independently associated with internet addiction even after adjusting for both demographics and the presence of psychopathology. Cheng et al [26] documented that internet addiction (compared with no internet addiction) was independently associated with suicidal ideation and attempts in a meta-analysis, but they had not compared the associations among groups with distinct levels of internet addiction severity and had not compared these associations after controlling for the presence of psychopathology apart from depression.

Third, the present study is, to our knowledge, the first to investigate the association of internet addiction with high somatic symptom severity or serious mental illness. Consequently, it was found that internet addiction was strongly associated not only with mental symptoms, serious mental illness, and suicidality but also with somatic symptoms, which are the core features of many medical illnesses [39]. Given that serious mental illness screened by the K6 scale may include most DSM-IV/CIDI-10 mental disorders that result in serious impairment [37], the strong association of internet addiction with serious mental illness found in the present study is an important addition to previous research findings on the association of internet addiction with individual psychopathologies.

The overall prevalence rates of internet addiction (6.5%, including moderate and severe cases), psychopathologies (eg, 6.5% for high somatic symptom severity, 4.1% for clinically significant depression), and serious mental illness (1.9%) in the present sample of university freshmen were lower than those reported in other recent surveys (eg, 10.4% for PIU in Anhui, China [40]; 9.3% for high somatic symptom severity and 4.2% for clinically significant depression in general community populations reported by Kocalevent et al [31] and Yu et al [33], respectively; and 4.0% for serious mental illness in the whole undergraduate population at the same university [37]). The reasons for these differences may be complex but may be at least partly explained as follows. On the one hand, Asian people are less willing to express their mental health problems than their Western counterparts, and it has been reported that Chinese people tend to provide lower scores on mental health screening questionnaires than other races/ethnicities [36]. On the other hand, it has been reported that healthy people, physically or psychologically, might achieve more success at school, and they are expected to be more likely to enter national comprehensive universities [36]. The choice of a Sichuan University sample is likely relevant, as the present student sample may have collectively invested more time in preparation for the National College Entrance Examination, resulting in their high scores and matriculating into a leading-level Chinese university, which might be an indication that they had less casual internet time. However, the lifetime prevalence rates for

suicidalities were not consistently lower than those of previous surveys. For example, the rate of suicidal ideation (36.3%) was similar to that found in a Polish adolescent group (31%) but was much higher than that reported for US adolescents (12.1%). The rate of suicidal plans (5.1%) was slightly higher than those in US adolescents (4.0%) and in a Korean adult population (3.3%). The rate of suicidal attempts (1.0%) was also slightly higher than that in a US adult population (0.8%), although it was much lower than that in US adolescents (4.1%) [41-43]. On a technical note, the reasons for the lower rate of suicidal attempts might partly be attributed to a translation problem relevant to cultural differences between Chinese and English-speaking people; although this issue is noteworthy, it has not been studied in previous research. That is, "attempted to kill myself" has usually been translated into "尝试自杀" (ChangShi ZiSha) in Chinese, which means that the "killing" behavior had already been conducted although it had not resulted in death; in contrast, the phrase in English may mean that the behavior of "killing" had not actually occurred.

Most previous studies have documented that either men are more vulnerable to internet addiction than women or that internet addiction is not gender-specific [36,37]. This is inconsistent with the present study, which found that the rates of moderate or severe internet addiction in women (7.2%) were higher than those in men (5.9%). This may be partially due to the increasing availability of more online games with gender equity or a female orientation (such as The Honor of Kings, Onmyoji) in recent years. A recent survey in the United Kingdom also reported that girls were more likely to be bothered by the content on the internet, which may be attributed to the fact that girls are more concerned about this problem than their male counterparts [44]. As the results have been inconsistent, it is too early to make conclusions regarding gender differences related to internet addiction. This study showed that the prevalence of internet addiction among freshmen increased in later years. This might be partly due to the increasing availability and convenience of the internet with the popularity of mobile phones and other mobile internet devices, and might indicate that increasing exposure to the internet environment may lead to a high risk of internet addiction.

Limitations

The present study has several limitations. As a cross-sectional survey, it does not help to inform us of a possible causal relationship between internet addiction and comorbidity with other mental health outcomes. The use of self-report scales, as in most previous studies, is another obstacle to the generation of a clinical diagnosis. It is possible that comparative data for this student population on several annual diagnoses of mental disorders may be useful for supporting our use of survey self-report data in contrast to standard clinical diagnoses, but there are two issues arising in the current context. On the one hand, the number of annual diagnoses of mental disorders at the university clinic is not available for research analysis due to institutional data access policies related to personal information and protection of privacy for individual students. On the other hand, we expect that such data (if available) would be only partially helpful in view of an expected limitation of correspondence of the data of this study and the university clinic data. Disparities between prevalence rates based on mental health clinic presentation–based estimates and community survey–based estimates are expected. Epidemiological studies have clearly documented that health care–seeking rates of individuals with mental disorders in general communities is low, especially in low and middle income countries such as China, which provides partial support for our current survey approach rather than a medical records–based analysis. It has been argued that cultural issues leading to increased stigmatization and alternate support–seeking behavior may underly this pattern of reporting [45,46].

This study investigated internet addiction but we were not able to identify IGD (which is the focus of the most recent revisions of international disease classification and diagnostic instruments) nor did we identify the proportion of the student sample using the internet predominantly to access online games as opposed to other content such as social media and information. Further information on the proportion of different internet activities that participants engaged in will be necessary to achieve any detailed understanding of the relationship of internet use with addiction and other psychopathologies that are a significant concern for the healthy development of our youth in the information age. To gain insight into the etiology and possible therapeutic interventions for internet addiction and IGD, further research must include data on the proportional access of participants to online gaming, social media, and social messaging as well as access to other risk-related content such as pornography.

We investigated a broader group of mental health outcomes as possibly being associated with internet addiction than previous studies, but did not include all mental health factors. Our participants were first-year students at a single university, although they were from all of the provincial administrative regions of China. The gross enrollment rate in higher education in China from 2015 to 2018 ranged from 40% to 48.1% [47]. Healthy people, physically or psychologically, with better socioecological status typically have higher school achievements, and are more likely to enter national comprehensive universities [48,49]. Given this demographic profile of our sample, we must acknowledge that our findings may not apply generally to the overall Chinese population for this age range. Although the sample size of this study is very large, the small proportion of participants with severe internet addiction (n=63) hindered further analyses to control for the confounding effects of more detailed covariates such as home provinces (n>30) of participants. Future research with larger representative sample sizes and those adopting a longitudinal design approach will be advantageous. The inclusion of a structured clinical diagnostic interview for detailed internet addiction patterns and mental disorders, along with measurement of other relevant markers (such as genetic, neuroimaging, and treatment response) will further shed light on how to treat internet addiction from a perspective of effectively improving human health in the era of "Internet Plus" and artificial intelligence.

Conclusions

The present study showed that internet addiction is strongly associated with a broad group of mental health problems (among

which clinically significant depression was the most strongly associated), which strongly supports the illness validity of moderate and severe, but not mild, internet addiction. This study is also meaningful for health policy makers and service suppliers to manage the so-called internet addiction or other mental health outcomes relevant to internet usage, especially from the perspective of resolving the overall human health burden in the current era of “Internet Plus” and artificial intelligence.

Acknowledgments

We express our appreciation to Sichuan University and all of the students who participated in this study for their cooperation and assistance.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Adjusted odds ratios (ORs) of the four psychopathologies and serious mental illness (SMI) of male and female respondents in the groups with mild, moderate, and severe internet addiction (IA).
[DOCX File , 20 KB - jmir_v22i8e17560_app1.docx]

Multimedia Appendix 2
Adjusted odds ratios (ORs) of suicidal ideation, suicidal plan, and suicidal attempt in the groups with mild, moderate, and severe internet addiction (IA) among males and females.
[DOCX File , 16 KB - jmir_v22i8e17560_app2.docx]

References

1. Saunders JB, Hao W, Long J, King DL, Mann K, Fauth-Bühler M, et al. Gaming disorder: Its delineation as an important condition for diagnosis, management, and prevention. J Behav Addict 2017 Sep 01;6(3):271-279 [FREE Full text] [doi: 10.1556/2006.6.2017.039] [Medline: 28816494]
2. Pells R. 'Giving your kid a smartphone like giving them a gram of cocaine': Expert. The Chronicle. URL: https://www.thechronicle.com.au/news/giving-you-kid-smartphone-giving-them-gram-cocaine/3211542/ [accessed 2017-06-06]
3. Willsher K. France to ban mobile phones in schools from September. The Guardian. 2017 Dec 11. URL: https://www.theguardian.com/world/2017/dec/11/france-to-ban-mobile-phones-in-schools-from-september [accessed 2017-10-16]
4. Research Report on Market Prospects and Investment Opportunities in China's Online Game Industry, 2019-2024. Chinabgao. URL: http://www.chinabgao.com/report/print4726461.html [accessed 2019-04-06]
5. Press conference on developments in the industrial communications industry, three quarters of 2018. Ministry of Industry and Information Technology of the People's Republic of China. URL: http://www.miit.gov.cn/Searchweb/news.jsp [accessed 2019-10-23]
6. Dubois M. Log in, Sign out. The World of Chinese. URL: http://www.theworldofchinese.com/2013/03/log-in-sign-out/ [accessed 2017-08-06]
7. Phillips T. 'Electronic heroin': China's boot camps get tough on internet addicts. The Guardian. 2017 Aug 28. URL: https://www.theguardian.com/world/2017/aug/28/electronic-heroin-china-boot-camps-internet-addicts [accessed 2018-02-08]
8. Urgent notice on Education and guiding work for preventing Primary and Middle School students from indulging in Internet. The Ministry of Education of People's Republic of China. URL: http://www.moe.gov.cn/srcsite/A06/s3325/201804/t20180424_334106.html?from=groupmessage&isappinstalled=0 [accessed 2018-04-20]
9. Young K, Abreu C. Internet addiction: A handbook and guide to evaluation and treatment. Hoboken, NJ: John Wiley & Sons Inc; 2011.
10. Vigna-Taglianti F, Brambilla R, Priotto B, Angelino R, Cuomo G, Diecidue R. Problematic internet use among high school students: Prevalence, associated factors and gender differences. Psychiatry Res 2017 Nov;257:163-171. [doi: 10.1016/j.psychres.2017.07.039] [Medline: 28759791]
11. Niu G, Sun X, Subrahmanyam K, Kong F, Tian Y, Zhou Z. Cue-induced craving for Internet among Internet addicts. Addict Behav 2016 Nov;62:1-5. [doi: 10.1016/j.addbeh.2016.06.012] [Medline: 27305097]
12. Meng Y, Deng W, Wang H, Guo W, Li T. The prefrontal dysfunction in individuals with Internet gaming disorder: a meta-analysis of functional magnetic resonance imaging studies. Addict Biol 2015 Jul;20(4):799-808. [doi: 10.1111/adb.12154] [Medline: 24889021]
13. Block JJ. Issues for DSM-V: internet addiction. Am J Psychiatry 2008 Mar;165(3):306-307. [doi: 10.1176/appi.ajp.2007.07101556] [Medline: 18316427]
14. Ahn J. Korean policy on treatment and rehabilitation for adolescents' internet addiction. 2007 Presented at: International Symposium on the Counseling and Treatment of Youth Internet Addiction; 2007; Seoul, Korea.

15. Firth J, Torous J, Stubbs B, Firth JA, Steiner GZ, Smith L, et al. The "online brain": how the Internet may be changing our cognition. World Psychiatry 2019 Jun;18(2):119-129. [doi: 10.1002/wps.20617] [Medline: 31059635]
16. Aarseth E, Bean AM, Boonen H, Colder Carras M, Coulson M, Das D, et al. Scholars' open debate paper on the World Health Organization ICD-11 Gaming Disorder proposal. J Behav Addict 2017 Sep 01;6(3):267-270 [FREE Full text] [doi: 10.1556/2006.5.2016.088] [Medline: 28033714]
17. Przybylski AK, Weinstein N, Murayama K. Internet Gaming Disorder: Investigating the Clinical Relevance of a New Phenomenon. Am J Psychiatry 2017 Mar 01;174(3):230-236. [doi: 10.1176/appi.ajp.2016.16020224] [Medline: 27809571]
18. Lam LT. Internet gaming addiction, problematic use of the internet, and sleep problems: a systematic review. Curr Psychiatry Rep 2014 Apr;16(4):444. [doi: 10.1007/s11920-014-0444-1] [Medline: 24619594]
19. Schou Andreassen C, Billieux J, Griffiths MD, Kuss DJ, Demetrovics Z, Mazzoni E, et al. The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: A large-scale cross-sectional study. Psychol Addict Behav 2016 Mar;30(2):252-262. [doi: 10.1037/adb0000160] [Medline: 26999354]
20. Király O, Griffiths MD, Urbán R, Farkas J, Kökönyei G, Elekes Z, et al. Problematic internet use and problematic online gaming are not the same: findings from a large nationally representative adolescent sample. Cyberpsychol Behav Soc Netw 2014 Dec;17(12):749-754 [FREE Full text] [doi: 10.1089/cyber.2014.0475] [Medline: 25415659]
21. Segal DL, Marty MA, Coolidge FL. Diagnostic Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). Washington DC: American Psychiatric Association; 2013.
22. ICD-11 for Mortality and Morbidity Statistics (ICD-11 MMS). 2018. URL: https://icd.who.int/browse11/l-m/en [accessed 2018-06-18]
23. Aarseth E, Bean AM, Boonen H, Colder Carras M, Coulson M, Das D, et al. Scholars' open debate paper on the World Health Organization ICD-11 Gaming Disorder proposal. J Behav Addict 2017 Sep 01;6(3):267-270 [FREE Full text] [doi: 10.1556/2006.5.2016.088] [Medline: 28033714]
24. Przybylski AK, Weinstein N, Murayama K. Internet Gaming Disorder: Investigating the Clinical Relevance of a New Phenomenon. Am J Psychiatry 2017 Mar 01;174(3):230-236. [doi: 10.1176/appi.ajp.2016.16020224] [Medline: 27809571]
25. Ho RC, Zhang MW, Tsang TY, Toh AH, Pan F, Lu Y, et al. The association between internet addiction and psychiatric co-morbidity: a meta-analysis. BMC Psychiatry 2014 Jun 20;14:183 [FREE Full text] [doi: 10.1186/1471-244X-14-183] [Medline: 24947851]
26. Cheng Y, Tseng P, Lin P, Chen T, Stubbs B, Carvalho AF, et al. Internet Addiction and Its Relationship With Suicidal Behaviors: A Meta-Analysis of Multinational Observational Studies. J Clin Psychiatry 2018 Jun 05;79(4):17r11761. [doi: 10.4088/JCP.17r11761] [Medline: 29877640]
27. Lau JTF, Walden DL, Wu AMS, Cheng K, Lau MCM, Mo PKH. Bidirectional predictions between Internet addiction and probable depression among Chinese adolescents. J Behav Addict 2018 Sep 01;7(3):633-643 [FREE Full text] [doi: 10.1556/2006.7.2018.87] [Medline: 30264608]
28. Wartberg L, Kriston L, Zieglmeier M, Lincoln T, Kammerl R. A longitudinal study on psychosocial causes and consequences of Internet gaming disorder in adolescence. Psychol Med 2019 Jan;49(2):287-294. [doi: 10.1017/S003329171800082X] [Medline: 29622057]
29. Cao H, Sun Y, Wan Y, Hao J, Tao F. Problematic Internet use in Chinese adolescents and its relation to psychosomatic symptoms and life satisfaction. BMC Public Health 2011 Oct 14;11:802 [FREE Full text] [doi: 10.1186/1471-2458-11-802] [Medline: 21995654]
30. Shek DTL, Yu L. Internet addiction phenomenon in early adolescents in Hong Kong. ScientificWorldJournal 2012;2012:104304. [doi: 10.1100/2012/104304] [Medline: 22778694]
31. Kocalevent R, Hinz A, Brähler E. Standardization of a screening instrument (PHQ-15) for somatization syndromes in the general population. BMC Psychiatry 2013 Mar 20;13:91 [FREE Full text] [doi: 10.1186/1471-244X-13-91] [Medline: 23514436]
32. Qian J, Ren ZQ, Yu DH, He XY, Li CB. The value of the Patient Health Questionnaire-15(PHQ-15) for screening somatic symptoms in general hospital. Chinese Mental Health Journal 2014;28(3):173-178. [doi: 10.3969/j.issn.1000-6729.2014.03.003]
33. Yu X, Tam WW, Wong PT, Lam TH, Stewart SM. The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. Compr Psychiatry 2012 Jan;53(1):95-102. [doi: 10.1016/j.comppsych.2010.11.002] [Medline: 21193179]
34. Zhang J, Zhang X. Chinese college students' SCL-90 scores and their relations to the college performance. Asian J Psychiatr 2013 Apr;6(2):134-140. [doi: 10.1016/j.ajp.2012.09.009] [Medline: 23466110]
35. Huang YP, Li L. SCL-90 as Measurement of Mental Health in College Students: a meta-analysis. Chinese Mental Health Journal 2009;23(5):366-371. [doi: 10.3969/j.issn.1000-6729.2009.05.015]
36. Kang Y, Guo W, Xu H, Chen Y, Li X, Tan Z, et al. The 6-item Kessler psychological distress scale to survey serious mental illness among Chinese undergraduates: Psychometric properties and prevalence estimate. Compr Psychiatry 2015 Nov;63:105-112. [doi: 10.1016/j.comppsych.2015.08.011] [Medline: 26555498]

37. Prochaska JJ, Sung H, Max W, Shi Y, Ong M. Validity study of the K6 scale as a measure of moderate mental distress based on mental health treatment need and utilization. Int J Methods Psychiatr Res 2012 Jun;21(2):88-97 [FREE Full text] [doi: 10.1002/mpr.1349] [Medline: 22351472]
38. Aloba O, Ojeleye O, Aloba T. The psychometric characteristics of the 4-item Suicidal Behaviors Questionnaire-Revised (SBQ-R) as a screening tool in a non-clinical sample of Nigerian university students. Asian J Psychiatr 2017 Apr;26:46-51. [doi: 10.1016/j.ajp.2017.01.017] [Medline: 28483090]
39. Gierk B, Kohlmann S, Kroenke K, Spangenberg L, Zenger M, Brähler E, et al. The somatic symptom scale-8 (SSS-8): a brief measure of somatic symptom burden. JAMA Intern Med 2014 Mar;174(3):399-407. [doi: 10.1001/jamainternmed.2013.12179] [Medline: 24276929]
40. Wu X, Zhang Z, Zhao F, Wang W, Li Y, Bi L, et al. Prevalence of Internet addiction and its association with social support and other related factors among adolescents in China. J Adolesc 2016 Oct;52:103-111. [doi: 10.1016/j.adolescence.2016.07.012] [Medline: 27544491]
41. Gmitrowicz A, Szymczak W, Kotlicka-Antczak M, Rabe-Jabłońska J. Suicidal ideation and suicide attempt in Polish adolescents: is it a suicidal process? Int J Adolesc Med Health 2003;15(2):113-124. [doi: 10.1515/ijamh.2003.15.2.113] [Medline: 12955813]
42. Nock MK, Green JG, Hwang I, McLaughlin KA, Sampson NA, Zaslavsky AM, et al. Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: results from the National Comorbidity Survey Replication Adolescent Supplement. JAMA Psychiatry 2013 Mar;70(3):300-310 [FREE Full text] [doi: 10.1001/2013.jamapsychiatry.55] [Medline: 23303463]
43. Jeon HJ, Lee J, Lee YM, Hong JP, Won S, Cho S, et al. Lifetime prevalence and correlates of suicidal ideation, plan, and single and multiple attempts in a Korean nationwide study. J Nerv Ment Dis 2010 Sep;198(9):643-646. [doi: 10.1097/NMD.0b013e3181ef3ecf] [Medline: 20823725]
44. Booker CL, Kelly YJ, Sacker A. Gender differences in the associations between age trends of social media interaction and well-being among 10-15 year olds in the UK. BMC Public Health 2018 Mar 20;18(1):321 [FREE Full text] [doi: 10.1186/s12889-018-5220-4] [Medline: 29554883]
45. Han J, Batterham PJ, Calear AL, Ma J. Seeking professional help for suicidal ideation: A comparison between Chinese and Australian university students. Psychiatry Res 2018 Dec;270:807-814. [doi: 10.1016/j.psychres.2018.10.080] [Medline: 30551329]
46. Yin H, Wardenaar KJ, Xu G, Tian H, Schoevers RA. Help-seeking behaviors among Chinese people with mental disorders: a cross-sectional study. BMC Psychiatry 2019 Nov 29;19(1):373 [FREE Full text] [doi: 10.1186/s12888-019-2316-z] [Medline: 31783825]
47. Statistical Bulletin on Education Development. Ministry of Education of the People's Republic of China. URL: http://www.moe.gov.cn/jyb_sjzl/sjzl_fztjgb/ [accessed 2019-07-24]
48. Gong X, Li B, Yan HF. Comparative Analysis of the Relationship Between Gross Enrollment Rate in Higher Education, Gender Equality Difference and Economy Development. Journal of Hubei Radio & Television University 2016 Jan 06;51(1):20-22. [doi: 10.1111/hea.12061_15]
49. Wang MJ, Tao QY. Study of relationship between academical achievement and mental health of adolescents. Chin J Behav Med Sci 2000;9(5):341-351. [doi: 10.3760/cma.j.issn.1674-6554.2000.05.009]

Abbreviations

ADHD: attention deficit and hyperactivity disorder.
CIDI: Composite International Diagnostic Interview
DSM: Diagnostic and Statistical Manual of Mental Disorders
IAT: Young 20-item Internet Addiction Test
ICD: International Classification of Diseases
IGD: internet gaming disorder
K6: six-item Kessler psychological distress scale
OPHS: Online Psychosomatic Health Survey
OR: odds ratio
PHQ: Patient Health Questionnaire
PIU: problematic internet use
SBQ-R: Suicidal Behaviors Questionnaire-Revised
SCL-90: Symptom Checklist 90

Edited by G Eysenbach; submitted 20.12.19; peer-reviewed by J Saunders, M Torok, Z Ma, T Muto; comments to author 10.03.20; revised version received 02.05.20; accepted 03.06.20; published 11.08.20.

Please cite as:
Guo W, Tao Y, Li X, Lin X, Meng Y, Yang X, Wang H, Zhang Y, Tang W, Wang Q, Deng W, Zhao L, Ma X, Li M, Chen T, Xu J, Li J, Hao W, Lee S, Coid JW, Greenshaw AJ, Li T
Associations of Internet Addiction Severity With Psychopathology, Serious Mental Illness, and Suicidality: Large-Sample Cross-Sectional Study
J Med Internet Res 2020;22(8):e17560
URL: https://www.jmir.org/2020/8/e17560
doi:10.2196/17560
PMID:32780029

Original Paper

Modeling Early Gambling Behavior Using Indicators from Online Lottery Gambling Tracking Data: Longitudinal Analysis

Gaëlle Challet-Bouju[1,2], PhD; Jean-Benoit Hardouin[2,3], PhD; Elsa Thiabaud[1], MSc; Anaïs Saillard[1], MSc; Yann Donnio[1], BSc; Marie Grall-Bronnec[1,2], MD, PhD; Bastien Perrot[2,3], PhD

[1]Addictology and Psychiatry Department, Centre Hospitalier Universitaire de Nantes, Nantes, France

[2]SPHERE, INSERM UMR1246, University of Nantes, University of Tours, Nantes, France

[3]Biostatistics and Methodology Unit,, Department of Clinical Research and Innovation, Centre Hospitalier Universitaire de Nantes, Nantes, France

Corresponding Author:
Gaëlle Challet-Bouju, PhD
Addictology and Psychiatry Department
Centre Hospitalier Universitaire de Nantes
Hôpital Saint Jacques, 85 rue Saint Jacques
Nantes, 44093
France
Phone: 33 253482532
Email: gaelle.bouju@chu-nantes.fr

Abstract

Background: Individuals who gamble online may be at risk of gambling excessively, but internet gambling also provides a unique opportunity to monitor gambling behavior in real environments which may allow intervention for those who encounter difficulties.

Objective: The objective of this study was to model the early gambling trajectories of individuals who play online lottery.

Methods: Anonymized gambling ‐ related records of the initial 6 months of 1152 clients of the French national lottery who created their internet gambling accounts between September 2015 and February 2016 were analyzed using a two-step approach that combined growth mixture modeling and latent class analysis. The analysis was based upon behavior indicators of gambling activity (money wagered and number of gambling days) and indicators of gambling problems (breadth of involvement and chasing). Profiles were described based upon the probabilities of following the trajectories that were identified for the four indicators, and upon several covariates (age, gender, deposits, type of play, net losses, voluntary self-exclusion, and Playscan classification—a responsible gambling tool that provides each player with a risk assessment: green for low risk, orange for medium risk and red for high risk). Net losses, voluntary self-exclusion, and Playscan classification were used as external verification of problem gambling.

Results: We identified 5 distinct profiles of online lottery gambling. Classes 1 (56.8%), 2 (14.8%) and 3 (13.9%) were characterized by low to medium gambling activity and low values for markers of problem gambling. They displayed low net losses, did not use the voluntary self-exclusion measure, and were classified predominantly with green Playscan tags (range 90%-98%). Class 4 (9.7%) was characterized by medium to high gambling activity, played a higher breadth of game types (range 1-6), and had zero to few chasing episodes. They had high net losses but were classified with green (66%) or orange (25%) Playscan tags and did not use the voluntary self-exclusion measure. Class 5 (4.8%) was characterized by medium to very high gambling activity, played a higher breadth of game types (range 1-17), and had a high number of chasing episodes (range 0-5). They experienced the highest net losses, the highest proportion of orange (32%) and red (39%) tags within the Playscan classification system and represented the only class in which voluntary self-exclusion was present.

Conclusions: Classes 1, 2, 3 may be considered to represent recreational gambling. Class 4 had higher gambling activity and higher breadth of involvement and may be representative of players at risk for future gambling problems. Class 5 stood out in terms of much higher gambling activity and breadth of involvement, and the presence of chasing behavior. Individuals in classes 4 and 5 may benefit from early preventive measures.

(J Med Internet Res 2020;22(8):e17675) doi:10.2196/17675

KEYWORDS

gambling; internet; trajectory; latent class analysis; growth mixture modeling; gambling tracking data; early detection

Introduction

The prevalence of past-year gambling problems varies from 0.1% to 5.8% worldwide [1]. In France, it was estimated at 2.7% of the general population between 15 and 75 years old [2]. This prevalence reached 22.4% among the population of individuals who gambled online in 2017 [3], compared to 17.0% in 2012 [4]. Internet gambling has intrinsic features that may facilitate excessive gambling, such as high accessibility, anonymity, high frequency of gambling outcomes, and digital payment modes [5-7]. Individuals who gamble online have a higher risk and a higher severity of gambling problems [8-12].

Online gambling also provides a unique opportunity to monitor gambling behavior in real environments [13]. The larger part of gambling research has been performed using questionnaire-reported subjective data, but self-reported data has been widely criticized (even beyond the framework of gambling) because it lacks both accuracy and validity and is prone to numerous biases [14-17]. It has been emphasized that too little research has been conducted in a real gambling environment with individuals who gamble [18,19]. Thus, a stream of gambling-related research has been undertaken using of gambling tracking data (naturalistic data) [20]. Large research programs [21-36] have been initiated including research [36] that focuses on rarely studied online lottery gambling tracking data (lottery draws, daily lotteries, and scratch cards). Indeed, such gambling is traditionally considered the least associated with gambling problems, but is the most prevalent form of gambling, leading to an overall sum of low-level harms that could be as important as the harm associated with more problematic but less prevalent forms of gambling [37].

Gambling tracking data allow researchers to easily access activity such as money wagered, number of gambling days, deposits, wins, and losses. Such indicators are informative, but do not, on their own, identify the potential for future gambling problems (since individuals who engage heavily in gambling are not necessarily those who develop gambling problems). Consequently, the combination of activity-based indicators with indicators that are related to core features of addiction may better capture individuals at risk for gambling problems. One potential indicator of addiction is chasing behavior which is defined as the continuation or intensification of gambling after a sequence of losses with the objective to recover previous losses [38]. This behavior is almost omnipresent in individuals with gambling problems and has been identified as the most significant step in the development of pathological gambling [38-40]. It is considered a key indicator of problem gambling behavior, especially in research that uses gambling tracking data [20,41]. The identification of chasing episodes may be performed in either a between- (over a long timescale spanning multiple sessions) or within-session (bet-by-bet behavior) dimension [38]; however, the latter appears to be better at capturing chasing behavior [20]. Researchers often use the modification of betting behavior depending on previous betting outcomes as a proxy to identify chasing episodes, as it is a behavior that is not directly observable [20]. In recent work [36,42], we proposed to approximate within-session chasing behavior by focusing on recurrent deposits within a short period of time or deposits that occurred immediately after a bet. When a deposit is made in these conditions, it likely means that the person has recently lost money; the deposit indicates the unplanned continuation of gambling in an attempt to recover losses. Another potential indicator of problem gambling is the breadth of involvement, which is generally defined as the number of different games played by an individual [30] and is considered a form of variability in gambling [20]. It has been found to be higher for those who gamble online compared to those who gamble offline [12,30,43] and could be a mediator in the relationship between online gambling and gambling problems [44].

Rather than using cross-sectional data, a longitudinal approach may be more relevant to identify individuals at risk for gambling problems. Gambling activity (frequency and intensity) and gambling variability (daily variability and trajectory, ie, the increasing or decreasing pattern of wagers) have been used to monitor gambling patterns in live-action sports betting by individuals during the first month after their gambling account is created and to determine the association of patterns with the development of later gambling problems [23]. The results indicated that individuals with high-activity and high-variability gambling were more likely to close their account due to gambling-related problems. Moreover, in previous work [36], we found that risky monthly behaviors were associated with larger deviations from the usual gambling activity in online lotteries. Finally, in a recent review, trajectory information has been reported as a possible method with good predictive performance in identifying problem gambling behavior through online gambling tracking data analysis [20].

In this study, the early gambling trajectories of online lottery gambling during the first month after the creation of an account is modeled. The objective was to identify distinct profiles of individuals who gamble online which can be distinguished by their early trajectories and to characterize their patterns in relation to their potential for gambling problems. This work is part of the first stage of EDEIN (*Etude de Dépistage des comportements Excessifs de jeu sur Internet*; *Screening for Excessive Gambling Behaviors on the Internet*) [42] (ClinicalTrials.gov NCT02415296; http://clinicaltrials.gov/ct2/show/NCT02415296).

Methods

Participants

An anonymized data set was used; this data set had also been used in a previous analysis [36]. The data were comprised of gambling - related records from a random sample of 10,000 clients of the French national lottery and included the age, gender, and Playscan classification—a responsible gambling tool that provides a low, medium, or high risk assessment—of each client. This operator is the only one in France that is

permitted to offer online lotteries and scratch games, and thus, represents 100% of the market for online lotteries and scratch games in France [36]. The Playscan classification is a risk assessment based on an individual's 5 previous weeks of gambling activity: green corresponds to a low risk of problem gambling, orange corresponds to a medium risk, and red corresponds to a high risk [45]. It uses a combination of quantitative (wagers, deposits, gambling session duration, etc) and qualitative (gambling in risky periods, etc) gambling behavior data to estimate risk of problem gambling. The initial dataset (n=10,000) included individuals who gambled at least once between September 2015 and August 2016. In this study, we were interested in early behaviors only; therefore, we restricted inclusion to individuals who created their account between September 2015 and February 2016 (n=1152). As a consequence, it was possible to observe gambling activity during the six months following account creation. We chose a period of six months in order to observe early gambling behaviors without focusing solely on the first few weeks, which are not necessarily representative of future gambling activity.

Data Reduction

To observe changes in early gambling trajectories, each variable was computed based on a 15-day unit of time, starting with $t=t_0$ on the day of account creation, resulting in 12 equidistant time points ($t=t_0, t_1, t_2,\ldots, t_{11}$) over the 6-month period for each newly registered individual.

To conduct the trajectory analysis, 4 measures were selected from the original data set and calculated for each 15-day unit of time: amount wagered (the sum of all the bets made), number of gambling days (the total number of days with at least one gambling session), the number of chasing episodes (the number of times that money was deposited into the gambling account was used as a proxy of chasing behavior and applicable only if criteria were met—deposits either 3 or more times within a 12-hour period or less than 1 hour after a bet) [36,42], and involvement (the number of different games played). The amount wagered and number of gambling days were used as indicators of gambling activity, while the number of chasing episodes and involvement were used as indicators of at-risk gambling behaviors.

In addition, we used the following covariates to characterize the profiles that were identified: gender, age, cumulative deposits over the 6-month period, largest single-day deposit during the 6-month period, percentage of bets on instant lotteries (scratch cards and instant draws) over the 6-month period, cumulative net loss over the 6-month period (calculated by subtracting winnings and promotional e-credits from wagers), voluntary self-exclusion during the 6-month period (a categorical variable defined as either yes, if there was at least one episode of self-exclusion, or no, if there were zero episodes of self-exclusion during the 6-month period), and the individual's highest Playscan classification during the 6-month period. Net loss, voluntary self-exclusion, and Playscan classification were chosen to serve as external verification of problem gambling [20]. All amounts were recorded in euros and have been converted to US dollars (a currency exchange rate of €1=US $1.084 was applicable at the time of publication).

Statistical Analysis

We used a two-step approach to establish typologies that group individuals who evolve differently over time (individuals with similar trajectories identified for the 4 indicators) using data from the initial six months of their online gambling subscription.

The first step of the analysis consisted of identifying trajectories for each of the 4 gambling indicators. To model the evolution of each indicator (measured at 12 discrete time points for each individual, in a highly heterogeneous population [46]), we used growth mixture models [47,48]. Models were selected based on statistical criteria and upon interpretation of the classes. For each indicator, we tested multiple models to determine the best number of trajectories (from 1 to 8). For each number of trajectories, we computed several models, starting with the most complex (linear, quadratic, and cubic time-dependent terms, an intercept term, and random effects on both intercept and slopes), which was then simplified, if necessary, based on statistical criteria (convergence and stability of the model, Bayesian information criterion, significance of the parameters). For each model, we randomly generated 400 sets of initial values, for which 10 iterations of the expectation-maximization algorithm were performed. For the 50 best solutions, the entire expectation-maximization algorithm was performed. A model was considered to be stable if the best log-likelihood was replicated (ie, if at least two solutions had the same final best log-likelihood). The selected trajectory models were compared with one another to determine the best model for each indicator. The outcomes of the growth mixture models were membership probabilities (the probability that an individual belonged to each modeled trajectory).

The second step was to perform classification by grouping individuals with similar indicator trajectories. A latent class analysis was performed using the trajectory membership probabilities as the observed variable. This strategy allowed us to define latent classes with distinct characteristics of gambling activity and at-risk behavior. Model selection was based on a trade - off between the Bayesian information criterion, the classification error rate (which reflects the precision of the classification), the interpretability of the classes, and the replicability of the model. The classes were described by the trajectory membership probabilities of each indicator and by the covariates.

Such a two-step analysis strategy has been previously used in studies [49,50] in various health areas, including behavioral addictions. Growth mixture models were estimated using Mplus software (version 8.1; Muthen and Muthen) [51]. Latent class analysis was conducted using Latent Gold software (Statistical Innovations Inc) [52].

Ethics

This study was approved by the research ethics committee (*Groupe Nantais d'Ethique dans le Domaine de la Santé*) on March 25, 2015. Because of the retrospective and noninterventional design of this study, consent of the individuals whose data were used was deemed unnecessary.

Results

Characteristics

The characteristics of the sample, which was composed of 1152 individuals who had newly registered for an online gambling account, are described in Table 1. The sample was mainly composed of men (male: 740/1152, 64.2%; female: 412/1152, 35.8%) and the sample had a mean of 39.83 (SD 12.65) years of age. Table 1 shows that the indicators of gambling activity varied highly both between individuals (between-subject SD) and over the 6-month period for a given individual (within-subject SD).

Table 1. Demographic and gambling characteristics of individuals who had newly registered for an online gambling account (over a period of 6 months).

Characteristics	Individuals (N=1152) n (%)	Mean[a]	SD	SD, between-subject[b]	SD, within-subject[c]	Minimum	Maximum
Age, years	—	39.83	12.65	N/A[d]	N/A	19	81
Gender							
Male	740 (64.2)	—	—	—	—	—	—
Female	412 (35.8)	—	—	—	—	—	—
Gambling activity							
Money wagered, €[e]	—	16.98	68.83	52.82	44.15	0.00	1918.00
Gambling days, n	—	1.27	2	1.48	1.36	0	15
Chasing, n	—	0.10	0.85	0.55	0.65	0	26
Involvement, n	—	1.15	2.23	1.5	1.66	0	31
Deposits, €	—	10.46	35.29	25.98	23.89	0.00	835.00
Largest single-day deposit, €	—	7.32	17.69	10.44	14.29	0.00	500.00
Losses, €	—	4.43	417.85	120.8	400	–48748.40	712.40
Instant lotteries, %	—	11	28	19	21	0	100
Voluntary self-exclusion	6 (0.5)	N/A	N/A	N/A	N/A	N/A	N/A
Playscan							
Green	1032 (89.6)	N/A	N/A	N/A	N/A	N/A	N/A
Orange	83 (7.2)	N/A	N/A	N/A	N/A	N/A	N/A
Red	36 (3.1)	N/A	N/A	N/A	N/A	N/A	N/A

[a]For quantitative variables, data were averaged at a monthly level over the 6-month period. This was intended to be more representative and meaningful than the 15-day unit of time used for trajectory analyses.

[b]between SD: represents the fluctuations of monthly gambling activity between the individuals (between-subject standard deviation).

[c]within SD: represents the fluctuations of monthly gambling activity within the 6-month period for a given individual (within-subject standard deviation).

[d]N/A: not applicable.

[e]At the time of publication, a currency exchange rate of €1=US $1.084 was applicable.

Growth Mixture Models

Seven trajectories were obtained for amount wagered and are shown in Figure 1. Trajectories 1 (13.0%), 3 (12.3%), and 6 (14.6%) exhibited a downward trend, initially showing a medium (trajectory 6) to high (trajectories 1 and 3) amount of money during initial weeks which gradually diminished to near zero amounts at the end of the 6-month period. Trajectories 2 (7.1%), 4 (26.4%), 5 (23.4%), and 7 (3.2%) were stable over the 6-month period. Trajectory 5 was characterized by low wagers—under €2 (US $2.17) per two weeks, trajectory 4 by medium wagers—€4-€12 (US $4.34-$13.01) per two weeks, trajectory 2 by high wagers—approximately €18 (US $19.51) per two weeks, and trajectory 7 by very high wagers—mainly €60-€100 (US $65.04-$108.40) per two weeks.

Three trajectories were obtained for the number of gambling days (Figure 2). Trajectory 1 (9.8%) was an inverted parabolic, increasing from 2.38 to 4.58 gambling days per two weeks, reaching a maximum at 120 days, and decreasing to reach 3.29 days at the end of the 6-month period. Trajectory 2 (9.8%) decreased rapidly from an initially high number of gambling days (6.8) which rapidly decreased to 0.98 days. Trajectory 3 represented the majority of individuals (83.4%) and had a relatively stable and low trajectory, with the number of gambling days fluctuating between 0.42 and 1.67.

Figure 1. Trajectories obtained from growth mixture models for amount wagered. The ordinate axis represents the log-transformation of amount wagered. A value of 1 corresponds to €2, 2 to €6, 3 to €19, 4 to €54, and 5 to €147.

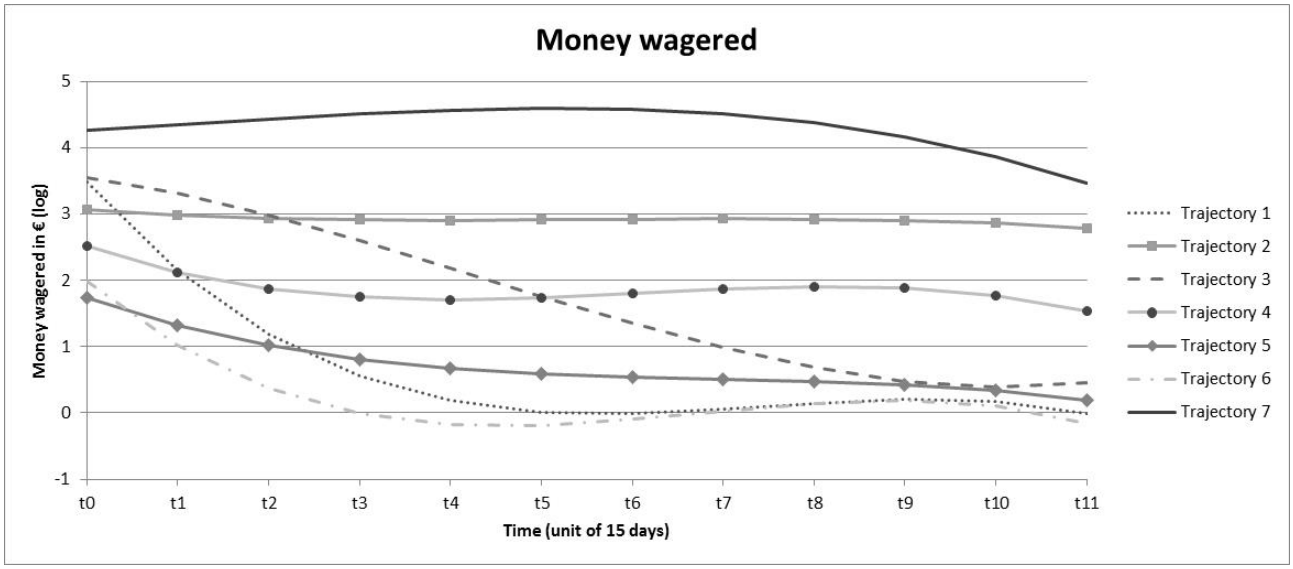
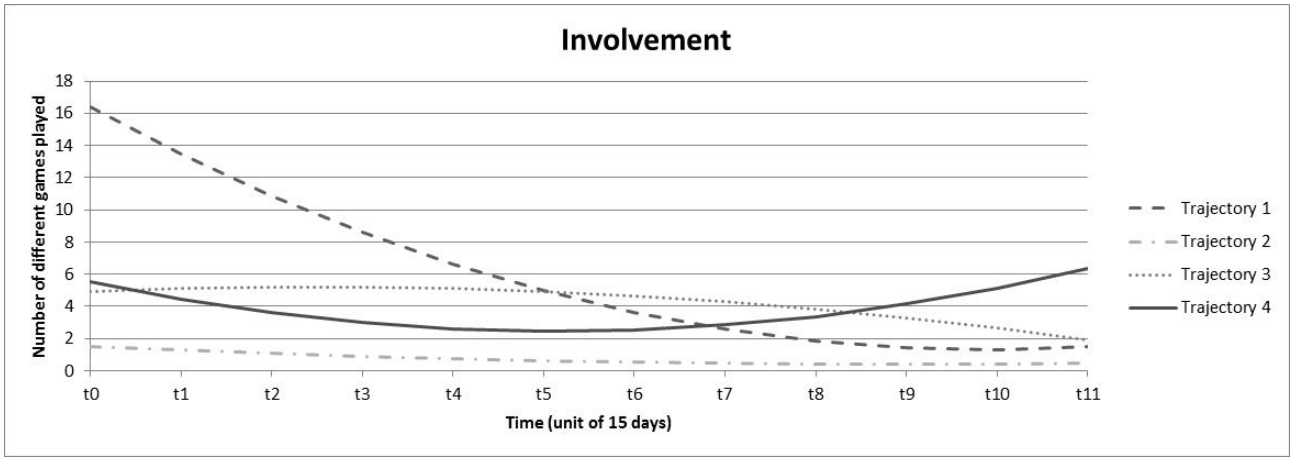


Figure 2. Trajectories obtained from growth mixture models for number of gambling days.

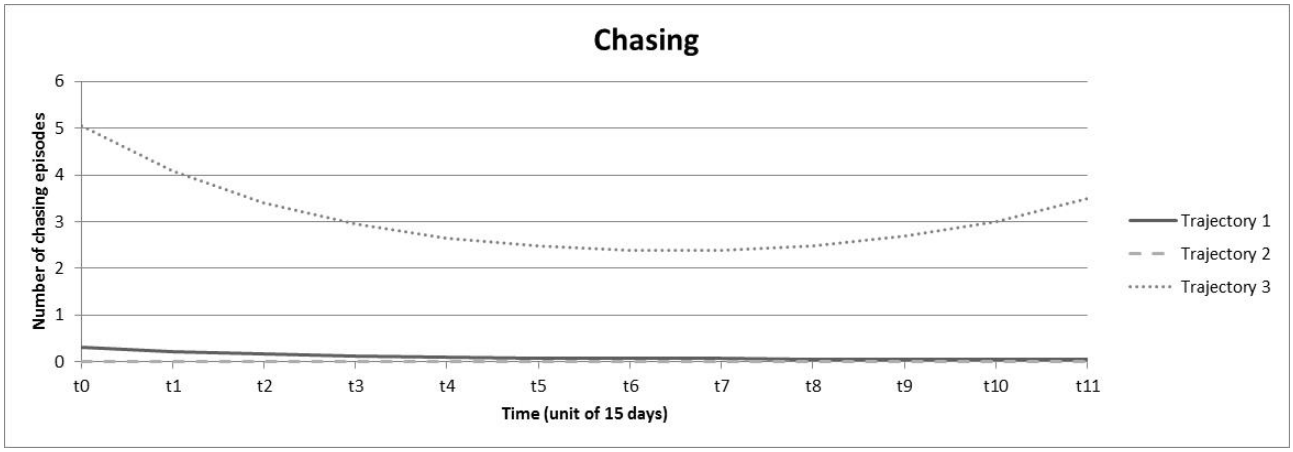
Four trajectories were obtained for the involvement indicator (Figure 3). Trajectory 1 represented a very low proportion of individuals (1.3%), initially showing a very high diversity of games played (16.38 per two weeks) that rapidly decreased, and then stabilized to reach between 1.29-1.84 games after 120 days. Trajectory 2 represented the majority of individuals (89.9%) and was stable and low with approximately 1 (range 0.43-1.49) game played throughout the 6-month period. Trajectory 3 (5.8%) was also relatively stable, but with a higher number of games played, varying between 1.91 and 5.21. Finally, trajectory 4 (3.0%) exhibited a reverse parabolic shape, decreasing from 5.52 to 2.61 games played during the first 60 days, followed by a more pronounced rise after 90 days to reach 6.37 games at the end of the 6-month period.

Figure 3. Trajectories obtained from growth mixture models for involvement.



Three trajectories were identified for chasing (Figure 4). Trajectories 1 and 2 were similar and represented the majority of individuals (trajectory 1: 33.9%; trajectory 2: 64.2%). They were characterized by the absence (trajectory 2: 0 episodes) or a very low number (trajectory 1: 0.06-0.31 episodes) of chasing episodes during the 6-month period. Trajectory 3 (1.9%) was an inverse parabolic and had a very high initial number (5.05 episodes per two weeks) which decreased to 2.39 episodes at 90 days and increased to 3.49 episodes at the end of the 6-month period.

Figure 4. Trajectories obtained from growth mixture models for chasing.



Latent Class Analysis

Fit indices of the models (1 to 8 classes) are given in Table 2. Bayesian information criteria decreased from the 1-class to the 8-class solution; however, the 7- and 8-class models were not stable. Because the 6-class solution did not yield a significant change in interpretation compared to that of the 5-class solution, the simplest model (ie, the 5-class model) was chosen.

Table 2. Fit indices of the 1- to 8-class models used to select the final model. A 5-class solution was selected.

Model type	Log-likelihood	Bayesian information criterion	Number of parameters	Classification errors
1-Class	–955	2150	34	0
2-Class	30077	–59667	69	0.0004
3-Class	38004	–75274	104	0.0006
4-Class	39901	–78823	139	0.0011
5-Class	43508	–85790	174	0.0016
6-Class	45205	–88937	209	0.0018
7-Class[a]	47402	–93085	244	0.0033
8-Class[a]	47837	–93707	279	0.0026

[a]Final log-likelihood not replicated.

Table 3 shows the distribution of the trajectories obtained for the gambling indicators and covariates for each class.

Class 1 (56.8%) was characterized by low gambling activity and low values for indicators of addiction; most individuals had the lowest trajectory of involvement and either a trajectory with no (p=.89) or few (p=.11) chasing episodes. Demographics of this class were similar to those of the overall sample (male: 64.8%; female: 35.2%; mean age: 39.06 years). Individuals from this class predominantly played deferred lotteries (79% of lotteries). They had low cumulative losses (€37.22) and deposits (€48.15) over the 6-month period. Almost all individuals (97.9%) were within the green Playscan classification.

Class 2 (14.8%) was characterized by medium to high and stable level of money wagered. For the number of gambling days, individuals in class 2 had either a low and stable trajectory (p=.61) or a medium and parabolic trajectory (p=.34). Moreover, they had similar indicators of addiction as those in class 1, but with a higher probability (proportion) for the trajectory that represented few chasing episodes (p=.59 compared to p=.11). The mean age of 43.41 years was higher than that of the overall sample and the proportion of women was lower than that in the overall sample (26.5%). These individuals were the only ones who had negative losses (ie, won €161.13), despite higher cumulative deposits (€154.97) than class 1. They predominantly played deferred lotteries (88% of lotteries) and were mainly (90.0%) within the green Playscan classification.

Table 3. Distribution of the trajectories for the gambling indicators and description of covariates among the 5 classes.

Model outcomes	Class 1	Class 2	Class 3	Class 4	Class 5
Probabilistic class size, in %	56.8	14.8	13.9	9.7	4.8
Probabilities, *p*					
Amount wagered					
Trajectory 1	.17[a]	<.001	.24[a]	<.001	.002
Trajectory 2	<.001	.37[a]	<.001	.001	.33[a]
Trajectory 3	.06	.02	.35[a]	.28[a]	.24[a]
Trajectory 4	.20[a]	.57[a]	.08	.52[a]	.13[a]
Trajectory 5	.33[a]	.03	.28[a]	.02	<.001
Trajectory 6	.24[a]	<.001	.05	<.001	<.001
Trajectory 7	<.001	.003	<.001	.18[a]	.30[a]
Gambling days					
Trajectory 1	.003	.34[a]	.004	.23[a]	.47[a]
Trajectory 2	.001	.05	.16[a]	.25[a]	.31[a]
Trajectory 3	.997[a]	.61[a]	.84[a]	.52[a]	.22[a]
Chasing					
Trajectory 1	.11[a]	.59[a]	.71[a]	.67[a]	.48[a]
Trajectory 2	.89[a]	.41[a]	.29[a]	.33[a]	.13[a]
Trajectory 3	<.001	<.001	<.001	<.001	.39[a]
Involvement					
Trajectory 1	<.001	<.001	<.001	<.001	.27[a]
Trajectory 2	.999[a]	.998[a]	.994[a]	.39[a]	.15[a]
Trajectory 3	.001	.002	.005	.41[a]	.35[a]
Trajectory 4	<.001	.001	.001	.19[a]	.23[a]
Covariates					
Age[b], years	39.06	43.41	38.21	4.33	41.68
Gender, %[c]					
Male	64.8	73.5	62.1	54.9	52.0
Female	35.2	26.5	37.9	45.1	48.0
Voluntary self-exclusion, %[c]	0	0	0	0	10.8
Cumulative losses[d], €[e]	37.22	–161.13	80.85	189.65	541.64
Cumulative deposits[d], €[e]	48.15	154.97	102.91	232.99	797.05
Largest single day deposit, €[e]	23.27	31.02	31.19	39.16	74.91
Instant lotteries, %[c]	21	12	39	65	78
Playscan, %[c]					
Missing	0.5	0	0.6	0	0
Green	97.9	90.0	9.6	66.4	28.7
Orange	1.5	8.8	18.1	24.6	31.8

Model outcomes	Class 1	Class 2	Class 3	Class 4	Class 5
Red	0.2	1.2	0.6	9.0	39.5

[a]These probabilities are the main trajectories represented within each class (p>.10).

[b]Values represent the mean for each class.

[c]Probability of belonging to each class. The percentages indicated refer to this probabilistic approach but do not represent a proportion of individuals.

[d]Cumulative over the 6-month period.

[e]At the time of publication, a currency exchange rate of €1=US $1.084 was applicable.

Class 3 (13.9%) was characterized by generally decreasing trends in gambling activity (amount wagered and the number of gambling days). Moreover, these individuals had similar indicators of addiction to those in class 2, but with a higher probability (proportion) for the trajectory that represented few chasing episodes (p=.71 compared to p=.59). Demographics of this class were similar to those of the overall sample (male: 62.1%; female: 37.9%; mean age: 38.21 years). These individuals displayed moderate cumulative losses (€80.85) and deposits (€102.91) over the 6-month period. They predominantly played deferred lotteries (61% of lotteries), but to a lesser extent than those in classes 1 and 2, and were mainly (90.6%) within the green Playscan classification.

Class 4 (9.7%) was characterized by medium to very high wagers, with the moderate and stable trajectory most represented (p=.52). For number of gambling days, individuals were distributed across the three trajectories with predominance in the low stable trajectory (p=.52). In contrast to classes 1, 2, and 3, this class was characterized by a combination of three trajectories for involvement with the majority in low and medium stable trajectories (p=.39 and p=.41, respectively). Moreover, individuals from class 4 had a similar pattern of chasing episodes as those from class 3. As a consequence, the pattern of class 4 was characterized by a diversification of games played but not by an increase in chasing episodes which remained relatively rare. The mean age of 40.33 years was similar to that of the overall sample and the proportion of women was higher than that in the overall sample (45.1%). These individuals displayed high cumulative losses (€189.65) and deposits (€232.99) over the 6-month period. In contrast to the first three classes, they predominantly played instant lotteries (65% of lotteries). The majority of individuals from this class were within the green Playscan classification (66.4%), but a significant proportion were within the orange (24.6%), and to a lesser extent, red (9.0%) classifications.

Class 5 (4.9%) was characterized by medium to high wagers, with the stable and high (p=.33) and the stable and very high (p=.30) trajectories most represented. Individuals from class 5 were also distributed in the three trajectories for the number of gambling days, but with a relative predominance of the high and parabolic trajectory (p=.47). Thus, individuals from this class had high or very high levels of gambling activity which remained high throughout the 6-month period. They were also characterized by a higher levels of involvement (all 4 trajectories) and were the only ones for which the highest-level trajectory of chasing was represented (p=.39). As a consequence, this class was characterized by both a diversification of games played and a large increase in chasing episodes. The mean age of 41.68 years was similar to that of the overall sample and the proportion of women was higher than that in the overall sample (48%). These individuals had very high cumulative losses (€541.64) and deposits (€797.05) over the 6-month period which were approximately 15 times more than the individuals in class 1 experienced and 3 times more than the individuals in class 4 experienced. They were also characterized by very high largest single-day deposits (€74.91) compared to those of the other classes. They predominantly played instant lotteries (78% of lotteries) and were mainly within orange (31.8%) and red (39.5%) Playscan classifications. Finally, this was the only class for which voluntary self-exclusions were present (10.8%).

Discussion

Principal Findings

The aim of this study was to investigate the early gambling trajectories of individuals over the initial 6 months of their subscription to the French online lottery website. We identified 5 distinct profiles of online lottery gambling, and we characterized each pattern in relation to indicators of gambling activity and gambling addiction.

The first three classes represented the majority of individuals (85.5%) and were characterized by low to medium gambling activity and low levels of problem gambling indicators. According to our assumptions, such individuals do not seem to encounter difficulty with their gambling practices and their gambling may be considered recreational. This was supported by findings of low losses (and sometimes wins), no voluntary self-exclusion, and low-risk Playscan classification (green tags).

In parallel, we identified 2 profiles that should be considered for early intervention or harm minimization. Class 4 represented approximately 10% of the sample and was characterized by medium to high gambling activity, a diversification of games played, and zero to few chasing episodes. The higher breadth of involvement may reflect more variable gambling activity and has previously been associated with account closure and classification of the individual as high risk by responsible gambling indicators [20]. Engagement in multiple online gambling activities has been identified as a potential predictor of high-risk gambling or the emergence of gambling problems [5]. Consequently, we can suggest that individuals in Class 4 may be at risk for future gambling problems but may not yet be gambling excessively. This assumption is supported by the fact that these individuals displayed high losses but were predominantly considered low and medium risk (green and orange, respectively) and did not use the voluntary self-exclusion measure. They played predominantly instant lotteries (65%), which have been found to be more associated with gambling

problems than deferred lotteries are, due to high event frequency [7,36,53].

Class 5 was the smallest (4.8%) and was characterized by medium to very high gambling activity, a higher diversification of games played, and a high number of chasing episodes. Higher breadth of involvement and chasing behavior have been consistently associated with problem gambling [20]. Chasing, in particular, is considered a key indicator of problem gambling [20,41] and has been found to be the best discriminator between social and problem gambling among women [54]. Because class 5 was the only class with nonzero chasing trajectories, there is strong reason to believe that such gambling may be problematic. This is supported by the fact that these individuals experienced higher net losses and had the highest proportion among their class (71.3%) who were considered medium and high risk (orange and red tags). Moreover, they played forms of lottery that are more associated with gambling problems (ie, instant lotteries). More importantly, this class was the only class in which voluntary self-exclusion was present in a proportion (10.8%) that was higher than that observed in general online gambling (1%) [55], but close to that found in at-risk individuals who gamble online (11%) [56]. Voluntary self-exclusion is a harm-minimization strategy for individuals who experience gambling problems. In the case of online gambling, it consists of voluntarily banishing oneself from gambling websites for a predefined period [57]. Voluntary self-exclusion is considered a valid proxy indicator to externally verify which individuals have gambling problems [20], and thus, reinforced our deduction that individuals in class 5 may have gambling problems.

The whole sample was characterized by a higher proportion of women (35.8%) than that found in other studies using gambling tracking data from online gambling: 5.5% [29], 8% [23], and 10% [35]. Given that games of chance, such as lotteries, are more appealing to women than skill-based games are [58], this higher proportion was expected. This finding is also consistent with a French survey on the prevalence of online gambling [3] which found that 38.8% of the individuals involved in online lotteries were women; however, it was surprising that we found a higher proportion of women in the two classes exhibiting at-risk profiles. This was unexpected because it has generally been shown that men gamble more than women [1,59] and that at-risk individuals are predominantly men [60,61]. This unexpectedly higher proportion of women in at-risk classes was previously described in the same data set in a study that did not restrict inclusion to newly registered individuals [36]. It is worth noting that this tendency was maintained in early trajectories, making women a particularly vulnerable population for problem gambling when they initially open online accounts. As reported in Perrot et al [36], women with gambling problems may prefer to gamble online because they experience less stigma [6] and because they may be more socially anxious [62]. Moreover, boredom has been found to be both a motivating factor for gambling and a factor associated with continued problem gambling among women [58,63]. One can hypothesize that, beyond social anxiety and stigma, women with gambling problems may tend to choose online gambling to avoid boredom because it is accessible 24/7 and may represent a way to stay occupied to avoid negative emotional states.

Limitations and Strengths

This study has several limitations. First, because we were not able to directly measure gambling problems, we were constrained to use proxies (net losses, Playscan status, and voluntary self-exclusion); however, these proxies have previously been used for external verification of gambling problems in gambling studies that use tracking data [20]. Otherwise, gambling problems have generally been measured using self-report questionnaires, such as in LaPlante et al [30], but self-reporting has limited representativeness given its low percentage of respondents [20]. Second, we only investigated a sample of individuals who play online lotteries. It would be interesting to replicate this work in other types of online gambling and will be the subject of future investigations; the results herein constitute only one part of the EDEIN research program and future EDEIN studies will address this limitation. Third, the choice of indicators for the trajectory analysis is questionable since other indicators exist; however, these choices had the advantage of combining information on gambling activity and on the potential for gambling problems. Fourth, the data set included only gambling activity related to online lotteries within a limited period. As a consequence, it is possible that some individuals in this study were also engaged in other types of gambling that were not captured by this study (other online or offline gambling activities) or that some individuals previously had an account on the same website which would bias the notion of describing “early” trajectories.

Despite these limitations, we must emphasize the strengths of this study. First, the naturalistic nature of gambling tracking data has high value for gambling research [13]. Second, we used a valid and robust method to explore trajectories whose utility has previously been demonstrated [49,50]. Trajectories as indicators for problem gambling have previously been defined using the slope of a linear regression that modeled wager size according to the sequence of active betting days during the initial month [23]. Such an approach may strongly reduce the information available because only positive or negative linear trends during the first month are observed and the following months, more complex trajectories, and other indicators of gambling are neglected. In our study, we used growth mixture modeling to model trajectories, which allowed us to capture the complexity of the evolution of gambling practice over time. Moreover, we performed this trajectory analysis on 4 indicators rather than only on wagers, which gave us access to the monetary variations over time as well as variations in gambling frequency, breadth of involvement, and chasing. Finally, the 6-month period allowed access to more in-depth trajectory information. Indeed, as shown in Figures 1 to 4, the variations of gambling activity during the first month did not necessarily reflect activity during the subsequent months. Third, Deng et al [20] emphasized that data aggregation often cannot capture chasing behavior, which requires fine-grained data. In our study, a chasing proxy was computed (based on the temporality between deposits and bets on a within-session basis) before data were aggregated. Because the chasing indicator best contributed to the identification of the problematic class, this definition may be useful for identifying individuals who are potentially at risk for gambling problems.

Conclusion and Perspectives

We demonstrated the importance of using longitudinal trajectory models rather than cross-sectional analysis to identify groups of individuals who are potentially at risk for gambling problems. High breadth of involvement and high and sustained chasing, combined with the use of trajectory models, may be used early on to identify individuals who are at risk of experiencing gambling problems. More specifically, the probabilities of individuals belonging to the 5 classes identified can serve to implement personalized preventive actions. Indeed, the effectiveness of self-regulation strategies from responsible gambling programs, such as the setting of gambling limits for oneself or voluntary self-exclusion, may be limited by the fact that they completely rely on individuals changing their own behavior. As argued by Haefeli et al [64], it is of high importance to detect excessive gambling as early as possible before individuals reach the late stages of gambling problems and experience too much damage to be receptive to preventive interventions. The ability to rapidly identify individuals who are at risk for future gambling problems may allow the implementation of targeted, personalized minimal interventions. Such interventions could rely on valuing help-seeking, or on informing at-risk individuals about existing tools to prevent excessive gambling and possible gambling-related damage. This is the ultimate goal of the EDEIN project [42] of which this study was a first step.

Acknowledgments

This research project has received funding from the Primary Prevention Call for Proposals that was issued by the French Institute for Public Health Research and the French National Cancer Institute in 2013 and funded by the French National Health Insurance Fund for Employees, the French Directorate General of Health, the Arc Foundation for Cancer Research, the French National Cancer Institute , the French National Institute for Prevention and Education in Health, the French National Institute of Health and Medical Research, the French Inter - Departmental Agency for the Fight against Drugs and Addictive Behaviors, and the French Social Security Scheme for Liberal Professionals.

We would like to thank *Française des Jeux* for giving us access to the gambling data and Playscan for allowing access to the risk assessment data.

This research was conducted on the initiative of and coordinated by the Department of Addiction and Psychiatry of the University Hospital of Nantes which sponsored this study. Neither the data providers (*Française des Jeux* and Playscan) nor the funders (French Institute for Public Health Research and its partners) had any influence on this work; they did not review or approve the manuscript. No constraints on publication were imposed.

Conflicts of Interest

GCB, ET, AS, YD, and MGB declare that the University Hospital of Nantes received funding from the gambling industry (Française des Jeux and Pari Mutuel Urbain) in the form of philanthropic sponsorship (donations that do not assign purpose of use). Scientific independence with respect to these gambling organizations is guaranteed, and the funding did not have any influence on this work. No conflicts of interest exist for BP.

References

1. Calado F, Griffiths MD. Problem gambling worldwide: an update and systematic review of empirical research (2000-2015). J Behav Addict 2016 Dec;5(4):592-613. [doi: 10.1556/2006.5.2016.073] [Medline: 27784180]
2. Costes JM, Eroukmanoff V, Richard JB, Tovar ML. Les jeux d'argent et de hasard en France en 2014. Les notes de l'Observatoire Des Jeux; n°6. Paris, France: Observatoire des Jeux; 2015 Apr 15. URL: https://www.economie.gouv.fr/files/note_6.pdf [accessed 2020-01-03]
3. Costes JM, Eroukmanoff V. Les pratiques de jeux d'argent sur Internet en France en 2017. Les notes de l'Observatoire des jeux; n°9. Paris, France: Observatoire des Jeux; 2018 Jul. URL: https://www.economie.gouv.fr/files/files/directions_services/observatoire-des-jeux/Note_ODJ_9.pdf [accessed 2020-01-03]
4. Tovar ML, Costes JM, Eroukmanoff V. Les jeux d'argent et de hasard sur Internet en France en 2012. Tendances; n°85. Saint-Denis-La-Plaine, France: Observatoire français des drogues et des toxicomanies (OFDT); 2013 Jun. URL: https://www.ofdt.fr/BDD/publications/docs/eftxmtt6.pdf [accessed 2020-01-03]
5. Gainsbury SM. Online gambling addiction: the relationship between internet gambling and disordered gambling. Curr Addict Rep 2015;2(2):185-193 [FREE Full text] [doi: 10.1007/s40429-015-0057-8] [Medline: 26500834]
6. Griffiths M. Internet gambling: issues, concerns, and recommendations. Cyberpsychol Behav 2003 Dec;6(6):557-568. [doi: 10.1089/109493103322725333] [Medline: 14756922]
7. McCormack A, Griffiths MD. A scoping study of the structural and situational characteristics of Internet gambling. International Journal of Cyber Behavior, Psychology and Learning 2013;3(1):29-49. [doi: 10.4018/ijcbpl.2013010104]
8. Effertz T, Bischof A, Rumpf HJ, Meyer C, John U. The effect of online gambling on gambling problems and resulting economic health costs in Germany. Eur J Health Econ 2018 Sep;19(7):967-978. [doi: 10.1007/s10198-017-0945-z] [Medline: 29362900]

9. Griffiths M, Wardle H, Orford J, Sproston K, Erens B. Internet gambling, health, smoking and alcohol use: findings from the 2007 british gambling prevalence survey. Int J Ment Health Addiction 2009 Sep 15;9(1):1-11. [doi: 10.1007/s11469-009-9246-9]
10. Kairouz S, Paradis C, Nadeau L. Are online gamblers more at risk than offline gamblers? Cyberpsychol Behav Soc Netw 2012 Mar;15(3):175-180. [doi: 10.1089/cyber.2011.0260] [Medline: 22132932]
11. Papineau E, Lacroix G, Sévigny S, Biron JF, Corneau-Tremblay N, Lemétayer F. Assessing the differential impacts of online, mixed, and offline gambling. International Gambling Studies 2017 Oct 03;18(1):69-91. [doi: 10.1080/14459795.2017.1378362]
12. Wood RT, Williams RJ. Internet Gambling: Prevalence, Patterns, Problems, and Policy Options. Guelph, Ontario: Final Report prepared for the Ontario Problem Gambling Research Centre; 2009 Jan 05. URL: https://opus.uleth.ca/bitstream/handle/10133/693/2009-InternetPPPP-OPGRC.pdf [accessed 2020-01-03]
13. Gainsbury S. Player account-based gambling: potentials for behaviour-based research methodologies. International Gambling Studies 2011 Aug;11(2):153-171. [doi: 10.1080/14459795.2011.571217]
14. Braverman J, Tom MA, Shaffer HJ. Accuracy of self-reported versus actual online gambling wins and losses. Psychol Assess 2014 Sep;26(3):865-877. [doi: 10.1037/a0036428] [Medline: 24708074]
15. Baumeister RF, Vohs KD, Funder DC. Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? Perspect Psychol Sci 2007 Dec;2(4):396-403. [doi: 10.1111/j.1745-6916.2007.00051.x] [Medline: 26151975]
16. Garber MC, Nau DP, Erickson SR, Aikens JE, Lawrence JB. The concordance of self-report with other measures of medication adherence: a summary of the literature. Med Care 2004 Jul;42(7):649-652. [Medline: 15213489]
17. Rundle‐Thiele S. Bridging the gap between claimed and actual behaviour: the role of observational research. Qualitative Mrkt Res 2009 Jun 12;12(3):295-306. [doi: 10.1108/13522750910963818]
18. Editorial. Science has a gambling problem. Nature 2018 Jan 25;553(7689):379. [doi: 10.1038/d41586-018-01051-z] [Medline: 29368723]
19. Ladouceur R, Shaffer P, Blaszczynski A, Shaffer HJ. Responsible gambling: a synthesis of the empirical evidence. Addiction Research & Theory 2016 Dec 14;25(3):225-235. [doi: 10.1080/16066359.2016.1245294]
20. Deng X, Lesch T, Clark L. Applying data science to behavioral analysis of online gambling. Curr Addict Rep 2019 Jul 15;6(3):159-164. [doi: 10.1007/s40429-019-00269-9]
21. Adami N, Benini S, Boschetti A, Canini L, Maione F, Temporin M. Markers of unsustainable gambling for early detection of at-risk online gamblers. International Gambling Studies 2013 Aug;13(2):188-204. [doi: 10.1080/14459795.2012.754919]
22. Braverman J, LaPlante DA, Nelson SE, Shaffer HJ. Using cross-game behavioral markers for early identification of high-risk internet gamblers. Psychol Addict Behav 2013 Sep;27(3):868-877. [doi: 10.1037/a0032818] [Medline: 24059836]
23. Braverman J, Shaffer HJ. How do gamblers start gambling: identifying behavioural markers for high-risk internet gambling. Eur J Public Health 2012 Apr;22(2):273-278 [FREE Full text] [doi: 10.1093/eurpub/ckp232] [Medline: 20110273]
24. Broda A, LaPlante DA, Nelson SE, LaBrie RA, Bosworth LB, Shaffer HJ. Virtual harm reduction efforts for Internet gambling: effects of deposit limits on actual Internet sports gambling behavior. Harm Reduct J 2008 Aug 06;5(27):1-9 [FREE Full text] [doi: 10.1186/1477-7517-5-27] [Medline: 18684323]
25. Gray HM, LaPlante DA, Shaffer HJ. Behavioral characteristics of internet gamblers who trigger corporate responsible gambling interventions. Psychol Addict Behav 2012 Sep;26(3):527-535. [doi: 10.1037/a0028545] [Medline: 22686960]
26. Labrie RA, Kaplan SA, Laplante DA, Nelson SE, Shaffer HJ. Inside the virtual casino: a prospective longitudinal study of actual Internet casino gambling. Eur J Public Health 2008 Aug;18(4):410-416. [doi: 10.1093/eurpub/ckn021] [Medline: 18434382]
27. LaBrie RA, LaPlante DA, Nelson SE, Schumann A, Shaffer HJ. Assessing the playing field: a prospective longitudinal study of internet sports gambling behavior. J Gambl Stud 2007 Sep;23(3):347-362. [doi: 10.1007/s10899-007-9067-3] [Medline: 17574522]
28. LaBrie R, Shaffer HJ. Identifying behavioral markers of disordered Internet sports gambling. Addiction Research & Theory 2010 Sep 07;19(1):56-65. [doi: 10.3109/16066359.2010.512106]
29. LaPlante DA, Kleschinsky JH, LaBrie RA, Nelson SE, Shaffer HJ. Sitting at the virtual poker table: a prospective epidemiological study of actual Internet poker gambling behavior. Computers in Human Behavior 2009 May;25(3):711-717. [doi: 10.1016/j.chb.2008.12.027]
30. LaPlante DA, Nelson SE, Gray HM. Breadth and depth involvement: understanding internet gambling involvement and its relationship to gambling problems. Psychol Addict Behav 2014 Jun;28(2):396-403. [doi: 10.1037/a0033810] [Medline: 23915365]
31. LaPlante DA, Schumann A, LaBrie RA, Shaffer HJ. Population trends in internet sports gambling. Comput Human Behav 2008 Sep;24(5):2399-2414. [doi: 10.1016/j.chb.2008.02.015]
32. Nelson SE, LaPlante DA, Peller AJ, Schumann A, LaBrie RA, Shaffer HJ. Real limits in the virtual world: self-limiting behavior of Internet gamblers. J Gambl Stud 2008 Dec;24(4):463-477. [doi: 10.1007/s10899-008-9106-8] [Medline: 18696219]

33. Philander KS. Identifying high-risk online gamblers: a comparison of data mining procedures. Int Gambl Stud 2013 Oct 23;14(1):53-63. [doi: 10.1080/14459795.2013.841721]
34. Dragicevic S, Tsogas G, Kudic A. Analysis of casino online gambling data in relation to behavioural risk markers for high-risk gambling and player protection. Int Gambl Stud 2011 Nov 17;11(3):377-391. [doi: 10.1080/14459795.2011.629204]
35. Luquiens A, Tanguy ML, Benyamina A, Lagadec M, Aubin HJ, Reynaud M. Tracking online poker problem gamblers with player account-based gambling data only. Int J Methods Psychiatr Res 2016 Dec;25(4):333-342. [doi: 10.1002/mpr.1510] [Medline: 27198992]
36. Perrot B, Hardouin JB, Grall-Bronnec M, Challet-Bouju G. Typology of online lotteries and scratch games gamblers' behaviours: a multilevel latent class cluster analysis applied to player account-based gambling data. Int J Methods Psychiatr Res 2018 Dec;27(4):e1746. [doi: 10.1002/mpr.1746] [Medline: 30338605]
37. Costes JM, Kairouz S, Monson E, Eroukmanoff V. Where lies the harm in lottery gambling? a portrait of gambling practices and associated problems. J Gambl Stud 2018 Dec;34(4):1293-1311. [doi: 10.1007/s10899-018-9761-3] [Medline: 29536292]
38. Breen RB, Zuckerman M. `Chasing' in gambling behavior: personality and cognitive determinants. Personality and Individual Differences 1999 Dec;27(6):1097-1111. [doi: 10.1016/s0191-8869(99)00052-5]
39. Lesieur HR. The compulsive gambler's spiral of options and involvement. Psychiatry 1979 Feb;42(1):79-87. [doi: 10.1080/00332747.1979.11024008] [Medline: 760136]
40. Toce-Gerstein M, Gerstein DR, Volberg RA. A hierarchy of gambling disorders in the community. Addiction 2003 Dec;98(12):1661-1672. [doi: 10.1111/j.1360-0443.2003.00545.x] [Medline: 14651495]
41. Ciccarelli M, Cosenza M, D'Olimpio F, Griffiths MD, Nigro G. An experimental investigation of the role of delay discounting and craving in gambling chasing behavior. Addictive Behaviors 2019 Jun;93:250-256. [doi: 10.1016/j.addbeh.2019.02.002]
42. Perrot B, Hardouin JB, Costes JM, Caillon J, Grall-Bronnec M, Challet-Bouju G. Study protocol for a transversal study to develop a screening model for excessive gambling behaviours on a representative sample of users of French authorised gambling websites. BMJ Open 2017 May 17;7(5):e014600 [FREE Full text] [doi: 10.1136/bmjopen-2016-014600] [Medline: 28515192]
43. LaPlante DA, Nelson SE, LaBrie RA, Shaffer HJ. Disordered gambling, type of gambling and gambling involvement in the British Gambling Prevalence Survey 2007. Eur J Public Health 2011 Aug;21(4):532-537 [FREE Full text] [doi: 10.1093/eurpub/ckp177] [Medline: 19892851]
44. Baggio S, Dupuis M, Berchtold A, Spilka S, Simon O, Studer J. Is gambling involvement a confounding variable for the relationship between internet gambling and gambling problem severity? Comput Human Behav 2017 Jun;71:148-152. [doi: 10.1016/j.chb.2017.02.004]
45. Forsström D, Jansson-Fröjmark M, Hesser H, Carlbring P. Experiences of Playscan: interviews with users of a responsible gambling tool. Internet Interv 2017 Jun;8:53-62 [FREE Full text] [doi: 10.1016/j.invent.2017.03.003] [Medline: 30135829]
46. Blaszczynski A, Nower L. A pathways model of problem and pathological gambling. Addiction 2002 May;97(5):487-499. [Medline: 12033650]
47. Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics 1999 Jun;55(2):463-469. [doi: 10.1111/j.0006-341x.1999.00463.x] [Medline: 11318201]
48. Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association 1996 Mar;91(433):217-221. [doi: 10.1080/01621459.1996.10476679]
49. Brédart A, Merdy O, Sigal-Zafrani B, Fiszer C, Dolbeault S, Hardouin JB. Identifying trajectory clusters in breast cancer survivors' supportive care needs, psychosocial difficulties, and resources from the completion of primary treatment to 8 months later. Support Care Cancer 2016 Jan;24(1):357-366. [doi: 10.1007/s00520-015-2799-1] [Medline: 26076962]
50. Montourcy M, Hardouin JB, Caillon J, Leboucher J, Rousselet M, Grall-Bronnec M, et al. Typology of patients with behavioral addictions or eating disorders during a one-year period of care: exploring similarities of trajectory using growth mixture modeling coupled with latent class analysis. PLoS One 2018;13(11):e0207398. [doi: 10.1371/journal.pone.0207398] [Medline: 30427931]
51. Muthén LK, Muthén BO. Mplus user‘s guide. Mplus. URL: https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf [accessed 2020-02-29]
52. Vermunt JK, Magidson J. Technical guide for Latent GOLD 5.1: basic, advanced, andsyntax. Statistical Innovations. URL: https://www.statisticalinnovations.com/wp-content/uploads/LGtecnical.pdf [accessed 2020-02-29]
53. Challet-Bouju G, Hardouin JB, Renard N, Legauffre C, Valleur M, Magalon D, et al. A gamblers clustering based on their favorite gambling activity. J Gambl Stud 2015 Dec;31(4):1767-1788. [doi: 10.1007/s10899-014-9496-8] [Medline: 25192752]
54. Temcheff CE, Paskus TS, Potenza MN, Derevensky JL. Which diagnostic criteria are most useful in discriminating between social gamblers and individuals with gambling problems? an examination of DSM-IV and DSM-5 criteria. J Gambl Stud 2016 Feb 5;32(3):957-968. [doi: 10.1007/s10899-015-9591-5]
55. Percy C, França M, Dragičević S, d'Avila Garcez A. Predicting online gambling self-exclusion: an analysis of the performance of supervised machine learning models. Int Gambl Stud 2016 Apr 10;16(2):193-210. [doi: 10.1080/14459795.2016.1151913]
56. Motka F, Grüne B, Sleczka P, Braun B, Örnberg JC, Kraus L. Who uses self-exclusion to regulate problem gambling? A systematic literature review. Journal of Behavioral Addictions 2018 Dec;7(4):903-916. [doi: 10.1556/2006.7.2018.96]

57. Caillon J, Grall-Bronnec M, Perrot B, Leboucher J, Donnio Y, Romo L, et al. Effectiveness of at-risk gamblers' temporary self-exclusion from internet gambling sites. J Gambl Stud 2019 Jun;35(2):601-615. [doi: 10.1007/s10899-018-9782-y] [Medline: 29974308]
58. McCormack A, Shorter GW, Griffiths MD. An empirical study of gender differences in online gambling. J Gambl Stud 2014 Mar;30(1):71-88. [doi: 10.1007/s10899-012-9341-x] [Medline: 23097131]
59. Auer M, Hopfgartner N, Griffiths MD. An empirical study of the effect of voluntary limit-setting on gamblers' loyalty using behavioural tracking data. Int J Ment Health Addiction 2019 Apr 26:1-12. [doi: 10.1007/s11469-019-00084-3]
60. Hing N, Cherney L, Gainsbury SM, Lubman DI, Wood RT, Blaszczynski A. Maintaining and losing control during internet gambling: a qualitative study of gamblers' experiences. New Media & Society 2014 Jan 27;17(7):1075-1095 [FREE Full text] [doi: 10.1177/1461444814521140]
61. Williams RJ, Volberg RA, Stevens RMG. The population prevalence of problem gambling: methodological influences, standardized rates, jurisdictional differences, and worldwide trends.: Report prepared for the Ontario Problem Gambling Research Centre & the Ontario Ministry of Health and Long Term Care; 2012 May 08. URL: https://pdfs.semanticscholar.org/e657/fb4399befa9a5cb68337d08b5675d04306bd.pdf?_ga=2.252403977.1281152942.1590565077-607814672.1580121986 [accessed 2020-01-03]
62. Corney R, Davis J. The attractions and risks of Internet gambling for women: a qualitative study. Journal of Gambling Issues 2010 Jul 01(24):121-139. [doi: 10.4309/jgi.2010.24.8]
63. Bruneau M, Guillou-Landreat M, Challet-Bouju G, Leboucher J, Sauvaget A, Caillon J, et al. Gambling problems in women: French specificities. In: Bowden-Jones H, Prever F, editors. Gambling Disorders In Women: An International Female Perspective On Treatment And Research. Abingdon (GB): Routledge; Jun 2017.
64. Haefeli J, Lischer S, Schwarz J. Early detection items and responsible gambling features for online gambling. Int Gambl Stud 2011 Aug 17;11(3):273-288. [doi: 10.1080/14459795.2011.604643]

Abbreviations

EDEIN: Etude de Dépistage des comportements Excessifs de jeu sur Internet (Screening for Excessive Gambling Behaviors on the Internet)

Edited by G Eysenbach; submitted 03.01.20; peer-reviewed by M Bellringer, D Hodgins; comments to author 13.02.20; revised version received 26.02.20; accepted 29.02.20; published 12.08.20.

Please cite as:
Challet-Bouju G, Hardouin JB, Thiabaud E, Saillard A, Donnio Y, Grall-Bronnec M, Perrot B
Modeling Early Gambling Behavior Using Indicators from Online Lottery Gambling Tracking Data: Longitudinal Analysis
J Med Internet Res 2020;22(8):e17675
URL: http://www.jmir.org/2020/8/e17675/
doi:10.2196/17675
PMID:32254041

Original Paper

Machine Learning Classifiers for Twitter Surveillance of Vaping: Comparative Machine Learning Study

Shyam Visweswaran[1,2], MD, PhD; Jason B Colditz[3], MEd; Patrick O'Halloran[1], MS; Na-Rae Han[4], PhD; Sanya B Taneja[2], MS; Joel Welling[5], PhD; Kar-Hai Chu[3], PhD; Jaime E Sidani[3], PhD, MPH; Brian A Primack[6], PhD, MD

[1]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, United States

[2]Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA, United States

[3]School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[4]Department of Linguistics, University of Pittsburgh, Pittsburgh, PA, United States

[5]Pittsburgh Supercomputing Center, Carnegie Mellon University, Pittsburgh, PA, United States

[6]College of Education and Health Professions, University of Arkansas, Fayetteville, AR, United States

Corresponding Author:
Shyam Visweswaran, MD, PhD
Department of Biomedical Informatics
University of Pittsburgh
The Offices at Baum
5607 Baum Blvd, Suite 523
Pittsburgh, PA, 15206
United States
Phone: 1 (412) 648 7119
Email: shv3@pitt.edu

Abstract

Background: Twitter presents a valuable and relevant social media platform to study the prevalence of information and sentiment on vaping that may be useful for public health surveillance. Machine learning classifiers that identify vaping-relevant tweets and characterize sentiments in them can underpin a Twitter-based vaping surveillance system. Compared with traditional machine learning classifiers that are reliant on annotations that are expensive to obtain, deep learning classifiers offer the advantage of requiring fewer annotated tweets by leveraging the large numbers of readily available unannotated tweets.

Objective: This study aims to derive and evaluate traditional and deep learning classifiers that can identify tweets relevant to vaping, tweets of a commercial nature, and tweets with provape sentiments.

Methods: We continuously collected tweets that matched vaping-related keywords over 2 months from August 2018 to October 2018. From this data set of tweets, a set of 4000 tweets was selected, and each tweet was manually annotated for relevance (vape relevant or not), commercial nature (commercial or not), and sentiment (provape or not). Using the annotated data, we derived traditional classifiers that included logistic regression, random forest, linear support vector machine, and multinomial naive Bayes. In addition, using the annotated data set and a larger unannotated data set of tweets, we derived deep learning classifiers that included a convolutional neural network (CNN), long short-term memory (LSTM) network, LSTM-CNN network, and bidirectional LSTM (BiLSTM) network. The unannotated tweet data were used to derive word vectors that deep learning classifiers can leverage to improve performance.

Results: LSTM-CNN performed the best with the highest area under the receiver operating characteristic curve (AUC) of 0.96 (95% CI 0.93-0.98) for relevance, all deep learning classifiers including LSTM-CNN performed better than the traditional classifiers with an AUC of 0.99 (95% CI 0.98-0.99) for distinguishing commercial from noncommercial tweets, and BiLSTM performed the best with an AUC of 0.83 (95% CI 0.78-0.89) for provape sentiment. Overall, LSTM-CNN performed the best across all 3 classification tasks.

Conclusions: We derived and evaluated traditional machine learning and deep learning classifiers to identify vaping-related relevant, commercial, and provape tweets. Overall, deep learning classifiers such as LSTM-CNN had superior performance and had the added advantage of requiring no preprocessing. The performance of these classifiers supports the development of a vaping surveillance system.

(J Med Internet Res 2020;22(8):e17478) doi:10.2196/17478

KEYWORDS

vaping; social media; infodemiology; infoveillance; machine learning; deep learning

Introduction

Background

Machine learning methods provide a valuable framework for systematic and automated processing and analysis of data on social media platforms such as Twitter for developing surveillance systems with application to public health. The continuous generation of an enormous amount of content by a vast number of users allows for efficient real-time monitoring of sources of information and user sentiment if it can be automated. Furthermore, such monitoring can lead to the discovery of emergent patterns of information flow and changes in sentiments that may occur in response to public health and policy interventions. In this study, we derived and evaluated traditional machine learning and deep learning classifiers that can be used to build a Twitter-based surveillance system to identify and monitor vaping-related content and sentiments.

Vaping and Public Health

Vaping is the inhalation of aerosols that often contain nicotine combined with flavorings where the aerosols are delivered through electronic delivery systems known as electronic cigarettes (e-cigarettes) or electronic vaporizers. Evidence suggests that vaping is safer than smoking tobacco and can help with successful smoking cessation [1]. However, emerging research indicates that vaping may cause cardiovascular and respiratory diseases and may pose health hazards from secondhand aerosol exposure [2]. More recently, vaping has been associated with e-cigarette or vaping product use–associated lung injury, which has caused hospitalization and even death [3,4]. There is a rising concern that vaping increases addiction among nonsmokers, especially adolescents [5], and many are unaware of the addictive potential until after they become nicotine dependent [6]. Thus, there is a strong need to measure and understand the risks, sentiments, and behavior related to vaping.

Surveillance Using Twitter

Twitter is a popular social media platform that is widely used by adolescents, young adults, and racial and ethnic minorities, all of whom are disproportionately affected by vaping [7-9]. Communication on Twitter is by short succinct messages, called tweets, which are limited to 280 characters. Twitter is an open platform that enables users to see information and messages from other public users without special permission. This results in high potential exposure to each tweet, which enables systematic assessment by investigators. Furthermore, tweets heavily use hashtags (eg, #vapelife) as searchable text, which allows users to click on a linked word or phrase and navigate to other mentions of it [10]. These factors make Twitter a relevant, valuable, and feasible social media platform to study.

Infoveillance is the application of surveillance methods to internet-related and other electronic content to inform public health and public policy. Traditional surveys around attitudes and beliefs are too slow to optimally capture rapid changes. Infoveillance methods that use web-based data streams have proven to be more effective for several areas of public health. Investigators have used Twitter data for the infoveillance of topics such as pharmacovigilance, vaccine information, and tracking health conditions [11-13]. For example, such data have been useful in characterizing outbreaks of food-related illness and influenza, factors surrounding prescription drug abuse [14], adverse drug events [15], sentiment toward the use of tobacco [16,17], and use of alcohol [18].

Objective

Our immediate objective was to derive and evaluate machine learning classifiers that can form the basis of a Twitter-based surveillance system that is focused on vaping-related tweets. Our ultimate goal is to use a surveillance system to assess key factors such as sentiment, marketing, procurement, health effects, and policy that will provide unique perspectives related to vaping. Furthermore, we plan to characterize changes over time in the volume of messaging related to vaping and other vaping-related characteristics of interest [19,20]. Leveraging Twitter as a complement to traditional surveillance will allow for real-time identification of changes that can be used by public health practitioners. For example, when positive sentiment toward vaping rises, practitioners may be able to determine reasons for this and respond accordingly. Similarly, when there is a notable spike in misinformation about vaping and health effects, they will be able to act immediately to correct this information. As a step toward the development of a Twitter-based vaping surveillance system, we derived machine learning classifiers to automatically identify tweets that are vaping-related, are noncommercial, and express provape sentiments. Using a data set of manually annotated tweets and a larger data set of unannotated tweets, we derived and evaluated traditional machine learning and deep learning classifiers.

Related Work

Natural language processing, classification, and sentiment analysis of Twitter data are more taxing than other kinds of text because of the limited length of the tweets. As tweets are limited to 280 characters and the language used is informal, the messages are interspersed with abbreviations, slang, Twitter-specific terms such as usernames and hashtags, and URLs.

Several investigators have derived classifiers using Twitter data in the context of vaping. For example, Han and Kavuluru [21] implemented support vector machines, logistic regression, and convolutional neural networks to identify marketing and nonmarketing e-cigarette tweets. Myslin et al [17] and Cole-Lewis et al [22] annotated tobacco-related tweets and derived several machine learning classifiers to predict sentiment. Huang et al [23] analyzed tweets using classifiers and found that tweets related to e-cigarettes were about 90% commercial

and about 10% mentioned smoking cessation. Resende and Culotta [24] derived a sentiment classifier for e-cigarette–related tweets that identified positive and negative tweets with 96% and 70% precision, respectively.

Compared with prior work, the main contributions of this paper are (1) exploration of a large range of classifiers, including deep learning classifiers; and (2) analysis of highly relevant features in classifiers using an algorithm that provides a unified approach to explain the output of any classifier.

Methods

Data Collection

Primary data were collected from the Twitter application programming interface using the open-source, real-time infoveillance of Twitter health messages (RITHM) software [19]. RITHM allows for the real-time collection of all publicly available tweets matching a specified set of keywords. We identified and collected all tweets that matched one or more keywords that are indicative of vaping-related tweets. The keywords that we used for data collection included *vape*, *vapes*, *vaper*, *vapers*, *vaping*, *juul*, *juuls*, and *juuling*. The vaping-related keywords are based on previous Twitter research [10], and, in particular, we included keywords to identify the highly popular JUUL e-cigarette [6].

Data Set for Annotation

We continuously collected all publicly available tweets that matched vaping-related keywords over 2 months from August 17, 2018, to October 19, 2018. This resulted in a data set of 1,892,722 tweets. From this data set, we removed *retweets* (rebroadcasted messages without original content), and from the remaining original 810,600 tweets, we randomly selected a subset of 4000 tweets for manual double coding and adjudication. The removal of retweets and the random selection ensured that the tweet content was lexically diverse and sufficiently representative of tweets related to vaping. This particular period was chosen as it also included salient health policy events related to vaping. In particular, the US Food and Drug Administration (FDA) sent warning letters to retailers and manufacturers (September 12, 2018) and seized documents from JUUL headquarters (October 05, 2018). In previous studies, data sets of 4000 to 7000 tweets have been adequate for the derivation of classifiers [17,25,26].

Unannotated Data Set for Deriving Word Vectors

Word vectors, also known as word embeddings, are derived from a large data set of text to capture semantic and syntactic similarity and context of each word as a vector of real numbers. Word vectors have become popular because they can improve the performance of deep learning classifiers and can reduce the volume of annotations that are needed. Word vectors have the advantage that they do not require annotations; instead, they leverage a large amount of unannotated data.

We continuously collected all publicly available tweets that matched vaping-related keywords over 7 months from January 01, 2018 to July 31, 2018. This resulted in a data set of 4,078,343 tweets, and from this data set, we removed retweets to obtain a set of 1,899,851 original tweets. We used this set to derive word vectors for deep learning. The period of data selected for word vectors represents 7 months of continuous data collection and provided a sufficiently large set of tweets for deriving word vectors and simultaneously ensuring that relevant context from the tweets, in terms of language and topical diversity, is captured in the word vectors. The period of data selected for word vectors was before the period of data selected for annotations, with no overlap, as a part of the annotated set was used for the evaluation of the classifiers.

Annotation

We developed a three-level hierarchical annotation schema, as shown in Figure 1. Descriptions of the labels used for annotation are provided in Table 1. The annotation procedure consisted of first annotating a tweet as vape relevant or not based on the content. A relevant tweet was further annotated as commercial or noncommercial, and a noncommercial tweet was further annotated for provape or not provape sentiments. A similar three-level hierarchical annotation schema has been used for annotating vaccination-related tweets. At the first level, a tweet is annotated as relevant or not; at the next level, only a relevant tweet is annotated as positive, negative, or neutral; and at the final level, only a negative tweet is annotated based on safety, efficacy, cost, etc [25,26]. A hierarchical annotation schema has the advantage that all tweets need not be annotated on all possible levels, thus allowing for a reduction in annotation effort. For example, nonrelevant tweets need not be annotated further, and relevant and commercial tweets need not be annotated further.

Trained annotators independently annotated 4000 tweets in batches of 100 to 200 and adjudicated annotation disagreements in the presence of a supervising investigator. Annotators considered tweet content that included both primary and secondary text (ie, quoted tweets within primary tweets). Furthermore, annotators had access to Twitter's native platform, where they could review the context of potentially confusing content. Cohen κ coefficient was used to assess interrater agreement [27] before adjudication and at regular intervals throughout the process. Initial κ coefficients were relatively modest (eg, κ=0.54 for relevance), but improved as annotators gained familiarity with the data and the domain. The κ coefficients for the final round of annotation (n=100) were 0.71 for relevance, 0.89 for commercial, and 0.70 for provape. Fully adjudicated annotations and tweet content including metadata were used for machine learning.

Figure 1. A hierarchical annotation scheme for vaping-related tweets.

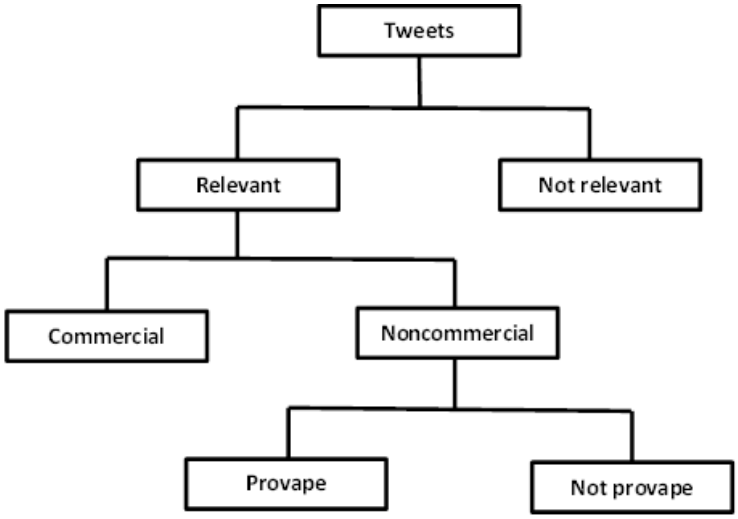


Table 1. Descriptions of labels used for annotating vaping-related tweets.

Labels	Descriptions
Relevant	Is the tweet in English and related to the vaping topic at hand (eg, vape use or users, vaping devices, or products)?
Not relevant	Tweets categorized as not relevant were typically in non-English or had referenced vaping cannabis products specifically, such as: • "Teens are smoking, vaping and eating cannabis" • "What if I vape weed?"
Commercial	Is the tweet selling, marketing, or advertising vaping products?
Noncommercial	Includes tweets that demonstrate favorability toward a product but do not directly advocate for purchasing it.
Provape	Is vaping associated with positive emotions or contexts? Such as: • The tweet author is currently using, has recently used, or intends to use a vape product. • The tweet author indicates acceptance of others' vaping or favorability toward others' positive perspectives of vaping. • The tweet author mentions vaping in association with other positive aspects of society or popular culture (eg, partying, sexuality, popularity, and attractiveness).
Not provape	Includes tweets that are antivape, neutral or fact based, or without subjective judgment about positive or acceptable aspects of vaping.

Machine Learning

In this section, we describe the steps in machine learning that consist of preprocessing, derivation of features, and training of classifiers.

Preprocessing and Vector Representation for Traditional Classifiers

Twitter data consist of tweet metadata and tweet content. Metadata includes information related to the user's profile (such as location, number of followers, number of friends, and tweeting frequency), information related to a tweet's status (such as the location of the tweet), media object contained in the tweet (such as audio, video, and image), and if the tweet was in reply to another tweet. As tweets are restricted to 280 characters, their content has, in addition to the standard text, abbreviations, usernames (that are annotated with the @), hashtags (topic tags annotated with the #), Unicode characters, URLs (typically shortened pseudorandom short URLs), and emojis (icons used to express an idea or emotion). Before preprocessing, we replaced usernames, hashtags, Unicode characters, and URLs with the textual placeholders _mention_, _hashtag_, _unicode_, and _url_, respectively. We also translated emojis into textual descriptions for better interpretability. This standardized text representation of tweets ensured that the preprocessing pipeline needed to handle only text.

The preprocessing pipeline consisted of 10 steps, including removal of textual placeholders (for usernames, hashtags, Unicode characters, and URLs), removal of textual descriptions of emojis, expansion of negations, removal of punctuation and digits, negation marking, normalization, stemming, removal of stopwords, and conversion to lowercase (Table 2).

After preprocessing, we created 2 types of tweet representations that are useful for machine learning. In the first representation, called the vector count representation, we identify unique words in the tweet data set and represent each tweet with a vector of numbers, where a number denotes the frequency (count) of the occurrence of a unique word in the tweet. Thus, each tweet is represented by a vector that contains as many counts as the number of unique words. We also investigated an alternative

vector representation called frequency-inverse document frequency (TF-IDF) where the number assigned to a word in the vector depends not only on its frequency in a tweet but also on its frequency in the entire data set. In this representation, words that occur in the majority of the tweets are considered to be of lower importance than words that occur more rarely. As preliminary results did not demonstrate improved performance with the TF-IDF representation, we did not perform extensive experiments with this representation.

Rather than applying the same set of preprocessing steps to every classifier, we searched all possible combinations of the 10 preprocessing steps for each classifier and identified the optimal set of preprocessing steps that gave the best classifier performance (Table 2).

Table 2. Description of preprocessing steps and options used in traditional classifiers.

Preprocessing steps	Descriptions	Options[a]
placeholder_remove	Remove textual placeholders such as _mention_, _hashtag_, _unicode_, and _url_	True, false
emoji_remove	Remove textual descriptions that denote emojis	True, false
negation_expand	Expand negative contractions, for example, “don’t” is expanded to “do not” and “can’t” is expanded to “cannot”	True, false
punctuation_remove	Remove all punctuation symbols	True, false
digits_remove	Remove all numeric digits (0-9)	True, false
negation_mark	Mark words that occur between a negation trigger and a punctuation mark with the NEG prefix [28]	True, false
normalize	Reduce to 2 characters all consecutive characters that appear more than twice, for example, “happppy” is reduced to “happy”	True, false
stemming	Reduce inflection in words (eg, troubled, troubles) to their root form (eg, trouble) using the Porter Stemmer [29]	True, false
stopwords_remove	Remove common words such as “the,” “a,” “on,” “is,” and “all” that are listed in the Natural Language Toolkit English stop words list [30]	True, false
lowercase	Change the case of all characters to lowercase	True, false

[a]If the option for a step is set to *true*, the corresponding preprocessing step will be applied in the preprocessing pipeline; if the option is set to *false*, the corresponding preprocessing step will be skipped in the pipeline.

Preprocessing and Vector Representation for Deep Learning Classifiers

For the deep learning classifiers, we used 2 alternative preprocessing methods: (1) a fixed preprocessing pipeline and (2) no preprocessing. The fixed preprocessing pipeline consisted of the following 5 steps (out of the possible 10 steps listed in Table 2): removal of textual placeholders, expansion of negations, removal of punctuation and digits, and conversion to lowercase. In contrast to vector count representation, which is used in traditional classifiers where a tweet is denoted by a vector of counts, in the deep learning classifiers, each word in a tweet is denoted by a word vector as described next, and each tweet is denoted by a vector of word vectors.

Word Vectors

Word vectors are derived from large unannotated tweet data (or other types of text data) and are increasingly used in deep learning classifiers. A word vector represents a word (not an entire tweet as in vector count representation) as a vector of numbers such that 2 words are considered to be similar in meaning if their vectors are close to each other mathematically. Word vectors capture the meaning and usage of words and are derived from patterns of how words co-occur in a large data set of tweets.

We investigated the performance of word vectors from 2 types of tweet data. First, we used word vectors that are derived from a large data set of tweets of all kinds; we call these vectors general or nondomain-specific word vectors. For general word vectors, we downloaded the 200-dimension Global Vectors for Word Representation (GloVe) word vectors. The GloVe vectors were derived from 2 billion tweets of all kinds, and each word was represented by a vector of size 200 [31]. Second, we used word vectors from a large data set of vaping-related tweets; we call these vectors vaping-related word vectors. We created vaping-related word vectors from a data set of tweets that were collected over 7 months from January 01, 2018 to July 31, 2018 using the vaping-related keywords. This data set contained 1,899,851 original tweets, and we used the Word2Vec algorithm [32] to derive 300-dimension word vectors (additional settings for the Word2Vec algorithm included a window size of 2 and 30 epochs).

Machine Learning Methods

We derived and evaluated 2 families of classifiers. The traditional classifiers included logistic regression (LR), random forest (RF), linear support vector machine (SVM), and multinomial naive Bayes (NB), and we used the implementations of these classifiers in scikit-learn version 0.23.1 [33]. The deep learning classifiers included convolutional neural network (CNN), long short-term memory (LSTM) network, combined LSTM and CNN (LSTM-CNN), and bidirectional LSTM (BiLSTM) network, and we used the implementations of these classifiers in Keras version 2.2.4 [34].

In contrast to traditional classifiers, CNNs automatically select words in tweets that are relevant. The LSTM network is a type of neural network that captures patterns of words in tweets. Conventional LSTM networks capture patterns in a single

direction, from left to right, whereas BiLSTM networks capture patterns in both directions, from left to right and from right to left. Both LSTM and BiLSTM have demonstrated good performance on social media data [35,36], and compared with CNNs, they can handle the variable lengths of tweets. The LSTM-CNN networks combine the advantages of CNNs and LSTM networks.

We derived and evaluated separate classifiers for 3 different tasks, that is, to identify which tweets are relevant, are noncommercial, and contain provape sentiment. For these tasks, the 3 binary targets and their corresponding values are (1) relevance: relevant (positive value) versus nonrelevant (negative value), (2) commercial: commercial (positive value) versus noncommercial (negative value), and (3) sentiment: provape (positive value) versus not provape (negative value).

Experimental Methods

From the annotated data set of 4000 tweets, we created 3 data sets to predict relevance, commercial, and sentiment that contained 4000, 3011, and 2175 tweets, respectively (Table 3). Each data set was randomly split into training and test sets (90:10 splits) such that the sets contained the same proportion of positive targets. A total of 3600, 2709, and 1957 tweets were used in the training data sets to derive relevance, commercial, and sentiment classifiers, respectively (Table 3). We used the training set to derive the best classifier (including the selection of hyperparameters if needed) for each type of classifier. The test data sets that were used to evaluate the relevance, commercial, and sentiment classifiers included 400, 302, and 218 tweets, respectively (Table 3).

Table 4 shows the traditional classifiers with parameter settings that we used in our experiments, and Table 5 shows the parameter settings of the deep learning classifiers that we used in our experiments.

Table 3. Description of training and test data sets.

Targets	Total number of tweets, n (%)	Number of tweets with positive target, n (%)	Number of tweets with negative target, n (%)
Relevance		Relevant	Nonrelevant
	• Total: 4000 (100) • Training: 3600 (100) • Test: 400 (100)	• Total: 3011 (75.28) • Training: 2709 (75.25) • Test: 302 (75.5)	• Total: 989 (24.72) • Training: 891 (24.75) • Test: 98 (24.5)
Commercial		Noncommercial	Commercial
	• Total: 3011 (100) • Training: 2709 (100) • Test: 302 (100)	• Total: 2175 (72.24) • Training: 1957 (72.24) • Test: 218 (72.2)	• Total: 836 (27.76) • Training: 752 (27.86) • Test: 84 (27.8)
Sentiment		Provape	Not provape
	• Total: 2175 (100) • Training: 1957 (100) • Test: 218 (100)	• Total: 1357 (62.39) • Training: 1221 (62.39) • Test: 136 (62.4)	• Total: 818 (37.61) • Training: 736 (37.61) • Test: 82 (37.6)

Table 4. Description of traditional classifiers and parameter settings used in the experiments (the same parameter settings were used for the following 3 targets: relevance, commercial, and sentiment).

Classifiers	Scikit-learn functions (version)	Parameter values
Logistic regression	sklearn.linear_model.LogisticRegression (0.20.3)	All default values except C=0.001
Random forest	sklearn.ensemble.RandomForestClassifier (0.20.3)	All default values except max_features="sqrt"
Support vector machine	sklearn.linear_model.SGDClassifier (0.20.3)	All default values except α=.01
Naive Bayes	sklearn.naive_bayes.MultinomialNB (0.20.3)	All default values

Table 5. Description of deep learning classifiers, target, and parameter settings used in the experiments.

Deep learning classifiers	Targets	Parameter values
Vaping-related word vectors		
CNN[a]	Relevance	max_features: 166,395, embed_size: 300, max_len: 75, optimizer: rmsprop, filters: 100, kernel_size: 1, epochs: 5, batch_size: 16
LSTM[b]	Relevance	max_features: 166,395, embed_size: 300, max_len: 75, optimizer: adam, epochs: 10, batch_size: 16
LSTM-CNN	Relevance	max_features: 166,395, embed_size: 300, max_len: 75, optimizer: adam, filters: 50, kernel_size: 2, epochs: 10, batch_size: 16
BiLSTM[c]	Relevance	max_features: 166,395, embed_size: 300, max_len: 75, optimizer: adam, epochs: 10, batch_size: 16
CNN	Commercial	max_features: 166,395, embed_size: 300, max_len: 75, optimizer: adam, filters: 100, kernel_size: 2, epochs: 10, batch_size: 16
LSTM	Commercial	max_features: 166,395, embed_size: 300, max_len: 75, optimizer: rmsprop, epochs: 5, batch_size: 32
LSTM-CNN	Commercial	max_features: 166,395, embed_size: 300, max_len: 75, optimizer: rmsprop, filters: 75, kernel_size: 2, epochs: 5, batch_size: 16
BiLSTM	Commercial	max_features: 166,395, embed_size: 300, max_len: 75, optimizer: adam, epochs: 5, batch_size: 64
CNN	Sentiment	max_features: 166,395, embed_size: 300, max_len: 75, optimizer: rmsprop, filters: 100, kernel_size: 2, epochs: 10, batch_size: 32
LSTM	Sentiment	max_features: 166,395, embed_size: 300, max_len: 75, optimizer: adam, epochs: 5, batch_size: 64
LSTM-CNN	Sentiment	max_features: 166,395, embed_size: 300, max_len: 75, optimizer: adam, filters: 75, kernel_size: 3, epochs: 5, batch_size: 64
BiLSTM	Sentiment	max_features: 166,395, embed_size: 300, max_len: 75, optimizer: rmsprop, epochs: 5, batch_size: 32
Global Vectors for Word Representation word vectors		
CNN	Relevance	max_features: 15,890, embed_size: 200, max_len: 75, optimizer: adam, filters: 100, kernel_size: 2, epochs: 10, batch_size: 16
LSTM	Relevance	max_features: 15,890, embed_size: 200, max_len: 75, optimizer: adam, epochs: 5, batch_size: 32
LSTM-CNN	Relevance	max_features: 15,890, embed_size: 200, max_len: 75, optimizer: adam, filters: 50, kernel_size: 2, epochs: 10, batch_size: 16
BiLSTM	Relevance	max_features: 15,890, embed_size: 200, max_len: 75, optimizer: adam, epochs: 5, batch_size: 64
CNN	Commercial	max_features: 10,842, embed_size: 200, max_len: 75, optimizer: rmsprop, filters: 50, kernel_size: 2, epochs: 5, batch_size: 16
LSTM	Commercial	max_features: 10,842, embed_size: 200, max_len: 75, optimizer: adam, epochs: 5, batch_size: 16
LSTM-CNN	Commercial	max_features: 10,842, embed_size: 200, max_len: 75, optimizer: adam, filters: 75, kernel_size: 2, epochs: 5, batch_size: 32
BiLSTM	Commercial	max_features: 10,842, embed_size: 200, max_len: 75, optimizer: adam, epochs: 5, batch_size: 64
CNN	Sentiment	max_features: 7979, embed_size: 200, max_len: 75, optimizer: rmsprop, filters: 100, kernel_size: 3, epochs: 5, batch_size: 64
LSTM	Sentiment	max_features: 7979, embed_size: 200, max_len: 75, optimizer: adam, epochs: 5, batch_size: 32
LSTM-CNN	Sentiment	max_features: 7979, embed_size: 200, max_len: 75, optimizer: rmsprop, filters: 75, kernel_size: 1, epochs: 10, batch_size: 64
BiLSTM	Sentiment	max_features: 7979, embed_size: 200, max_len: 75, optimizer: adam, epochs: 5, batch_size: 32

[a]CNN: convolutional neural network.

[b]LSTM: long short-term memory.

[c]BiLSTM: bidirectional long short-term memory.

Evaluation of Classifier Performance

We assessed the performance of the classifiers with the area under the receiver operating characteristic curve (AUC), precision, recall, and F1 scores. The AUC is a measure of discrimination, that is, how well a classifier differentiates between the positive and negative tweets, and larger values indicate better performance. Precision is the number of correctly classified positive tweets divided by the number of all positive tweets returned by the classifier, and recall is the number of correctly classified positive tweets divided by the number of all positive tweets. The F1 score is the harmonic average of the precision and recall; the F1 score achieves the best value at 1

when both precision and recall are perfect and the worst value at 0.

Evaluation of Relevance

To identify relevant words (features) in each classifier, we applied SHapley Additive exPlanations (SHAP), which is an algorithm for interpreting the relevance of features used in classifiers [37]. SHAP assigns each feature an average relevance value based on predictions on a data set. We examined the top 10 ranked features for each classifier.

Results

Proportions of Tweet Categories in the Annotated Data Set

In the annotated data set, 75.28% were relevant to vaping, and of the vaping-relevant tweets, 72.24% were of a noncommercial nature. Of the noncommercial vaping-relevant tweets, 62.39% contained provape sentiments.

Performance of Classifiers

Relevance Classifiers

Application of traditional classifiers yielded AUC values of 0.84 to 0.95, application of deep learning classifiers with vaping-related word vectors yielded AUC values of 0.90 to 0.93, and application of deep learning classifiers with GloVe word vectors yielded AUC values of 0.93 to 0.96. LR had the highest recall, whereas RF and the deep learning classifiers with GloVe word vectors had the highest F1 value. LSTM-CNN with GloVe word vectors performed the best overall with the highest AUC and precision values.

Commercial Classifiers

Overall, the AUC values were similar across all classifiers. Application of traditional classifiers yielded AUC values of 0.96 to 0.98, application of deep learning classifiers with vaping-related word vectors yielded AUC values of 0.97 to 0.98, and application of deep learning classifiers with GloVe word vectors yielded AUC values of 0.99. LSTM-CNN and BiLSTM with GloVe word vectors performed the best overall with the highest AUC, precision, recall, and F1 values.

Sentiment Classifiers

Application of traditional classifiers yielded AUC values of 0.69 to 0.78, application of deep learning classifiers with vaping-related word vectors yielded AUC values of 0.74 to 0.75, and application of deep learning classifiers with GloVe word vectors yielded AUC values of 0.78 to 0.83. BiLSTM and LSTM-CNN with GloVe word vectors performed the best overall with the highest AUC, precision, and F1 values.

Preprocessing

Our experiments showed that some traditional classifiers performed best with minimal preprocessing compared with others. LR and NB did not use any of the 10 preprocessing steps for any of the 3 targets (Multimedia Appendix 1). On the other hand, RF and SVM used 5 preprocessing steps on average (Multimedia Appendix 1). The deep learning classifiers performed better with no preprocessing compared with the fixed preprocessing pipeline. Furthermore, in addition to the standard text in tweets, information such as URLs, usernames, hashtags, and Unicode characters was found to be important and was included in most of the classifiers.

Feature Relevance

We applied the SHAP algorithm to the 12 classifiers for each target (corresponding to the classifiers in Tables 6,7, and 8) to generate 10 top-ranked features. The feature relevance plots for each classifier and target are shown in Multimedia Appendix 1. The word *vape* and its variations *vapes*, *vaping*, or *vapelife* appear in the 10 top-ranked features in all classifiers except RF relevance and commercial classifiers. Several textual placeholders appear in traditional classifiers, whereas several Unicode characters representing emojis appear in the deep learning classifiers. Interestingly, common simple words such as *we*, *as*, *was*, and *no* appear in many classifiers.

Table 6. Performance of relevance classifiers.

Classifiers	Area under the receiver operating characteristic curve (95% CI)	Precision	Recall	F1
Logistic regression	0.84 (0.78-0.89)	0.80	1.00	0.92
Random forest	0.95 (0.93-0.98)	0.93	0.97	*0.98*
Support vector machine	0.92 (0.88-0.96)	0.91	0.97	0.95
Naive Bayes	0.88 (0.83-0.93)	0.88	0.99	0.93
CNN[a] (vaping-related word vectors)	0.94 (0.91-0.97)	0.90	0.97	0.98
LSTM[b] (vaping-related word vectors)	0.91 (0.88-0.95)	0.89	0.98	0.96
LSTM-CNN (vaping-related word vectors)	0.89 (0.85-0.93)	0.93	0.87	0.95
BiLSTM[c] (vaping-related word vectors)	0.89 (0.85-0.94)	0.90	0.96	0.94
CNN (GloVe[d] word vectors)	0.95 (0.92-0.97)	0.93	0.95	0.98
LSTM (GloVe word vectors)	0.95 (0.92-0.98)	0.95	0.95	0.98
LSTM-CNN (GloVe word vectors)	0.96 (0.93-0.98)	0.96	0.93	0.98
BiLSTM (GloVe word vectors)	0.95 (0.93-0.98)	0.92	0.96	0.98

[a]CNN: convolutional neural network.

[b]LSTM: long short-term memory.

[c]BiLSTM: bidirectional long short-term memory.

[d]GloVe: Global Vectors for Word Representation.

Table 7. Performance of commercial classifiers.

Classifiers	Area under the receiver operating characteristic curve (95% CI)	Precision	Recall	F1
Logistic regression	0.98 (0.95-0.99)	0.93	0.83	0.96
Random forest	0.97 (0.96-0.99)	0.95	0.82	0.97
Support vector machine	0.98 (0.91-0.99)	0.92	0.86	0.92
Naive Bayes	0.96 (0.94-0.99)	0.83	0.89	0.92
CNN[a] (vaping-related word vectors)	0.98 (0.96-0.99)	0.93	0.75	0.94
LSTM[b] (vaping-related word vectors)	0.97 (0.95-0.99)	0.88	0.81	0.94
LSTM-CNN (vaping-related word vectors)	0.97 (0.94-0.99)	0.92	0.85	0.94
BiLSTM[c] (vaping-related word vectors)	0.98 (0.96-0.99)	0.84	0.87	0.95
CNN (GloVe[d] word vectors)	0.99 (0.98-0.99)	0.93	0.89	0.98
LSTM (GloVe word vectors)	0.99 (0.98-0.99)	0.89	0.94	0.98
LSTM-CNN (GloVe word vectors)	0.99 (0.98-0.99)	0.86	0.96	0.99
BiLSTM (GloVe word vectors)	0.99 (0.98-0.99)	0.97	0.88	0.98

[a]CNN: convolutional neural network.

[b]LSTM: long short-term memory.

[c]BiLSTM: bidirectional long short-term memory.

[d]GloVe: Global Vectors for Word Representation.

Table 8. Performance of sentiment classifiers.

Classifiers	Area under the receiver operating characteristic curve (95% CI)	Precision	Recall	F1
Logistic regression	0.78 (0.71-0.84)	0.73	0.88	0.82
Random forest	0.78 (0.70-0.83)	0.78	0.79	0.82
Support vector machine	0.69 (0.64-0.78)	0.66	0.98	0.75
Naive Bayes	0.75 (0.66-0.82)	0.75	0.79	0.80
CNN[a] (vaping-related word vectors)	0.74 (0.66-0.81)	0.73	0.85	0.80
LSTM[b] (vaping-related word vectors)	0.74 (0.69-0.82)	0.75	0.81	0.81
LSTM-CNN (vaping-related word vectors)	0.75 (0.71-0.84)	0.74	0.91	0.83
BiLSTM[c] (vaping-related word vectors)	0.74 (0.68-0.81)	0.72	0.91	0.82
CNN (GloVe[d] word vectors)	0.81 (0.75-0.87)	0.72	0.96	0.86
LSTM (GloVe word vectors)	0.78 (0.71-0.84)	0.76	0.82	0.84
LSTM-CNN (GloVe word vectors)	0.80 (0.74-0.86)	0.83	0.84	0.84
BiLSTM (GloVe word vectors)	0.83 (0.78-0.89)	0.79	0.79	0.88

[a]CNN: convolutional neural network.

[b]LSTM: long short-term memory.

[c]BiLSTM: bidirectional long short-term memory.

[d]GloVe: Global Vectors for Word Representation.

Discussion

Principal Findings

The relative prevalence of the 3 categories that we annotated in our data set reflects the general level of vaping-related discussions on Twitter. The high proportion of tweets of a commercial nature (72% of vaping-related tweets) reflects the observation that manufacturers of vaping products marketed their products heavily on Twitter. However, this percentage has likely decreased significantly since the beginning of 2020 because of the introduction of advertising restrictions by federal and state authorities. A high proportion of noncommercial tweets contained provape sentiments (62.39% of noncommercial tweets), suggesting that among Twitter users who post about vaping, the sentiment is overall more positive than negative in our data set, after the exclusion of marketing tweets. This reflects the growing prevalence of vaping, especially among adolescents who post more on Twitter than other age groups [38]. However, as this study used data before the FDA banned a range of flavored e-cigarette cartridges, both vaping and positive sentiments related to vaping may have decreased significantly.

Classifiers that we derived from our data set demonstrated high levels of performance, indicating that currently available machine learning methods can produce high-performing classifiers on a data set of only several thousand annotated tweets. Compared with traditional classifiers, deep learning classifiers had superior performance with AUC values of 0.96, 0.99, and 0.83 for predicting vaping-relevant, commercial, and provape tweets. Furthermore, our results indicate that deep learning classifiers performed the best with no preprocessing and with nondomain-specific GloVe word vectors. A few studies have shown that no preprocessing may provide better performance with Twitter data [39,40]. More generally, additional research is needed to systematically examine alternate preprocessing regimes for Twitter and other types of text data [41]. Although deep learning classifiers are computationally more expensive to derive compared with traditional classifiers, the lack of preprocessing and derivation of domain-specific word vectors offsets the computational cost. Moreover, the application of deep learning classifiers to new Twitter data is as computationally efficient as traditional classifiers.

Analyses of the 10 top-ranked features show that similar features appear across the classifiers. In addition to English terms, emojis and Unicode characters were often identified as useful features. Several common simple terms also appear as important features; these terms may interact with other features rather than being discriminatory on their own.

Limitations

Our study has several limitations. First, we used a small list of keywords to restrict our data, rather than using the full Twitter feed. As vaping products and their discussions evolve, the list of keywords will likely become stale and will need to be updated. Second, our annotated data set was of moderate size, though the sample size of 4000 tweets was adequate for obtaining classifiers with high performance. Third, the expression of tweet sentiments related to vaping is likely to vary over time [42]. It would be useful to evaluate the performance of the classifiers on data that are obtained from a different period to assess the generalizability of the classifiers over time. Fourth, there may be geographical variation in sentiments regarding vaping [43], and it would be useful to evaluate the performance of the classifiers on data that are obtained from different locations. In future work, we plan to address the limitations of

evaluating the classifiers over time and location. Fifth, it is not clear if individuals with certain personality traits make them more predisposed to express positive or negative sentiments [44]. More research is needed to assess the degree to which sentiment reflects variance in psychological traits versus the situational context in which those traits were expressed. Finally, this study uses data before the FDA banned a range of flavored e-cigarette cartridges that were likely to have been popular among frequent Twitter users, such as adolescents. In future work, we plan to derive classifiers from data that were collected after the FDA ban on flavored e-cigarette cartridges.

Future Surveillance Research

Machine learning classifiers, especially deep learning classifiers, show promising performance over strictly keyword-based approaches for identifying vaping-related tweets and sentiments related to vaping. This observation provides support for the development of a vaping surveillance system. Twitter surveillance can provide relatively inexpensive opportunities for monitoring the evolution of use and sentiment toward vaping and the effects of regulations on the marketing of vaping products. We plan to develop a surveillance system that will apply the classifiers to tweets to produce daily counts of vaping-related tweets, noncommercial tweets, and provape tweets. These daily counts will be used for future behavioral and attitudinal research related to vaping as well as for correlating changes in behavior and attitudes to changes in policy, such as those issued by the FDA. We plan to use the classifiers derived in this study as a basis for comparison with classifiers that we plan to derive from data obtained after the FDA ban to understand whether the ban has altered vaping-related health attitudes and behaviors. Furthermore, we plan to develop methods to infer the age group of the authors of tweets that will enable the daily tracking of vaping and related sentiments in adolescents.

Conclusions

We derived and evaluated machine learning classifiers to identify vaping-related relevant, commercial, and provape tweets. We developed a hierarchical classification scheme for vaping-related tweets and applied it to a data set of 4000 selected tweets to manually annotate them. We evaluated both traditional machine learning and deep learning classifiers using the annotated data set of 4000 tweets as well as vaping-related word vectors and GloVe word vectors that are derived from large unannotated tweet data sets. Overall, deep learning classifiers such as LSTM-CNN had superior performance and had the added advantage of requiring no preprocessing. These classifiers pave the way for the development of a vaping surveillance system.

Acknowledgments

The authors thank Erica Barrett, Daria Williams, and Sarah Matheny for data annotation. This work was supported by awards from the National Cancer Institute of the National Institutes of Health (R01-CA225773), the National Library of Medicine of the National Institutes of Health (R01-LM012095), and the National Science Foundation (ACI-1548562 and ACI-1445606 to the Pittsburgh Supercomputing Center). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional information.
[DOCX File , 2152 KB - jmir_v22i8e17478_app1.docx]

References

1. Rigotti NA. Balancing the benefits and harms of e-cigarettes: a national academies of science, engineering, and medicine report. Ann Intern Med 2018 May 1;168(9):666-667. [doi: 10.7326/M18-0251] [Medline: 29435573]
2. Czogala J, Goniewicz M, Fidelus B, Zielinska-Danch W, Travers M, Sobczak A. Secondhand exposure to vapors from electronic cigarettes. Nicotine Tob Res 2014 Jun;16(6):655-662 [FREE Full text] [doi: 10.1093/ntr/ntt203] [Medline: 24336346]
3. Layden JE, Ghinai I, Pray I, Kimball A, Layer M, Tenforde MW, et al. Pulmonary illness related to e-cigarette use in Illinois and Wisconsin - final report. N Engl J Med 2020 Mar 5;382(10):903-916. [doi: 10.1056/NEJMoa1911614] [Medline: 31491072]
4. Triantafyllou GA, Tiberio PJ, Zou RH, Lamberty PE, Lynch MJ, Kreit JW, et al. Vaping-associated acute lung injury: a case series. Am J Respir Crit Care Med 2019 Dec 1;200(11):1430-1431. [doi: 10.1164/rccm.201909-1809LE] [Medline: 31574235]
5. Farzal Z, Perry MF, Yarbrough WG, Kimple AJ. The adolescent vaping epidemic in the United States-how it happened and where we go from here. JAMA Otolaryngol Head Neck Surg 2019 Aug 22;145(10):885. [doi: 10.1001/jamaoto.2019.2410] [Medline: 31436792]

6. Sidani JE, Colditz JB, Barrett EL, Shensa A, Chu K, James AE, et al. I wake up and hit the JUUL: analyzing Twitter for JUUL nicotine effects and dependence. Drug Alcohol Depend 2019 Nov 1;204:107500. [doi: 10.1016/j.drugalcdep.2019.06.005] [Medline: 31499242]
7. Kim AE, Hansen H, Murphy J, Richards A, Duke J, Allen J. Methodological considerations in analyzing Twitter data. J Natl Cancer Inst Monogr 2013 Dec;2013(47):140-146. [doi: 10.1093/jncimonographs/lgt026] [Medline: 24395983]
8. Duggan M, Brenner J. The Demographics of Social Media Users — 2012. Pew Research Center. 2013. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2013/PIP_SocialMediaUsers.pdf [accessed 2020-07-18]
9. Eaton DK, Kann L, Kinchen S, Shanklin S, Ross J, Hawkins J, Centers for Disease Control and Prevention (CDC). Youth risk behavior surveillance - United States, 2009. MMWR Surveill Summ 2010 Jun 4;59(5):1-142 [FREE Full text] [Medline: 20520591]
10. Colditz JB, Welling J, Smith NA, James AE, Primack BA. Journal of clinical outcomes management.World vaping day: contextualizing vaping culture in online social media using a mixed methods approach. J Mix Methods Res 2017 Apr 9;13(2):196-215. [doi: 10.1177/1558689817702753]
11. Powell GE, Seifert HA, Reblin T, Burstein PJ, Blowers J, Menius JA, et al. Social media listening for routine post-marketing safety surveillance. Drug Saf 2016 May;39(5):443-454. [doi: 10.1007/s40264-015-0385-6] [Medline: 26798054]
12. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying multiple data collection tools to quantify human Papillomavirus vaccine communication on Twitter. J Med Internet Res 2016 Dec 5;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
13. McClellan C, Ali MM, Mutter R, Kroutil L, Landwehr J. Using social media to monitor mental health discussions - evidence from Twitter. J Am Med Inform Assoc 2017 May 1;24(3):496-502. [doi: 10.1093/jamia/ocw133] [Medline: 27707822]
14. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013 Sep 6;15(9):e189 [FREE Full text] [doi: 10.2196/jmir.2741] [Medline: 24014109]
15. Bian J, Topaloglu U, Yu F. Towards Large-scale Twitter Mining for Drug-Related Adverse Events. In: Proceedings of the 2012 International Workshop on Smart Health and Wellbeing. 2012 Presented at: SHB'12; October 29, 2012; Maui, HI, USA p. 25-32. [doi: 10.1145/2389707.2389713]
16. Prier KW, Smith M, Giraud-Carrier C, Hanson C. Identifying Health-Related Topics on Twitter. In: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction. 2011 Presented at: SBP'11; March 29-31, 2011; College Park, MD, USA p. 18-25. [doi: 10.1007/978-3-642-19656-0_4]
17. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013 Aug 29;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
18. Burton S, Dadich A, Soboleva A. Competing voices: marketing and counter-marketing alcohol on Twitter. J Nonprofit Public Sect Mark 2013 Apr;25(2):186-209. [doi: 10.1080/10495142.2013.787836]
19. Colditz JB, Chu K, Emery SL, Larkin CR, James AE, Welling J, et al. Toward real-time infoveillance of Twitter health messages. Am J Public Health 2018 Aug;108(8):1009-1014. [doi: 10.2105/AJPH.2018.304497] [Medline: 29927648]
20. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
21. Han S, Kavuluru R. Exploratory Analysis of Marketing and Non-Marketing E-cigarette Themes on Twitter. In: International Conference on Social Informatics. 2016 Presented at: SocInfo'16; November 11-14, 2016; Bellevue, USA p. 307-322. [doi: 10.1007/978-3-319-47874-6_22]
22. Cole-Lewis H, Varghese A, Sanders A, Schwarz M, Pugatch J, Augustson E. Assessing electronic cigarette-related tweets for sentiment and content using supervised machine learning. J Med Internet Res 2015 Aug 25;17(8):e208 [FREE Full text] [doi: 10.2196/jmir.4392] [Medline: 26307512]
23. Huang J, Kornfield R, Szczypka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. Tob Control 2014 Jul;23(Suppl 3):iii26-iii30 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051551] [Medline: 24935894]
24. Resende EC, Culotta A. A Demographic and Sentiment Analysis of E-cigarette Messages on Twitter. Computer Science Department, Illinois Institute of Technology. 2015. URL: http://cs.iit.edu/~culotta/pubs/resende15demographic.pdf [accessed 2020-07-21]
25. Du J, Xu J, Song H, Liu X, Tao C. Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets. J Biomed Semantics 2017 Mar 3;8(1):9 [FREE Full text] [doi: 10.1186/s13326-017-0120-6] [Medline: 28253919]
26. Du J, Xu J, Song H, Tao C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. BMC Med Inform Decis Mak 2017 Jul 5;17(Suppl 2):69 [FREE Full text] [doi: 10.1186/s12911-017-0469-6] [Medline: 28699569]
27. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr;20(1):37-46. [doi: 10.1177/001316446002000104]

28. Pang B, Lee L, Vaithyanathan S. Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. 2002 Presented at: EMNLP'02; July 6-7, 2002; Philadelphia, PA, USA. [doi: 10.3115/1118693.1118704]
29. Porter M. An algorithm for suffix stripping. Program 2006 Jul;40(3):211-218. [doi: 10.1108/00330330610681286]
30. Loper E, Bird S. NLTK: The Natural Language Toolkit. In: Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics (Vol. 1). 2002 Presented at: EMNLP '02; July 6-7; Philadelphia, PA, USA. [doi: 10.3115/1118108.1118117]
31. Pennington J, Socher R, Manning C. GloVe: Global Vectors for Word Representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: EMNLP'14; October 25-29, 2014; Doha, Qatar p. 1532-1543. [doi: 10.3115/v1/d14-1162]
32. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv preprint 2013:- epub ahead of print [FREE Full text]
33. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. J Mach Learn Res 2015 Jun;19(1):29-33. [doi: 10.1145/2786984.2786995]
34. Gulli A, Pal S. Deep Learning with Python. Birmingham, United Kingdom: Packt Publishing Ltd; 2017.
35. Ballesteros M, Dyer C, Smith NA. Improved Transition-based Parsing by Modeling Characters instead of Words with LSTMs. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015 Presented at: EMNLP 2015; September 17-21; Lisbon, Portugal. [doi: 10.18653/v1/d15-1041]
36. Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint 2015:- epub ahead of print [FREE Full text]
37. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Advances in Neural Information Processing Systems 30. 2017 Presented at: NIPS'17; December 4-9, 2017; Long Beach, CA p. 4765-4774 URL: https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions
38. Kim AE, Chew R, Wenger M, Cress M, Bukowski T, Farrelly M, et al. Estimated ages of JUUL Twitter followers. JAMA Pediatr 2019 Jul 1;173(7):690-692 [FREE Full text] [doi: 10.1001/jamapediatrics.2019.0922] [Medline: 31107511]
39. Hamidian S, Diab M. Rumor detection and classification for Twitter data. arXiv preprint 2019:- epub ahead of print [FREE Full text]
40. Zilincik M, Navrat P, Koskova G. Exploratory search on Twitter utilizing user feedback and multi-perspective microblog analysis. PLoS One 2013;8(11):e78857 [FREE Full text] [doi: 10.1371/journal.pone.0078857] [Medline: 24265724]
41. Denny MJ, Spirling A. Program-electron libtext preprocessing for unsupervised learning: why it matters, when it misleads, and what to do about it. Polit Anal 2018 Mar 19;26(2):168-189. [doi: 10.1017/pan.2017.44]
42. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
43. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
44. Auer EML. Detecting Deceptive Impression Management Behaviors in Interviews Using Natural Language Processing. ODU Digital Commons. 2018. URL: https://digitalcommons.odu.edu/psychology_etds/70/ [accessed 2020-07-18]

Abbreviations

AUC: area under the receiver operating characteristic curve
BiLSTM: bidirectional long short-term memory
CNN: convolutional neural network
e-cigarettes: electronic cigarettes
FDA: Food and Drug Administration
GloVe: Global Vectors for Word Representation
LR: logistic regression
LSTM: long short-term memory
NB: naive Bayes
RF: random forest
RITHM: real-time infoveillance of Twitter health messages
SHAP: SHapley Additive exPlanations
SVM: support vector machine
TF-IDF: frequency-inverse document frequency

Edited by G Eysenbach; submitted 16.12.19; peer-reviewed by R Gore, K Binger, N Lambert; comments to author 10.03.20; revised version received 05.06.20; accepted 11.06.20; published 12.08.20.

Please cite as:
Visweswaran S, Colditz JB, O'Halloran P, Han NR, Taneja SB, Welling J, Chu KH, Sidani JE, Primack BA
Machine Learning Classifiers for Twitter Surveillance of Vaping: Comparative Machine Learning Study
J Med Internet Res 2020;22(8):e17478
URL: https://www.jmir.org/2020/8/e17478
doi:10.2196/17478
PMID:32784184

Original Paper

Tactics for Drawing Youth to Vaping: Content Analysis of Electronic Cigarette Advertisements

Laura L Struik[1], BSCN, MSN, PhD; Sarah Dow-Fleisner[2], BA, BSc, MA, PhD; Michelle Belliveau[2], BA; Desiree Thompson[1], BScN; Robert Janke[3], BA, MLIS

[1]School of Nursing, Department of Health and Social Development, University of British Columbia Okanagan, Kelowna, BC, Canada

[2]School of Social Work, Department of Health and Social Development, University of British Columbia Okanagan, Kelowna, BC, Canada

[3]Library, University of British Columbia Okanagan, Kelowna, BC, Canada

Corresponding Author:
Laura L Struik, BSCN, MSN, PhD
School of Nursing
Department of Health and Social Development
University of British Columbia Okanagan
ARTS 140
1147 Research Road
Kelowna, BC, V1V1V7
Canada
Phone: 1 2508077792
Email: laura.struik@ubc.ca

Abstract

Background: The use of electronic cigarettes (e-cigarettes), also known as vaping, has risen exponentially among North American youth in recent years and has become a critical public health concern. The marketing strategies used by e-cigarette companies have been associated with the uptick in use among youth, with video advertisements on television and other electronic platforms being the most pervasive strategy. It is unknown how these advertisements may be tapping into youth needs and preferences.

Objective: The aim of this 2-phase study was to examine the marketing strategies that underpin e-cigarette advertisements, specifically in the context of television.

Methods: In phase 1, a scoping review was conducted to identify various influences on e-cigarette uptake among youth. Results of this scoping review informed the development of a coding framework. In phase 2, this framework was used to analyze the content of e-cigarette advertisements as seen on 2 popular television channels (Discovery and AMC).

Results: In phase 1, a total of 20 articles met the inclusion criteria. The resultant framework consisted of 16 key influences on e-cigarette uptake among youth, which were categorized under 4 headings: personal, relational, environmental, and product-related. In phase 2, 38 e-cigarette advertisements were collected from iSpot.tv and represented 11 popular e-cigarette brands. All of the advertisements tapped into the cited influences of youth e-cigarette uptake, with the most commonly cited influences (product and relational) tapping into the most, at 97% (37/38) and 53% (20/38), respectively.

Conclusions: The findings highlight the multidimensional influences on youth uptake of e-cigarettes, which has important implications for developing effective antivaping messages, and assist public health professionals in providing more comprehensive prevention and cessation support as it relates to e-cigarette use. The findings also bring forward tangible strategies employed by e-cigarette companies to recruit youth into vaping. Understanding this is vital to the development of cohesive strategies that combat these provaping messages.

(J Med Internet Res 2020;22(8):e18943) doi:10.2196/18943

KEYWORDS

qualitative research; electronic nicotine delivery systems; marketing; advertisement; youth; vaping

Introduction

The use of electronic cigarettes (e-cigarettes), also known as vaping, among youth is a pressing public health concern in the United States and Canada. Importantly, the initiation and use of vaping products among youth have been associated with immediate and lasting health consequences [1-3]. Yet, according to a recent report, nearly 1 in 3 high school–aged students and 1 in 7 middle school–aged students report vaping [4,5], which reflects a significant increase in use. Specifically, between 2017 and 2019, vaping among youth in the United States has more than doubled, with reported use increasing from 4% to 9% among 8th graders, 8% to 20% among 10th graders, and 11% to 25% among 12th graders [5]. Canada reflects similar trends, with past 30-day use of e-cigarettes up from 10% in 2016-2017 to 20% in 2018-2019 among youth in grades 7 to 12 [4].

While e-cigarettes entered the North American market as a cessation tool in 2008, e-cigarette use among non-smoking youth suggests that cessation is not a primary reason for use in this age demographic. According to recent statistics, only 3% of Canadian youth in grades 7 to 12 are current smokers, and 20% are current e-cigarettes users, suggesting that upwards of 17% of e-cigarette users were originally non-smokers [4]. In addition, among youth who do smoke combustible cigarettes, fewer than 8% report using e-cigarettes for smoking cessation [6]. In relation to smoking cessation, emerging evidence is inconclusive as to the effectiveness of e-cigarettes as a cessation method for youth [7], and some research even suggests that it contributes to ongoing nicotine addiction [8]. Even more concerning is that vaping among youth has been linked to a 3 times greater likelihood to try traditional cigarettes [9-12]. These findings have contributed to what the US Department of Health and Human Services Secretary Alex Azar described as “an epidemic of youth e-cigarette use, which threatens to engulf a new generation in nicotine addiction” [13].

Marketing strategies used by e-cigarette companies have been associated with the uptick in vaping among youth [10]. Many studies have found that the promotion of e-cigarettes through various channels (eg, television [TV], social media influencers) has lent to increased positive perceptions of vaping and intentions to use vaping products and contributed to e-cigarette uptake among youth [14-20]. The use of TV as a mode of marketing has been the most recent focus for e-cigarette companies. Between 2018 and 2019, JUUL, the most popular e-cigarette brand in North America [21,22], spent US $57 million on TV advertisements to promote their products [23]. While JUUL claims that these advertisements are aimed at helping adults find a healthier alternative to smoking, many public health advocates are concerned these advertisements may also attract youth [24], and research confirms that these advertisements increase exposure to their products, which subsequently increases the likelihood of use by youth [25,26]. For example, in a recent study, it was found that exposure to e-cigarette advertisements among youth is not uncommon, with 28% of youth in the United States, 17% of youth in Canada, and 21% of youth in England having seen a vaping advertisement [25]. Furthermore, nearly 40% of youth across all three countries reported that the advertisements made vaping look appealing, and about 44% of youth perceived that vaping advertisements targeted non-smokers [25]. Indeed, skepticism is warranted in relation to how e-cigarettes are portrayed in advertisements.

Subject to the Food and Drug Administration Family and Tobacco Control Act in the United States [27] and the Tobacco and Vaping Products Act in Canada [28], e-cigarette companies are not permitted to market or sell their products to youth. While this broad stipulation is helpful, we know little about what marketing to youth truly looks like in order to deem advertisements to be appropriately adhering to this provision. Given that advertisements clearly influence uptake among youth, it is important that the ways in which the advertisements are tapping into youth needs and preferences be examined. Thus, the aim of this study was two-fold: first, to conduct a scoping review to identify the known influences associated with youth vaping and second, to analyze the content of e-cigarette advertisements using content analysis to identify and describe overlapping themes.

Methods

Aim 1: Scoping Review of Vaping Influences on Youth

A scoping review of the literature was conducted to identify various empirically supported influences on e-cigarette uptake among youth, which was then used to develop the coding framework to analyze influences present in vaping ads. The search strategy was developed by the research team, including a research librarian and experts in youth vaping and substance use, and based on a methodology reported by Kinouani and colleagues [29]. The scoping review used two databases, Medline on the OVID platform and CINAHL on the EBSCO platform, using a combination of keywords and database-specific subject headings. The search strategy from Medline is included in Multimedia Appendix 1. Inclusion criteria were studies that were in North America, published in English, focused on correlates of and influences on e-cigarette use, and focused on youth. Also, both qualitative and quantitative methodologies and literature reviews were included. Exclusion criteria were studies that were a surveillance of e-cigarette use, focused on traditional cigarette use, biomedical studies, or not focused on youth.

The initial search, limited to articles in English with an abstract available, resulted in 937 articles across both databases. Results were uploaded to RefWorks, a bibliographic management software package, and removal of duplicates resulted in 855 articles left for screening. Using the exclusion and inclusion criteria, two research assistants screened the article titles and abstracts independently, with disagreements resolved in collaboration with the first and second authors. After abstract and title screening, the research team members screened 37 articles at the full-text level, which resulted in 20 final articles for analysis. For the final 20 articles, two research assistants extracted all significant influences on youth vaping. Using Excel, the cited influences on youth e-cigarette uptake were listed. Using conventional content analysis, categories and subcategories were inductively derived [30]. Two research team members extracted data from the first 5 articles, generating

initial codes and categories for the different influences on youth uptake of e-cigarettes. The whole research team met to achieve consensus on the approach to data extraction. The team then generated an overall framework of categories and subcategories. Two research team members then extracted data about the influences on vaping among youth from the remaining articles. Once all the articles were reviewed and data were extracted, the team met again and refined the coding framework into 4 overall categories associated with youth vaping influences.

Aim 2: Application of Coding Framework to e-Cigarette Advertisements

Data Collection of e-Cigarette Advertisements

A total of 38 e-cigarette advertisements were collected using the freely available iSpot.tv, which is a real-time television advertisement platform [31]. Using the iSpot.tv platform, we searched for industry-driven advertisements related to e-cigarettes that were aired in North America between December 15, 2019 and January 15, 2020. Using this platform ensured that the advertisements collected for this study were generated by e-cigarette companies versus informal or individually developed e-cigarette promotions, like those posted on social media sites. Initially, common e-cigarette brands were searched including BLU, JUUL, FIN, LOGIC, and VUSE, which resulted in 29 nationally aired advertisements. The terms "vape," "vaping," and "e-cigarettes" were also used to search for advertisements, which resulted in 8 additional advertisements for the brands CUE, FreeBoxMod, O2PUR, VCHIC, and VERO. The search for advertisements was completed between December 15, 2019 and January 15, 2020.

Advertisement Content Analysis

The coding framework developed from the scoping review of influences on youth vaping was used to deductively analyze the content of the advertisements. Deductive content analysis is the process of applying data to a pre-existing framework [30]. First, two researchers coded the same 2 advertisements using the a priori themes developed through the systematic review of the literature on vaping influence for youths. The research team then met to discuss the application of the coding framework and reached consensus regarding the coding process. The two researchers then finished applying the framework to the advertisements. In addition to the framework, we coded the look and feel of the ads to capture demographic and other contextual data (eg, age, race, and sex of individuals in the ads and location setting of the ads). Finally, we noted key messages in the narratives and taglines for each advertisement.

Results

Scoping Review Description

The scoping review generated a coding framework that revealed 4 overall categories associated with youth vaping, including personal, relational, environmental, and product-related influences, with 16 subcategories. Personal influences were related to whether youth reported vaping to remove negative affect (eg, daily stress, anxiety, boredom), recreation was related to youth vaping for the purposes of having fun (eg, doing tricks), and finally curiosity was related to vaping to experiment with something new. Relational influences were related to family approval, such as getting positive message about vaping from parents or siblings, and included whether parents, siblings, or peers were using e-cigarettes. This category also included whether youth viewed vaping as enhancing social capital (eg, viewing vaping as a necessary part of social events) and enhancing social acceptance (eg, using e-cigarettes to fit in with peers). Environmental factors were related to external factors influencing ease of access or use (eg, easy to use and obtain e-cigarettes) and the impact of cost related to vaping. Finally, we found that elements related to vaping products were key influencers for youth. These included the ability to use the product discreetly (eg, no bad smell), have a positive sensory experience (eg, good flavors and better "buzz" than cigarettes), and use a new or innovative product. Lastly, perceptions that vaping was less harmful than cigarettes and that it could be used to support smoking cessation influenced e-cigarette uptake among youth.

Of the 20 articles reviewed, there were 57 different cited examples of product-related influences, 41 relational influences, 19 personal influences, and 11 environmental influences. Of note, an article could have more than one example of an influence type. For example, Ickes and colleagues [32] reported ease of use or access, peer use, good flavors, and low cost as some influences of current use and initial use. Among subcategories, perceptions of vaping being less harmful than cigarettes and peer use were among the most cited influences for youth e-cigarette use, followed closely by the perception of vaping as a positive sensory experience and use as a smoking cessation aid. For instance, 81% of youth reported initiation of e-cigarette use because a friend vaped, and 80% reported continued e-cigarette use because of the good flavors. See Table 1 for final categories and subcategories with citations and examples and Multimedia Appendix 2 for references of articles from the scoping review.

Table 1. Influences on youth electronic cigarette (e-cigarette) uptake. See Multimedia Appendix 2 for the reference list.

Influences	Articles	Example
Personal		
Removal of negative affect	[5,7,9,11,18]	Participant reports: "I have issues with anxiety…sometimes if I'm dealing with sensory overload…[vaping] really helps" [5]
Recreation	[2,5,9,11,18]	22.4% reported vaping to have a good time, 21.6% to relax, and 23.5% to reduce boredom [18]
Curiosity	[2,8,9,11-14,17,18]	95% of youth reported curiosity as the reason for initiating vaping [11]
Relational		
Family approval	[2,6,12,13,17,19,20]	6.8 times greater risk of vaping if there is an e-cigarette user at home [6]
Parent use	[1,4,8,10]	Higher rate of youth vaping (14%) associated with maternal e-cigarette use [10]
Sibling use	[2,10,17]	Participant reports: "I got it [e-cigarette] from my older brother; he was with his friends…he told me I should try it" [2]
Peer use	[2,4-6,8,10,12-15,17,19,20]	Friend vaping associated with an increased frequency of use (r=.30, P<.001) [20]
Enhance social capital	[5,7,13,17,18]	Participant reports: "[Vaping] tasted good and it was mostly a social thing. It looked cool, and I wanted other people to think that I looked cool" [5]
Enhance social acceptance	[1,5,6,8,10,14,15,17,20]	28% of youth who ever vaped and 46% with current use reported vaping to feel more comfortable in social situations [7]
Environmental		
Easy to access or use	[2,4,10,11,13,16,19,20]	91% of youth reported "ease of use" as their reason for continued use of e-cigarettes [11]
Cost	[8,13,19]	2.5%-3.9% reported vaping because they cost less than cigarettes [19]
Product		
Discreet	[8-13,18,19]	1.76 times more likely to try vaping because it can be hidden from adults [13]
Positive sensory experience	[2,5,8-14,17-20]	42% youth reported 'good flavors' as a reason for first use [8]
Less harmful	[2-8,10,11,13,15-17,19]	52-54% youth with past 30-day use reported vaping was not harmful to their health [7]
New or novel product	[2,12-15,17,18,20]	72% reported trying e-cigarettes because they were something new, cool, or fun [12]
Smoking cessation	[4,5,8-13,15,17-19]	8.5% report using e-cigarettes to quit smoking [9]

Advertisement Descriptions and Context

Among the 38 e-cigarette advertisements reviewed, 11 advertisements were for BLU, 7 for JUUL, 6 for VUSE, 1 for CUE, 3 for FIN, 3 for LOGIC, 3 for O2PUR, 2 for VERO, 1 for FreeBoxMod, and 1 for VCHIC. Of the 38 advertisements reviewed, 73.7% (28/38) included people, with 16 of 28 advertisements featuring individuals as couples or in groups. Most advertisements featured people who appeared to be white (25/38) or black (10/38), with 9 advertisements showing more than one race in the advertisement. Among the 28 advertisement showing people, 17 included both male and female actors, 6 had male actors only, and 5 had female actors only. Additionally, among advertisements featuring people, most showed individuals who appeared to be 19-30 years old (15/28) or 31-40 years old (14/28). The most common settings for advertisements were during recreational activities (eg, party or camping; 13/38), in a city (11/38), or within an individual's home (10/38), followed by advertisements showing a person's workplace (5/38). Many advertisements showed individuals using e-cigarettes in a variety of settings. For example, advertisements showed individuals using e-cigarettes at home, work, school, in social settings (eg, at a bar), and while doing recreational activities (eg, biking, camping). Many of the advertisements (eg, VUSE) used incredible graphic designs with the use of

vibrant colors, animation, music, and setting transitions. While all the advertisements utilized action-oriented videography to market their products, they varied in terms of being virtual animation only, real-life settings, or a mix of both. Finally, advertisements also frequently included narration and taglines related to the product. Phrases included, "Take back your freedom" (BLU), "Make the switch" (JUUL), "Satisfying; It's that simple" (LOGIC), and "Real draw, real taste, real satisfaction" (VUSE).

Advertisements and Vaping Influences

Among the 38 advertisements reviewed, all 4 of the main influences identified in the scoping review (personal, relational, environmental, and product) and the majority of subthemes in the framework were present. Parental and sibling use and exposure to advertisements were not present in the advertisements analyzed. However, there was mention of siblings and spouses being more accepting of vaping than smoking (eg, JUUL). The majority of advertisements (97%; 37/38) had at least one influence related to the product, and 20 advertisements (20/38, 53%) included at least one element related to relational influences. Additionally, 16 advertisements (16/38, 42%) included elements related to personal influences, and 9 (9/38, 24%) included environmental influences. The most common influences present in vaping advertisements included the product as new or innovative (30/38), the positive sensory experience of vaping (20/38), the ability of vaping to enhance social acceptance (18/38), and vaping as an alternative to smoking (18/38). In sum, all the advertisements included at least one influence, with an average of 4.39 (SD 0.31) influences per advertisement (range 1-10).

Furthermore, advertisements often emphasized the ability to derive "satisfaction" from the product (eg, LOGIC, VUSE, CUE, and FIN). Common words and phrases included "100% flavor," "satisfaction at last," "satisfaction," "real satisfaction," "unrivaled taste satisfaction," and "a truly satisfying taste" and encouraged consumers to "draw" and "taste" the array of flavors. Additionally, companies like BLU, JUUL, VERO, and VCHIC presented their product as an "alternative" to smoking and included phrases like "make the switch" and "rise from the ashes." See Table 2 for the number of advertisements with a particular influence and examples of how the influence was presented within the advertisement.

Table 2. Content analysis of vaping advertisements.

Influences	Number of ads	Examples
Personal (n=16)		
Removal of negative affect	4	"Now that I've switched to BLU, I feel so much better about myself" (BLU 1056056)
Recreation	4	Actor (man) hiking, bicycling, and intent to race car while vaping (BLU 1075930)
Curiosity	10	"Try CUEThese risk free, and change your life" (CUE 1717816)
Relational (n=20)		
Family approval	4	Adult son on the switch from smoking to vaping based on his mother's suggestion. "This [the switch to JUUL] came from her [Mother], really." (JUUL 2059571)
Parent use	0	No explicit mention of parent use leading to vaping.
Sibling use	0	No explicit mention of sibling use leading to vaping.
Peer use	3	"It was a friend of mine that said, why wouldn't you just try the JUUL?" (JUUL, 2060997)
Enhance social capital	12	"I can whip out my BLU and not worry about scaring that special someone away," (BLU 1056056).
Enhance social acceptance	18	"There was a time when no one was offended by it [smoking]. That time has come again" (FIN, 1044824)
Environmental (n=14)		
Easy to access or use	9	"Truly vaping made easy." (CUE 171816)
Cost	13	"VCHIC saves me over $150 dollars a month" (VCHIC 1089981)
Product (n=37)		
Discreet	17	"A better way to enjoy everything you love about smoking, only without the smell" (CUE 1717816)
Positive sensory experience	20	"Real draw, real taste, real satisfaction" (VUSE, 2148838) "Four new flavors to awaken your senses (VUSE, 1266750)
Less harmful	5	"Definitely has all the stuff I want, and not all the bad stuff" (FIN 1148496)
New or novel product	30	"Innovation has changed the world, moving us all forward, isn't it time smoking changed too?" (VUSE 2148838)
Smoking cessation	18	"I was a pack -a-day smoker for more than 30 years….(until I switched to JUUL)," (JUUL, 2346024)

Discussion

Principal Findings

The factors influencing youth vaping are varied, complex, and multidimensional. A scoping review of the extant literature identified 16 major influences associated with youth vaping, which fell under the categories of personal, relational, environmental, and product-related factors. These findings indicate that vaping among youth is prompted by a variety of factors and not just one factor alone. To date, however, prevention campaigns have typically focused on addressing perceptions of harmlessness. For example, a recent review of 21 prevention interventions across North America confirmed that these efforts frequently appeared to be designed to teach youth about the dangers of vaping and encouraging them to refrain from or to stop vaping [33]. While these efforts are a much-needed step forward, the findings from this scoping review present more than just perceptions of harmlessness as influencing youth uptake and include curiosity, social factors, and stress and anxiety, to name a few. The provision of information on the harms is a common response to substance use [34], but with little evidence to support its effectiveness [35,36]. While it is important to relay information and implement protective policies, it is also important to acknowledge other factors at play. Hyshka [34] suggested that prevention interventions for substance use target the social determinants rather than the individual behavior to improve young people's health and well-being. In this regard, prevention efforts for vaping would benefit from a deeper understanding of the various and relevant reasons and pathways to vaping among youth populations and should develop holistic interventions driven by youth.

The TV advertisements reviewed were found to tap into almost all of the reasons that youth cite for taking up e-cigarettes. The most highly cited reasons were most prominent in the ads, including a focus on relational aspects of vaping and product-related benefits, such as a positive sensory experience. Similar to the findings of the present study, a recently published focus group study that analyzed e-cigarette advertisements with 39 non-vaping adolescents found that the perceived social benefits, like increased friendships, and product-related appeals, like the innovative design and variety of flavors, presented in the advertisements were major draws to trying e-cigarettes [37]. The present study builds on these findings even further and adds to a growing evidence base that adolescents are indeed the target market for e-cigarette advertisers, despite claims that this is not the case. In addition, findings from this study shed light on how advertisements are successfully drawing this demographic into vaping, offering tangible results that tobacco control advocates can draw upon to inform the development of policies and interventions to combat these marketing strategies.

Rather than emphasizing the use of e-cigarettes to address nicotine addiction, the advertisements emphasized e-cigarettes as a solution to maintain nicotine dependence by portraying e-cigarettes as an innovative way to get the nicotine fix most commonly associated with cigarettes. While phrases like "make the switch" by JUUL may appear to suggest e-cigarettes as tools for smoking cessation, it does not appear to be the intention of these advertisements or the e-cigarette companies. For example, JUUL advertisements (eg, JUUL, 2060997) include a warning that says: "JUUL is not a smoking-cessation product and has not been approved by the FDA for the treatment, prevention, or cure of any specific disease or condition." Despite this, e-cigarette companies like JUUL still benefit from lax regulations and the ability to advertise their products as an alternative nicotine product. Adding to this, many of the advertisements ridiculed traditional cigarettes for being outdated. In this regard, e-cigarette companies are differentiating themselves from traditional cigarettes by minimizing, and even attempting to eradicate, any relationship with the stigmatized cigarette. To promote e-cigarettes as a solution to get a nicotine fix and, at the same time, purposefully convey e-cigarettes as unrelated to cigarettes are in direct opposition to the mandate that they are marketed as a cessation aid for cigarette smokers. Additionally, most e-cigarette pods contain up to three times the amount of nicotine compared to a pack of combustible cigarettes [38], which further contradicts the notion that e-cigarettes are for the purposes of cessation.

Comparison to Prior Work

That the TV advertisements largely tapped into the influence of peers and relationships, one of the most commonly cited influences on uptake in the scoping review results, is noteworthy. Peers become an important influence during the adolescent stage of development, with social networks and social acceptance essential parts of positive identity development [39-41]. This establishment of identity and acceptance in a social group has been termed entitativity [42,43]. A high sense of entitativity is associated with improved peer relationships and self-efficacy, and a low sense of entitativity is associated with poor views of self and peers [42]. Entitativity occurs with youth as they identify with a social group to gain social acceptance and respect [41]. For youth, entitativity alters perceptions and beliefs and increases willingness to participate in risky health-related activities [41]. For example, entitativity in smoking groups has been found to lend to more homogenous opinions and reinforces positive perceptions of smoking [43,44]. Furthermore, individuals that strongly identify with smoking groups are more likely to resist antismoking campaigns [44]. Similar to smoking, e-cigarettes allow youth to form a community based on the inclusion criteria of vaping, which enables opportunities for increasing social networks and establishing a group identity [45,46]. Narratives of social identity and belonging around e-cigarette use have a powerful impact on youth, not only for shaping positive attitudes around vaping but also for resisting prevention efforts. Indeed, the advertisements reviewed in the present study may serve as a pre-emptive measure to keep youth vaping despite the roll out of anticipated antivaping campaigns and interventions.

Despite most of the advertisements being 30-60 seconds long, they still managed to convey e-cigarettes as less harmful than traditional cigarettes by presenting them as the cleaner alternative to cigarettes or as a cessation aid. This is similar to the findings of Richardson and colleagues [47], who found that online advertisements prioritized messages of reduced harm within their short window of advertising time. This is concerning

because there is no basis to make claims of reduced harm, and a lack of knowledge should not be conflated with no harm. Even more concerning is that youth are particularly vulnerable to these messages [24,48]. Making unfounded claims around harmlessness supports misperceptions of e-cigarette safety, promotes uptake, delays cessation, and encourages dual use of tobacco products [49,50], the latter of which has recently been found to significantly compound long-term health risks, including chronic obstructive pulmonary disease [51] and stroke [52]. Making modified risk claims, therefore, contributes significantly to the burden of population-level harm.

It is also noteworthy that the advertisements promoted the ability to use e-cigarettes in a variety of contexts, including contexts where traditional cigarette smoking is banned, such as indoors or in urban areas, because of their sleek design and lack of odor [53,54]. Vaping is essentially a way to avoid smoke-free policies due to a current lack of regulations [55,56]. In this regard, these advertisements reinforce already problematic patterns of use among youth, such as hiding it from parents [54,57] or using e-cigarettes in challenging contexts, like schools [58,59]. By reinforcing the ability to vape anywhere, harmful patterns of use will likely increase.

A prominent theme in the ads was the incitement of curiosity through presentation of vaping as something new. Curiosity is a predictor of youth starting to vape, and curiosity about vaping is often associated with increased exposure to advertisements [60-62]. e-Cigarette use in the advertisements is framed in such a way as to peak curiosity through increasing novelty, going against the norm, and messages that these products are innovative. Curiosity is a powerful experience for youth that appears to inform both smoking and vaping behavior. These findings add further support to the idea that exposure to advertisements is informing youth vaping behavior.

Using stunning sensory appeals, such as through graphical displays that emphasize the visual appeal of sleek technological designs, and the pleasurable taste and smell of e-cigarettes, the advertisements play to the sensation-seeking needs of youth. Adolescence is a period of psychosocial development where it is more likely that an individual will engage in sensation seeking than at other developmental stage [63]. Other researchers confirm that e-cigarette advertisements capitalize on this sensation-seeking behavior by emphasizing pleasurable sensations, such as flavors or odor produced by the e-juice [37,64-66]. The advertisements reinforce the belief that e-cigarettes will satisfy the sensation-seeking needs of youth by highlighting the sensory experience that comes with e-cigarette use in their animations, graphics, and music. The high rates of vaping among youth can be attributed to these strategic, developmentally targeted, marketing tactics.

Future Research

Youth vaping is a concern, and there is a growing need for comprehensive strategic plans to curtail youth vaping. One strategy is to provide stronger and clearer regulations for vaping ads. Given the findings of this study, we see that many advertisements inadvertently or directly included elements in advertisements that tap into key influences of vaping among youth. Thus, we can conclude that the current regulations are not having the intended impact. Findings from this study suggest that future research should explore the development and impact of regulations that consider the personal, relational, environmental, and product influences that are most meaningful for youth and ensure that these are limited in vaping advertisements. For example, regulations should restrict the sensory-appeal tactics in advertisements.

In addition, while this study looked at influences on e-cigarette use among the general population of youth and how advertisements tapped into these influences, it would be interesting to examine differences according to developmental stage. Broadly, adolescence is a time when youth are learning to assert their independence, try new tasks, and seek out support from peers more so than parents, which influences risky health behaviors like vaping [67]. Thus, future research should look at stages of adolescent development in relation to susceptibility to advertising and motivations around vaping (eg, early adolescence [10-13 years old], middle adolescence [13-15 years old], late adolescence [15-19 years old]). As seen in other tobacco prevention efforts [68], this would lend to more targeted efforts to curb youth vaping.

Research aimed at developing and evaluating interventions to curtail youth vaping is also needed. The findings of this study provide preliminary evidence for influencing factors that may be taken into consideration in interventions. Further research is needed to explore and understand how different populations of youth perceive e-cigarette advertisements and how the advertisements appeal to them. Finally, research that examines how different aspects of the advertisements may be playing to vulnerable populations, such as low-income groups, needs to be explored further.

Limitations

The findings of this study are limited to North American applications. Factors that influence use and the tactics of e-cigarette advertisements in other countries may be different. In addition, we do not know if reasons for uptake of e-cigarettes are the same for adults and is an area for future research. It is also important to note that the advertisement selection was limited to the iSpot.tv platform, and there may be other ads with nuances not captured in this study. Furthermore, while not all e-cigarette companies have TV advertisements, it is important to note that this study represents 10 out of over 450 brands now available [69]. Finally, the advertising landscape for e-cigarettes is a rapidly evolving space, in terms of content and delivery channels, and it is important to consider the study findings in light of changes in the field.

Conclusions

The findings of this study reveal the multidimensional influences on youth uptake of e-cigarettes. Importantly, this study also reveals how these influences are leveraged in e-cigarette advertisements. This contributes to a more nuanced understanding of how e-cigarette companies market their products and to whom. Understanding this is vital to our understanding of how to combat them.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Medline search strategy.
[DOCX File , 13 KB - jmir_v22i8e18943_app1.docx]

Multimedia Appendix 2
Scoping review references.
[DOCX File , 17 KB - jmir_v22i8e18943_app2.docx]

References

1. Chaumont M, de Becker B, Zaher W, Culié A, Deprez G, Mélot C, et al. Differential Effects of E-Cigarette on Microvascular Endothelial Function, Arterial Stiffness and Oxidative Stress: A Randomized Crossover Trial. Sci Rep 2018 Jul 10;8(1):10378. [doi: 10.1038/s41598-018-28723-0] [Medline: 29991814]
2. Government of Canada. Information Update - Health Canada warns of potential risk of pulmonary illness associated with vaping products. Recalls and safety alerts. 2019 Oct 17. URL: https://healthycanadians.gc.ca/recall-alert-rappel-avis/hc-sc/2019/70919a-eng.php [accessed 2020-03-18]
3. Kuntic M, Oelze M, Steven S, Kröller-Schön S, Stamm P, Kalinovic S, et al. Short-term e-cigarette vapour exposure causes vascular oxidative stress and dysfunction: evidence for a close connection to brain damage and a key role of the phagocytic NADPH oxidase (NOX-2). Eur Heart J 2020 Jul 07;41(26):2472-2483 [FREE Full text] [doi: 10.1093/eurheartj/ehz772] [Medline: 31715629]
4. Canadian Tobacco, Alcohol and Drugs Survey (CTADS): summary of results for 2017. Health Canada. 2019 Dec 23. URL: https://www.canada.ca/en/health-canada/services/canadian-tobacco-alcohol-drugs-survey/2017-summary.html [accessed 2020-03-18]
5. Miech R, Patrick ME, O'Malley PM, Johnston LD. What are kids vaping? Results from a national survey of US adolescents. Tob Control 2017 Jul 25;26(4):386-391 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053014] [Medline: 27562412]
6. Tsai J, Walton K, Coleman BN, Sharapova SR, Johnson SE, Kennedy SM, et al. Reasons for Electronic Cigarette Use Among Middle and High School Students - National Youth Tobacco Survey, United States, 2016. MMWR Morb Mortal Wkly Rep 2018 Feb 16;67(6):196-200 [FREE Full text] [doi: 10.15585/mmwr.mm6706a5] [Medline: 29447148]
7. Barrington-Trimis JL, Kong G, Leventhal AM, Liu F, Mayer M, Cruz TB, et al. E-cigarette Use and Subsequent Smoking Frequency Among Adolescents. Pediatrics 2018 Dec 05;142(6):e20180486 [FREE Full text] [doi: 10.1542/peds.2018-0486] [Medline: 30397165]
8. Lee S, Grana RA, Glantz SA. Electronic cigarette use among Korean adolescents: a cross-sectional study of market penetration, dual use, and relationship to quit attempts and former smoking. J Adolesc Health 2014 Jun;54(6):684-690 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.11.003] [Medline: 24274973]
9. Hammond D, Reid JL, Cole AG, Leatherdale ST. Electronic cigarette use and smoking initiation among youth: a longitudinal cohort study. CMAJ 2017 Oct 30;189(43):E1328-E1336 [FREE Full text] [doi: 10.1503/cmaj.161002] [Medline: 29084759]
10. Hammond D, Wackowski O, Reid J, O'Connor R. Use of JUUL E-cigarettes Among Youth in the United States. Nicotine Tob Res 2020 Apr 21;22(5):827-832 [FREE Full text] [doi: 10.1093/ntr/nty237] [Medline: 30371838]
11. National Academies of Sciences, Engineering, and Medicine. Public health consequences of e-cigarettes. Washington, DC: The National Academies Press; 2018.
12. Soneji S, Barrington-Trimis JL, Wills TA, Leventhal AM, Unger JB, Gibson LA, et al. Association Between Initial Use of e-Cigarettes and Subsequent Cigarette Smoking Among Adolescents and Young Adults: A Systematic Review and Meta-analysis. JAMA Pediatr 2017 Aug 01;171(8):788-797 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.1488] [Medline: 28654986]
13. Results from 2018 National Youth Tobacco Survey show dramatic increase in e-cigarette use among youth over past year. U.S. Food & Drug Administration. 2018 Nov 15. URL: https://www.fda.gov/news-events/press-announcements/results-2018-national-youth-tobacco-survey-show-dramatic-increase-e-cigarette-use-among-youth-over [accessed 2020-03-18]
14. Duke JC, Allen JA, Eggers ME, Nonnemaker J, Farrelly MC. Exploring Differences in Youth Perceptions of the Effectiveness of Electronic Cigarette Television Advertisements. Nicotine Tob Res 2016 May 26;18(5):1382-1386. [doi: 10.1093/ntr/ntv264] [Medline: 26706908]
15. Farrelly MC, Duke JC, Crankshaw EC, Eggers ME, Lee YO, Nonnemaker JM, et al. A Randomized Trial of the Effect of E-cigarette TV Advertisements on Intentions to Use E-cigarettes. Am J Prev Med 2015 Nov 29;49(5):686-693. [doi: 10.1016/j.amepre.2015.05.010] [Medline: 26163170]

16. Hansen J, Hanewinkel R, Morgenstern M. Electronic cigarette advertising and teen smoking initiation. Addict Behav 2020 Apr;103:106243. [doi: 10.1016/j.addbeh.2019.106243] [Medline: 31855726]
17. Mantey DS, Cooper MR, Clendennen SL, Pasch KE, Perry CL. E-Cigarette Marketing Exposure Is Associated With E-Cigarette Use Among US Youth. J Adolesc Health 2016 Jun;58(6):686-690 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.03.003] [Medline: 27080732]
18. Pu J, Zhang X. Exposure to advertising and perception, interest, and use of e-cigarettes among adolescents: findings from the US National Youth Tobacco Survey. Perspect Public Health 2017 Nov;137(6):322-325. [doi: 10.1177/1757913917703151] [Medline: 28379069]
19. Singh T, Agaku IT, Arrazola RA, Marynak KL, Neff LJ, Rolle IT, et al. Exposure to Advertisements and Electronic Cigarette Use Among US Middle and High School Students. Pediatrics 2016 May 25;137(5):e20154155-e20154155 [FREE Full text] [doi: 10.1542/peds.2015-4155] [Medline: 27244815]
20. Villanti AC, Rath JM, Williams VF, Pearson JL, Richardson A, Abrams DB, et al. Impact of Exposure to Electronic Cigarette Advertising on Susceptibility and Trial of Electronic Cigarettes and Cigarettes in US Young Adults: A Randomized Controlled Trial. Nicotine Tob Res 2016 May;18(5):1331-1339. [doi: 10.1093/ntr/ntv235] [Medline: 26574551]
21. Craver R. Juul ends 2018 with 76 percent market share. Winston-Salem Journal. 2019 Jan 08. URL: https://www.journalnow.com/business/juul-ends-with-percent-market-share/article_6f50f427-19ec-50be-8b0c-d3df18d08759.html [accessed 2020-03-19]
22. Huang J, Duan Z, Kwok J, Binns S, Vera LE, Kim Y, et al. Vaping versus JUULing: how the extraordinary growth and marketing of JUUL transformed the US retail e-cigarette market. Tob Control 2018 May 31;28(2):146-151 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054382]
23. Baysinger T. E-Cigarette Brands Spent $57 Million on TV Ads This Year, Including 54% to Networks That Now Reject Them. TheWrap. 2019 Sep 19. URL: https://www.thewrap.com/e-cigarette-brands-spent-57-million-on-tv-ads-this-year-including-54-to-networks-that-now-reject-them/ [accessed 2020-03-16]
24. Andrews J, Mays D, Netemeyer R, Burton S, Kees J. Effects of E-Cigarette Health Warnings and Modified Risk Ad Claims on Adolescent E-Cigarette Craving and Susceptibility. Nicotine Tob Res 2019 May 21;21(6):792-798. [doi: 10.1093/ntr/nty076] [Medline: 29669064]
25. Cho YJ, Thrasher JF, Reid JL, Hitchman S, Hammond D. Youth self-reported exposure to and perceptions of vaping advertisements: Findings from the 2017 International Tobacco Control Youth Tobacco and Vaping Survey. Prev Med 2019 Sep;126:105775. [doi: 10.1016/j.ypmed.2019.105775] [Medline: 31323286]
26. Loukas A, Paddock EM, Li X, Harrell MB, Pasch KE, Perry CL. Electronic Nicotine Delivery Systems Marketing and Initiation Among Youth and Young Adults. Pediatrics 2019 Sep 26;144(3):e20183601. [doi: 10.1542/peds.2018-3601] [Medline: 31451608]
27. Family Smoking Prevention and Tobacco Control and Federal Retirement Reform. Public Law 111–31. 2009 Jun 22. URL: http://www.gpo.gov/fdsys/pkg/PLAW-111publ31/pdf/PLAW-111publ31.pdf [accessed 2020-03-13]
28. Tobacco and Vaping Products Act. Minister of Justice. 2019 Nov 09. URL: https://laws-lois.justice.gc.ca/PDF/T-11.5.pdf [accessed 2020-03-18]
29. Kinouani S, Leflot C, Vanderkam P, Auriacombe M, Langlois E, Tzourio C. Motivations for using electronic cigarettes in young adults: A systematic review. Subst Abus 2020 Oct 22;41(3):315-322. [doi: 10.1080/08897077.2019.1671937] [Medline: 31638872]
30. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
31. Muller S, Happell S. U.S. Patent No 10405037: Television audience measurement method and apparatus. Washington, DC: US Patent and Trademark Office; 2014 Sep 17. URL: https://patents.google.com/patent/US9402111B2/en [accessed 2019-12-08]
32. Ickes M, Hester JW, Wiggins AT, Rayens MK, Hahn EJ, Kavuluru R. Prevalence and reasons for Juul use among college students. J Am Coll Health 2020 Jul 26;68(5):455-459 [FREE Full text] [doi: 10.1080/07448481.2019.1577867] [Medline: 30913003]
33. O'Connor S, Pelletier H, Bayoumy, D, Schwartz R. Interventions to Prevent Harms from Vaping: Report for the Central East TCAN. Ontario Tobacco Research Unit. 2019 May 01. URL: https://www.otru.org/documents/interventions-to-prevent-harms-from-vaping/ [accessed 2020-04-22]
34. Hyshka E. Applying a social determinants of health perspective to early adolescent cannabis use – An overview. Drugs: Education, Prevention and Policy 2013 Jan 14;20(2):110-119. [doi: 10.3109/09687637.2012.752434]
35. Degenhardt L, Hall W. Extent of illicit drug use and dependence, and their contribution to the global burden of disease. The Lancet 2012 Jan;379(9810):55-70. [doi: 10.1016/s0140-6736(11)61138-0]
36. Room R, Reuter P. How well do international drug conventions protect public health? The Lancet 2012 Jan;379(9810):84-91. [doi: 10.1016/s0140-6736(11)61423-2]

37. Chen Y, Tilden C, Vernberg DK. Adolescents' interpretations of e-cigarette advertising and their engagement with e-cigarette information: results from five focus groups. Psychol Health 2020 Feb 16;35(2):163-176. [doi: 10.1080/08870446.2019.1652752] [Medline: 31418593]
38. Jackler RK, Ramamurthi D. Nicotine arms race: JUUL and the high-nicotine product market. Tob Control 2019 Nov 06;28(6):623-628. [doi: 10.1136/tobaccocontrol-2018-054796] [Medline: 30733312]
39. Arslan G. Psychological Maltreatment, Social Acceptance, Social Connectedness, and Subjective Well-Being in Adolescents. J Happiness Stud 2017 Feb 9;19(4):983-1001. [doi: 10.1007/s10902-017-9856-z]
40. El Bouhaddani S, van Domburgh L, Schaefer B, Doreleijers TAH, Veling W. Peer status in relation to psychotic experiences and psychosocial problems in adolescents: a longitudinal school-based study. Eur Child Adolesc Psychiatry 2018 Jun 13;27(6):701-710. [doi: 10.1007/s00787-017-1063-2] [Medline: 29030694]
41. Hogg MA, Siegel JT, Hohman ZP. Groups can jeopardize your health: Identifying with unhealthy groups to reduce self-uncertainty. Self and Identity 2011 Jul;10(3):326-335. [doi: 10.1080/15298868.2011.558762]
42. Dang J, Liu L, Du Y. Benefits of a Highly Entitative Class for Adolescents' Psychological Well-Being in School. School Mental Health 2019 Mar 30;11(4):766-776. [doi: 10.1007/s12310-019-09319-z]
43. Park M. How Smoking Advocates are Connected Online: An Examination of Online Social Relationships Supporting Smoking Behaviors. J Health Commun 2020;25(1):82-90. [doi: 10.1080/10810730.2019.1709924] [Medline: 31885336]
44. Blondé J, Falomir-Pichastor JM. Smoker identity and resistance to antismoking campaigns: The role of group entitativity. Group Processes & Intergroup Relations 2019 Nov 21:136843021988463. [doi: 10.1177/1368430219884633]
45. Padon A, Lochbuehler K, Maloney E, Cappella J. A Randomized Trial of the Effect of Youth Appealing E-Cigarette Advertising on Susceptibility to Use E-Cigarettes Among Youth. Nicotine Tob Res 2018 Jul 09;20(8):954-961 [FREE Full text] [doi: 10.1093/ntr/ntx155] [Medline: 29106669]
46. Yule JA, Tinson JS. Youth and the sociability of "Vaping". J. Consumer Behav 2016 Aug 31;16(1):3-14. [doi: 10.1002/cb.1597]
47. Richardson A, Ganz O, Vallone D. Tobacco on the web: surveillance and characterisation of online tobacco and e-cigarette advertising. Tob Control 2015 Jul;24(4):341-347. [doi: 10.1136/tobaccocontrol-2013-051246] [Medline: 24532710]
48. Pollay RW, Siddarth S, Siegel M, Haddix A, Merritt RK, Giovino GA, et al. The Last Straw? Cigarette Advertising and Realized Market Shares among Youths and Adults, 1979–1993. Journal of Marketing 2018 Nov 28;60(2):1-16. [doi: 10.1177/002224299606000201]
49. Cobb NK, Byron MJ, Abrams DB, Shields PG. Novel Nicotine Delivery Systems and Public Health: The Rise of the "E-Cigarette". Am J Public Health 2010 Dec;100(12):2340-2342. [doi: 10.2105/ajph.2010.199281]
50. Cobb NK, Abrams DB. E-Cigarette or Drug-Delivery Device? Regulating Novel Nicotine Products. N Engl J Med 2011 Jul 21;365(3):193-195. [doi: 10.1056/nejmp1105249]
51. Bhatta DN, Glantz SA. Association of E-Cigarette Use With Respiratory Disease Among Adults: A Longitudinal Analysis. Am J Prev Med 2020 Feb;58(2):182-190. [doi: 10.1016/j.amepre.2019.07.028] [Medline: 31859175]
52. Parekh T, Pemmasani S, Desai R. Risk of Stroke With E-Cigarette and Combustible Cigarette Use in Young Adults. Am J Prev Med 2020 Mar;58(3):446-452. [doi: 10.1016/j.amepre.2019.10.008] [Medline: 31924460]
53. Dai H, Hao J. Flavored Tobacco Use Among U.S. Adults by Age Group: 2013-2014. Subst Use Misuse 2019 Oct 31;54(2):315-323. [doi: 10.1080/10826084.2018.1521428] [Medline: 30380969]
54. Ramamurthi D, Chau C, Jackler R. JUUL and other stealth vaporisers: hiding the habit from parents and teachers. Tob Control 2018 Sep 15:610-616. [doi: 10.1136/tobaccocontrol-2018-054455] [Medline: 30219794]
55. Berg CJ, Barr DB, Stratton E, Escoffery C, Kegler M. Attitudes toward E-Cigarettes, Reasons for Initiating E-Cigarette Use, and Changes in Smoking Behavior after Initiation: A Pilot Longitudinal Study of Regular Cigarette Smokers. Open J Prev Med 2014 Oct;4(10):789-800 [FREE Full text] [doi: 10.4236/ojpm.2014.410089] [Medline: 25621193]
56. Glantz SA, Bareham DW. E-Cigarettes: Use, Effects on Smoking, Risks, and Policy Implications. Annu Rev Public Health 2018 Apr 01;39(1):215-235 [FREE Full text] [doi: 10.1146/annurev-publhealth-040617-013757] [Medline: 29323609]
57. Hilton S, Weishaar H, Sweeting H, Trevisan F, Katikireddi SV. E-cigarettes, a safer alternative for teenagers? A UK focus group study of teenagers' views. BMJ Open 2016 Nov 16;6(11):e013271 [FREE Full text] [doi: 10.1136/bmjopen-2016-013271] [Medline: 27852721]
58. Kirkcaldy A, Fairbrother H, Weiner K, Curtis P. Young people's perspectives of e-cigarette use in the home. Health Place 2019 May;57:157-164. [doi: 10.1016/j.healthplace.2019.04.005] [Medline: 31054499]
59. Schillo BA, Cuccia AF, Patel M, Simard B, Donovan EM, Hair EC, et al. JUUL in School: Teacher and Administrator Awareness and Policies of E-Cigarettes and JUUL in U.S. Middle and High Schools. Health Promot Pract 2020 Jan 18;21(1):20-24. [doi: 10.1177/1524839919868222] [Medline: 31530185]
60. Goniewicz ML, Boykan R, Messina CR, Eliscu A, Tolentino J. High exposure to nicotine among adolescents who use Juul and other vape pod systems ('pods'). Tob Control 2019 Nov 07;28(6):676-677. [doi: 10.1136/tobaccocontrol-2018-054565] [Medline: 30194085]
61. Margolis K, Nguyen A, Slavit W, King B. E-cigarette curiosity among U.S. middle and high school students: Findings from the 2014 national youth tobacco survey. Prev Med 2016 Aug;89:1-6 [FREE Full text] [doi: 10.1016/j.ypmed.2016.05.001] [Medline: 27155440]

62. Portnoy DB, Wu CC, Tworek C, Chen J, Borek N. Youth curiosity about cigarettes, smokeless tobacco, and cigars: prevalence and associations with advertising. Am J Prev Med 2014 Aug;47(2 Suppl 1):S76-S86 [FREE Full text] [doi: 10.1016/j.amepre.2014.04.012] [Medline: 25044199]
63. Khurana A, Romer D, Betancourt LM, Hurt H. Modeling Trajectories of Sensation Seeking and Impulsivity Dimensions from Early to Late Adolescence: Universal Trends or Distinct Sub-groups? J Youth Adolesc 2018 Sep 6;47(9):1992-2005 [FREE Full text] [doi: 10.1007/s10964-018-0891-9] [Medline: 29980957]
64. Bevan I. E-cigarettes. Contemporary Drug Problems 2016 Jul 20;43(3):228-241. [doi: 10.1177/0091450916657348]
65. Leventhal AM, Goldenson NI, Cho J, Kirkpatrick MG, McConnell RS, Stone MD, et al. Flavored E-cigarette Use and Progression of Vaping in Adolescents. Pediatrics 2019 Nov 28;144(5):e20190789. [doi: 10.1542/peds.2019-0789] [Medline: 31659004]
66. Thrul J, Gubner NR, Tice CL, Lisha NE, Ling PM. Young adults report increased pleasure from using e-cigarettes and smoking tobacco cigarettes when drinking alcohol. Addict Behav 2019 Jun;93:135-140 [FREE Full text] [doi: 10.1016/j.addbeh.2019.01.011] [Medline: 30710807]
67. Crone EA, Dahl RE. Understanding adolescence as a period of social-affective engagement and goal flexibility. Nat Rev Neurosci 2012 Sep 20;13(9):636-650. [doi: 10.1038/nrn3313] [Medline: 22903221]
68. Sussman S. A Lifespan Developmental-Stage Approach to Tobacco and Other Drug Abuse Prevention. ISRN Addict 2013;2013:745783-745719 [FREE Full text] [doi: 10.1155/2013/745783] [Medline: 25298961]
69. Zhu S, Sun JY, Bonnevie E, Cummins SE, Gamst A, Yin L, et al. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. Tob Control 2014 Jul 16;23(Suppl 3):iii3-iii9 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051670] [Medline: 24935895]

Abbreviations

e-Cigarettes: electronic cigarettes
TV: television

Edited by G Eysenbach; submitted 28.03.20; peer-reviewed by K McCausland, D Levine, E Westling; comments to author 12.06.20; revised version received 15.06.20; accepted 15.06.20; published 14.08.20.

Please cite as:
Struik LL, Dow-Fleisner S, Belliveau M, Thompson D, Janke R
Tactics for Drawing Youth to Vaping: Content Analysis of Electronic Cigarette Advertisements
J Med Internet Res 2020;22(8):e18943
URL: https://www.jmir.org/2020/8/e18943
doi:10.2196/18943
PMID:32663163

Original Paper

Racial and Ethnic Disparities in Patient Experiences in the United States: 4-Year Content Analysis of Twitter

Yulin Hswen[1,2,3,4], ScD; Jared B Hawkins[4,5], MMSc, PhD; Kara Sewalk[4], MPH; Gaurav Tuli[4], PhD; David R Williams[6,7], MPH, PhD; K Viswanath[6,8], PhD; S V Subramanian[6,7], PhD; John S Brownstein[4,5], PhD

[1]Boston Children's Hospital, Boston, MA, United States

[2]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, United States

[3]Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, CA, United States

[4]Innovation Program, Boston Children's Hospital, Boston, MA, United States

[5]Computational Epidemiology Lab, Harvard Medical School, Boston, MA, United States

[6]Department of Social and Behavioral Sciences, Harvard TH Chan School of Public Health, Boston, MA, United States

[7]Harvard Center for Population and Development Studies, Harvard University, Cambridge, MA, United States

[8]Center for Community-Based Research, Dana-Farber Cancer Institute, Boston, MA, United States

Corresponding Author:
Yulin Hswen, ScD
Department of Epidemiology and Biostatistics
University of California San Francisco
550 16th Street
San Francisco, CA
United States
Phone: 1 617 775 1889
Email: yhswen@gmail.com

Abstract

Background: Racial and ethnic minority groups often face worse patient experiences compared with the general population, which is directly related to poorer health outcomes within these minority populations. Evaluation of patient experience among racial and ethnic minority groups has been difficult due to lack of representation in traditional health care surveys.

Objective: This study aims to assess the feasibility of Twitter for identifying racial and ethnic disparities in patient experience across the United States from 2013 to 2016.

Methods: In total, 851,973 patient experience tweets with geographic location information from the United States were collected from 2013 to 2016. Patient experience tweets included discussions related to care received in a hospital, urgent care, or any other health institution. Ordinary least squares multiple regression was used to model patient experience sentiment and racial and ethnic groups over the 2013 to 2016 period and in relation to the implementation of the Patient Protection and Affordable Care Act (ACA) in 2014.

Results: Racial and ethnic distribution of users on Twitter was highly correlated with population estimates from the United States Census Bureau's 5-year survey from 2016 (r^2=0.99; P<.001). From 2013 to 2016, the average patient experience sentiment was highest for White patients, followed by Asian/Pacific Islander, Hispanic/Latino, and American Indian/Alaska Native patients. A reduction in negative patient experience sentiment on Twitter for all racial and ethnic groups was seen from 2013 to 2016. Twitter users who identified as Hispanic/Latino showed the greatest improvement in patient experience, with a 1.5 times greater increase (P<.001) than Twitter users who identified as White. Twitter users who identified as Black had the highest increase in patient experience postimplementation of the ACA (2014-2016) compared with preimplementation of the ACA (2013), and this change was 2.2 times (P<.001) greater than Twitter users who identified as White.

Conclusions: The ACA mandated the implementation of the measurement of patient experience of care delivery. Considering that quality assessment of care is required, Twitter may offer the ability to monitor patient experiences across diverse racial and ethnic groups and inform the evaluation of health policies like the ACA.

(J Med Internet Res 2020;22(8):e17048) doi:10.2196/17048

KEYWORDS

racial disparities; race; patient experience; policy; social media; digital epidemiology; social determinants of health; health disparities; health inequities

Introduction

In the United States, racial and ethnic minority populations experience suboptimal access to quality health care [1]. Since patient experience is strongly associated with quality health care [2-4], lower quality health care within these groups may be attributed to poorer patient experiences. Because of these poorer experiences by racial and ethnic minorities, authentic patient experience can be difficult to capture in traditional health care research. For instance, a systematic review of 44 articles showed that mistrust, stigmatization, fears, and lack of access to information prevented racial and ethnic minority groups from participating in health research [5].

Most research surrounding patient experience has been conducted by the health care system, whereby quality assessments about patient experience are retrieved from the health care institution that is providing these health care services to the patients [1,6-8]. Therefore, fear of consequences of reporting negative feedback to the health care institution that is responsible for respondents' care may bias the results of these studies [9-11]. The validity of health care–based surveys about patient experience are put into question because a strong correlation between positive patient satisfaction and response rate has been documented [12-14]. Research suggests that levels of reported patient satisfaction with health care may be inflated due to high risks of response bias, especially among racial and ethnic minorities who have historically received poor treatment and care [13]. For example, White patients are more likely to respond to health care surveys compared with racial and ethnic minorities [15,16] and are overrepresented in many health care studies about patient experience and care [17].

Measurement of patient experience has also gained recognition as a component of the Patient Protection and Affordable Care Act (ACA) of 2010 [18]. The ACA was enacted in order to improve the access to health care through the expansion of public health coverage, improve affordability of health insurance, and make the health care system more accountable to diverse patient populations, like racial and ethnic minorities that have historically had a lack of health coverage [19]. Since the ACA's major provisions came into force in 2014, as of 2016, it is estimated that up to 24 million additional people have received insurance coverage [20]. Additionally, the law repeatedly refers to the importance of patient-centeredness, patient satisfaction, and the measurement of patient experience of care, highlighting the importance of accurately measuring these patient-reported experiences [21,22]. Therefore, novel approaches are needed to better understand patient experience, especially among underrepresented racial and ethnic minority groups, which could also shed light on the impact of new health care policies and changing legislation, such as the implementation of the ACA.

Data derived from Twitter may offer an opportunity to capture authentic patient experiences and broaden the view of existing health care surveys on patient satisfaction. Twitter is a popular social networking and microblogging service that has over 330 million monthly active users worldwide and approximately 500 million tweets per day [23,24]. Tweets are limited to 280 characters and have been recognized as a source of organic sentiment and opinions [25]. The Twitter platform has been widely recognized as a source to monitor public sentiment across a spectrum of health-related issues, including mental health [26-28], vaccination [29], and smoking [30]. Most recently, Twitter has emerged as a potential source of information for capturing hospital and health care experiences of sexual minorities [31,32]. Considering that a greater proportion of Black and Hispanic users are on Twitter [33,34], data collected from this social media platform may make it possible to capture patient experiences from minority populations that are typically underrepresented in traditional research on patient satisfaction.

This study seeks to evaluate the ability of Twitter to monitor patient experiences in racial and ethnic minority groups in the United States. Specifically, we examined tweets about patient experience among racial and ethnic groups, including White, Black, Hispanic/Latino, Asian/Pacific Islander, and American Indian/Alaska Native patients in the United States. We further investigated trends in sentiment of these tweets across race and ethnicity from 2013 to 2016. Lastly, we explored changes in patient experience sentiment postimplementation of the ACA (2014-2016) to preimplementation of the ACA (2013) across racial and ethnic groups to understand the potential impact of this health policy on racial and ethnic disparities in patient experience.

Methods

Patient Experience Data Set

This study used a previously established Twitter patient experience data set to investigate racial and ethnic disparities in patient experience in the United States [35]. This patient experience data set was created from a support vector machine–based supervised machine learning classifier that was iteratively built to specifically identify tweets related to patient experience. A patient experience tweet was defined as any tweet that discussed any exposure to health care, such as care received in a hospital, urgent care, or any other health institution.

We used a geolocation inference engine validated by a previous study for the patient experience data set [35] that used a combination of users' profiles, GPS, and the Google Maps Geocoding application programming interface to identify the geographic location of the user. A total of 2,759,257 tweets were labeled as patient experience tweets from February 1, 2013, to February 28, 2017, out of which 876,384 (31.76%) tweets were inferred to one of the 50 US states, the District of Columbia, or the US Virgin Islands by the geolocation inference engine [35]. To align our analyses with the 2016 census data, we excluded data from 2017. In total, 851,973 geolocation inference tweets were used in this analysis.

Patient Experience Sentiment

We used natural language processing to measure the sentiment of all patient experience tweets. The sentiment of patient experience tweets was determined using a widely accepted lexicon and rule-based sentiment classifier for microblogs, Valence Aware Dictionary for Sentiment Reasoner (VADER). VADER is based on a pattern library that is trained from human-annotated words commonly found in product reviews [36,37]. VADER is often used for product reviews and news articles and therefore does not always contain similar text-based characteristics to Twitter. Therefore, we appended VADER's dictionary and rules to provide broader representation for Twitter, such as incorporating over 110 emojis and their respective sentiment scores [38]. VADER computes sentiment for each word and generates a compound score for the sentence by summing the sentiment score of each word. We considered sentiment score to be positive if the mean compound score was greater than or equal to 0.3 and negative if the score was less than or equal to –0.3. Mean compound scores between –0.3 and 0.3 were considered neutral.

Classification of Race and Ethnicity

The relationship between surname distribution and population structure dates to the 19th century [39,40]. More recently, the Human Genome Diversity Project has shown a strong correlation of surnames and genetic linkages between human groups [41], which has caused name-based classification of race and ethnicity to be frequently used in population-based studies to identify racial and ethnic identities when information is not directly available [39]. In 2010, the United States Census Bureau conducted a study to develop a classification system to identify racial and ethnic identities associated with a list of names from the 2010 decennial survey [42]. In this study, surnames were recorded for 295 million people (95.5% of the population) and surnames with a frequency of 100 or more were used to identify race and ethnicity. These 162,253 names cover 90% of the people recorded in the United States Census Bureau decennial survey in 2010.

Therefore, we used the United States Census Bureau surname classification system [43] to build a surname classifier to identify race and ethnicity in our Twitter population. The profile names of Twitter users were collected and matched to the United States Census Bureau surname classification database (N=162,253) to identify the race and ethnicity of Twitter users in our study [42]. Racial and ethnic categorizations were based on the categories used in the United States Census Bureau 2010 decennial survey, which included White, Black, Hispanic/Latino, Asian/Pacific Islander, and American Indian/Alaska Native persons. Out of the 851,973 geolocation tweets, 392,215 tweets were used that had racial and ethnic inferred information in the analyses presented in this study.

Statistical Analysis

We used two methods to compare the racial and ethnic distribution of our Twitter data with the 5-year US census estimates for 2016. We first used the Kendall τ rank correlation to compare the ranking of the states by number of users that tweeted about patient experience and compared this with the rank of populations by state based on United States Census Bureau's 2016 American Community Survey 5-year estimates. Second, we quantified the strength of the correlation of distribution across race and ethnicity for each state with a Pearson correlation test and reported the r^2 coefficient variable.

We identified the relative difference in sentiment of patient experience across race and ethnicity using ordinary least squares regression. To examine the change in sentiment in patient experience across race and ethnicity from 2013 to 2016, we included a year × race/ethnicity interaction term and tested its significance in the ordinary least squares regression model. To evaluate the potential impact of the implementation of the ACA's full provisions in 2014, we generated a dichotomous dummy variable to distinguish between the two time periods of preimplementation (2013) versus postimplementation of the ACA's provisions (2014-2016), along with the interaction terms between that variable and the variable of race and ethnicity. This study received approval from the Boston Children Hospital's Institutional Review Board.

Results

Geographic Distribution of Patient Experience Across Race and Ethnicity

The results between the distribution of Twitter users per state and the population estimates from the 2016 census 5-year survey show that Kendall τ was 0.845 (P<.001) and Pearson coefficient r^2 was 0.99 (P<.001), as illustrated in Figure 1. The Kendall τ rank correlation was highest for the distribution of Hispanic/Latino patients (tau-b=0.74), followed by Black (tau-b=0.62), White (tau-b=0.57), Asian/Pacific Islander (tau-b=0.47), and lastly American Indian/Alaska Native patients (tau-b=0.40), with all P values <.001. The Pearson r^2 coefficient for the correlation between the patient experience user counts per state and the corresponding population estimates by the US census was highest for Hispanic/Latino patients (r^2=0.98), followed by Asian/Pacific Islander (r^2=0.91), White (r^2=0.85), Black (r^2=0.91), and lastly American Indian/Alaska Native patients (r^2=0.59), with all P values <.001, as seen in Figure 2.

Figure 1. Correlation between the patient experience user counts per state and the corresponding population estimates by US census. PTE: patient experience.

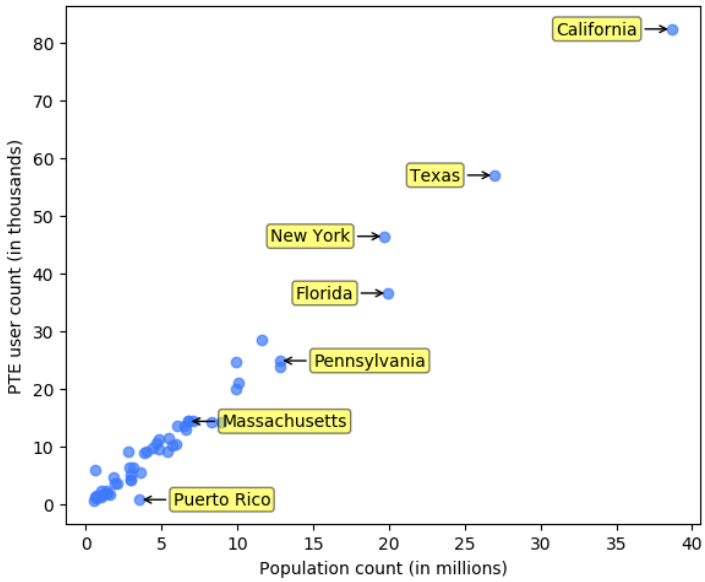
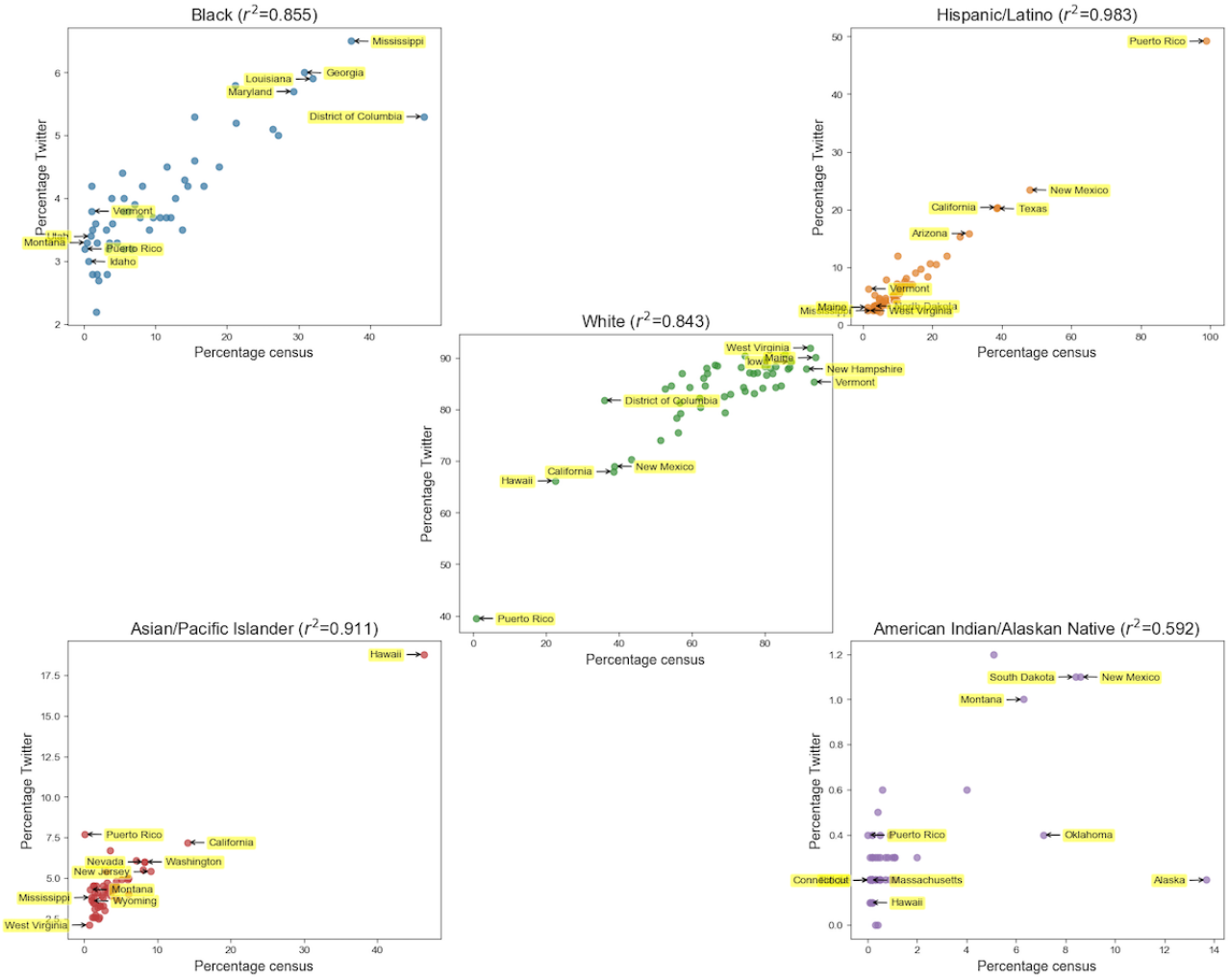


Figure 2. Correlation between the patient experience user counts per state and the corresponding population estimates by the US census by racial and ethnic group from 2013 to 2016.



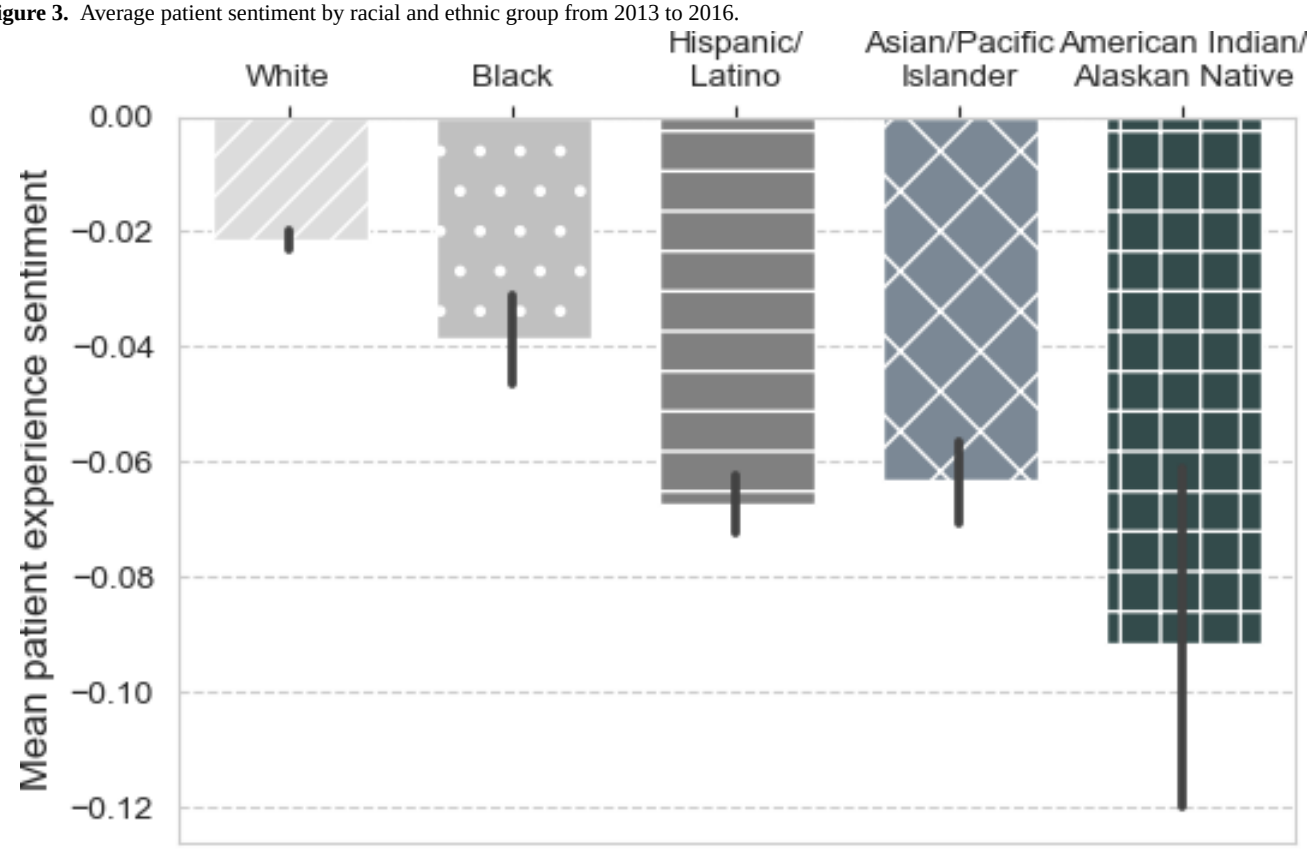
Patient Experience Sentiment Across Race and Ethnicity From 2013 to 2016

Table 1 shows the descriptive statistics of the mean, median, and standard deviation across each racial and ethnic group for each year and the overall average. The overall mean sentiment from 2013 to 2016 was highest for White patients, followed by Black, Asian/Pacific Islander, Hispanic/Latino, and lastly American Indian/Alaska Native patients, as shown graphically in Figure 3. Baseline patient experience sentiment in 2013 was the highest for White patients, followed by Black, Asian/Pacific Islander, Hispanic/Latino, and then American Indian/Alaska Native patients. In 2014, the patient experience sentiment was the highest for Black patients, followed by White, Hispanic/Latino, Asian/Pacific Islander, and then American Indian/Alaska Native patients. In 2015, the patient experience sentiment was the highest for White patients, followed by Black, Hispanic/Latino, Asian/Pacific Islander, and then American Indian/Alaska Native patients. Lastly, in 2016 the patient experience sentiment was the highest for White patients, followed by Black, Asian/Pacific Islander, Hispanic/Latino, and then American Indian/Alaska Native patients.

Table 1. Patient experience sentiment raw scores by racial and ethnic group from 2013 to 2016.

Race and year	Tweets, n	Sentiment, mean (SD)
White		
2013	109,656	–0.043 (0.478)
2014	89,540	–0.018 (0.478)
2015	63,459	–0.004 (0.480)
2016	58,079	–0.005 (0.481)
Black		
2013	5552	–0.087 (0.478)
2014	5121	0.004 (0.489)
2015	3216	–0.036 (0.484)
2016	2766	–0.024 (0.492)
Hispanic/Latino		
2013	12,825	–0.094 (0.487)
2014	9495	–0.069 (0.491)
2015	6254	–0.037 (0.490)
2016	5654	–0.038 (0.492)
Asian/Pacific Islander		
2013	5734	–0.092 (0.478)
2014	4797	–0.059 (0.480)
2015	3949	–0.056 (0.500)
2016	3708	–0.033 (0.484)
American Indian/Alaskan Native		
2013	310	–0.154 (0.456)
2014	304	–0.085 (0.477)
2015	258	–0.078 (0.465)
2016	238	–0.034 (0.485)

Figure 3. Average patient sentiment by racial and ethnic group from 2013 to 2016.



The linear regression model using race as the predictor variable is shown in Table 2. Compared with White patients, other racial and ethnic groups had statistically significant lower patient experience sentiments. American Indian/Alaska Native patients had the lowest patient experience sentiment, with a patient experience sentiment score that was 0.070 points less than White patients (P<.001). This was followed by Hispanic/Latino patients, with 0.046 fewer patient experience sentiment points (P<.001), then Asian/Pacific Islander patients, who had 0.042 fewer patient experience sentiment points (P<.001), and then Black patients, with 0.017 fewer patient experience sentiment points than White patients (P<.001). A graph of the average patient experience sentiment by race from 2013 to 2016 is depicted in Figure 3.

Table 2. Ordinary least squares modeling effects of year and racial and ethnic group on patient experience sentiment.

Predictor variables	Model 1: combined years		Model 2: yearly change		Model 3: pre-post ACA[a,b]	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Race						
White	–0.0215[c]	<.001	–28.21[c]	<.001	—[d]	—
Black	–0.0171	<.001	–9.08	.20	—	—
Hispanic/Latino	–0.0460	<.001	–14.19	.01	—	—
Asian/Pacific Islander	–0.0419	<.001	–8.77	.18	—	—
American Indian/Alaska Native	–0.0702	<.001	–46.71	.08	—	—
Change over years 2013-2016						
Race						
White	—	—	0.014[c]	<.001	—	—
Black	—	—	0.0045	.21	—	—
Hispanic/Latino	—	—	0.0070	.01	—	—
Asian/Pacific Islander	—	—	0.0043	.19	—	—
American Indian/Alaska Native	—	—	0.0232	.08	—	—
Baseline Pre-ACA (2013)						
Race						
White	—	—	—	—	–0.0433[c]	<.001
Black	—	—	—	—	–0.0433	<.001
Hispanic/Latino	—	—	—	—	–0.0487	<.001
Asian/Pacific Islander	—	—	—	—	–0.0511	<.001
American Indian/Alaska Native	—	—	—	—	–0.1112	<.001
Change Post-ACA (2014-2017)						
Race						
White	—	—	—	—	0.033[c]	<.001
Black	—	—	—	—	0.0388	<.001
Hispanic/Latino	—	—	—	—	0.0086	.28
Asian/Pacific Islander	—	—	—	—	0.0098	.08
American Indian/Alaska Native	—	—	—	—	0.0540	.09

[a]ACA: Affordable Care Act.

[b]Change from preimplementation to postimplementation of the ACA.

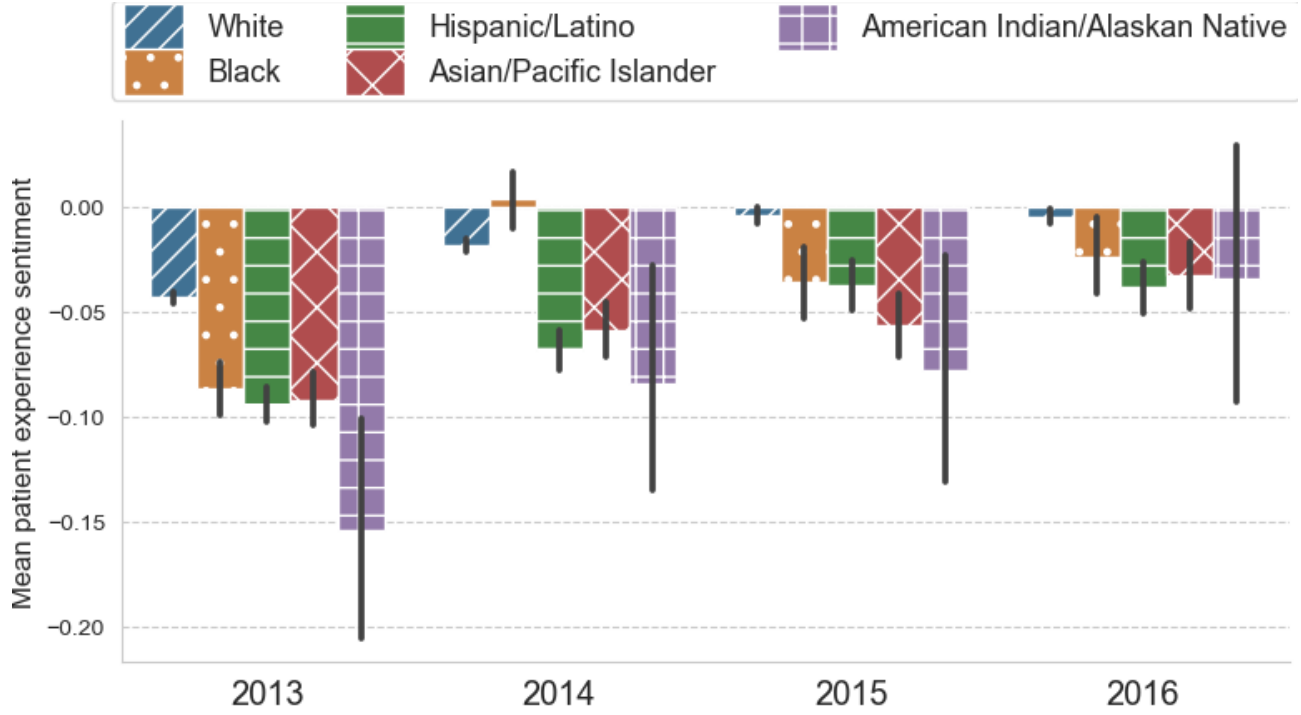
[c]Reference category.

[d]Not applicable.

Changes in Patient Experience Sentiment Across Race and Ethnicity

The changes in slope for patient experience sentiment across each racial and ethnic group are shown in regression model 2 in Table 2 and visually displayed in Figure 4. The largest increase in mean sentiment for all racial and ethnic groups was seen from 2013 to 2014, and mean sentiment continued to increase in 2015 and 2016, as illustrated in Figure 4. The ordinary squares regression model 2 showed that the yearly increase in patient experience sentiment was 0.014 points (*P*<.001), and for Hispanic/Latino patients it was 0.021 (*P*<.001). For Black patients, it was 0.0185 points (*P*=.21), for Asian/Pacific Islander patients it was 0.0183 (*P*=.19), and for American Indian/Alaska Native patients the yearly increase was 0.037 (*P*=.08).

Figure 4. Overall average patient experience sentiment by racial and ethnic group for the years from 2013 to 2016.



Patient experience increased in all racial and ethnic groups postimplementation of the ACA's full provisions (2014-2016) compared with preimplementation of the ACA (2013), which is shown in regression model 3 in Table 2. Twitter users who identified as Black experienced a 0.0718-point (P<.001) change in patient experience sentiment when comparing preimplementation and postimplementation of the ACA's full provisions. Twitter users who identified as White experienced a 0.0330-point (P<.001) change in patient experience sentiment from preimplementation to postimplementation of the ACA's full provisions.

Discussion

Monitoring Patient Experience Sentiment of Racial and Ethnic Minority Groups Using Twitter

This study used a machine learning classifier on patient experience with data captured from Twitter to measure changes in patient experience sentiment in the United States across racial and ethnic groups. Over 2.8 million tweets about patient experience were collected and a strong correlation between the distribution of users by race and ethnicity per state and the corresponding population estimates by the US census was seen. These findings showcase a promising approach for using Twitter to capture patient health care experiences of minority racial and ethnic groups. We were also able to compare patient experiences across racial and ethnic groups and found similar disparities in patient experience that have been shown in previous patient satisfaction surveys [44,45]. Twitter users who identified as White tweeted more positively about their health care experiences compared with all other racial and ethnic groups in our study. The Hospital Consumer Assessment of Healthcare Providers and Systems, a survey overseen by the Agency for Healthcare Research and Quality that asks patients to report their health care experiences, has yielded similar results, with non-Hispanic White patients reporting better experiences compared with all other racial and ethnic groups [7]. However, this same survey found that Asian patients reported the worst disparities in hospital experiences compared with White patients [7].

There is a large absence of data from American Indian/Alaska Native patients about their experiences with health care [46]. The limited studies on American Indian/Alaska Native patients have shown that they report high levels of discrimination [47], mistrust, and low satisfaction [46]. Additionally, American Indian/Alaska Native patients have the highest health burden of illness, injury, and premature mortality [48-50]. Results from our study suggest that Twitter users who identified as American Indian/Alaska Native tweeted most negatively about their patient experience, which appears consistent with existing findings. This also indicates that Twitter may be a valuable source to capture information about this vulnerable patient group that is often difficult to capture in traditional surveys [46].

Fear about the consequences of reporting negative feedback about health care experiences is a common concern because patients who do not have good experiences with their health care system or medical providers are less likely to respond to a survey administered directly by their health care institution that provided them the care [13]. On social media platforms like Twitter, users may feel unhindered in posting content and therefore share their authentic narratives and perspectives, such as their interactions with health care services, which may be especially important in the context of experiences among racial and ethnic minorities [51]. For instance, Black Twitter users have used Twitter to foster a community network to generate public conversation about their experiences. In particular, they highlight inequities and discrimination they face [51], which can include interactions with health care services. Results from our study may overlap with this community network of Black

Twitter users, whereby more negative reporting of patient experiences was observed among Black users compared with White users.

The Affordable Care Act and Patient Experiences Across Race and Ethnicity

From 2013 to 2016, tweets about patient experience declined in negative sentiment across all racial and ethnic groups. This may suggest that users experienced fewer negative experiences in health care during this time. Twitter users who identified as Hispanic/Latino experienced a 1.5 times greater decline in negative sentiment in our study than White Twitter users, a change that may be attributed to improvements in care delivery and potentially increased access and insurance coverage within this population group between the years of 2013 and 2016. The Patient Protection and Affordable Care Act's major provisions came into effect in 2014, and they specifically aimed to expand the quality of care and public health coverage for racial and ethnic minorities [52,53], who historically have been less likely to receive quality care and insurance compared with White patients. As a result of the implementation of the ACA, racial and ethnic minorities showed the largest gains in health insurance coverage [54-58]. From 2013 to 2016, a report from the Kaiser Family Foundation showed that Hispanic/Latino patients experienced the greatest decrease in uninsured rates compared with any other racial or ethnic group [58]. Greater insurance coverage among Hispanic/Latino patients may have led to improvement in patient experiences within this group. While it is difficult to draw conclusions from the findings in our study, it is promising that the greatest increases in patient experience sentiment on Twitter from 2013 to 2016 were observed among Hispanic/Latino patients, which parallels the reports of changing insurance coverage during this period. The Kaiser report also reported that Black and Asian/Pacific Islander patients experienced a reduction in uninsured rates, but this change was not as great as the change experienced among Hispanic/Latino patients [58].

Regression results showed that patient experience sentiment increased postimplementation of the ACA's full provisions (2014-2016) compared with preimplementation of the ACA (2013). While our yearly regression model identified that the highest change in sentiment by year was among Hispanic/Latino patients, Twitter users who identified as Black appeared to show the most improvement in patient experience in postimplementation compared with preimplementation of the ACA's full provisions. This change in sentiment among Black Twitter users was 2.2 times greater than the change among White Twitter users. Our results are consistent with the 2016 National Healthcare Quality and Disparities Report from the Agency for Healthcare Research and Quality, which reported a narrowing of racial and ethnic disparities in care in 2014 [59,60]. In this report, it was documented that uninsured people received worse care, but after the implementation of the Affordable Care Act's full provisions in 2014, there was an expansion of health coverage, especially among Black and Hispanic patients [1,59,61]. This decrease in the percentage of uninsured patients within these populations appears consistent with the rise in patient experience sentiment observed in our study following the full implementation of the ACA in 2014.

Limitations

Several limitations exist in our study. Certain populations may be excluded, which reduces the external validity of our results. Previous studies have indicated that a higher proportion of Twitter users are located in urban areas compared with rural areas, which may introduce further biases [62]. Additionally, when compared with the general population, Twitter is overrepresented by a younger population group aged 18 to 29 years [63,64], people who live in urban areas [43], and racial and ethnic minority groups [65]. On the other hand, specifically for this patient experience study, this sampling bias may have been beneficial for studying underrepresented minority populations [66,67]. Many disadvantaged groups have been shown to use social media and the internet to access health information, potentially as a means for overcoming their lack of access to adequate health care and health information [68,69]. For instance, Twitter has been used to document lesbian, gay, bisexual, transgender, and queer (LGBTQ) disparities in experience because of the higher representation of LGBTQ persons on Twitter compared with traditional surveys [31,32]. Our study used a nontraditional data set of free-forming discussions from Twitter that may capture a more authentic and truthful account of patient experiences than existing restricted hospital-based surveys, which are known to have issues of response and social desirability bias, especially among minority racial and ethnic groups [13].

We used a name-based classification system to identify race and ethnicity in this study. Therefore, our study may be limited, as this method may not have been able to fully capture race. For instance, names that could not be identified with a particular race or ethnicity using this classifier were excluded, and there is also the possibility that some names were identified incorrectly as the wrong race [39]. Additionally, name changes by marriage or for other legal reasons could not be accounted for with this system of classification. Furthermore, this name-based classification is based on preconceived race and ethnic group classifications defined by previous survey data from the United States Census Bureau's 2010 Decennial Census. Thus, this classifier measured race and ethnicity as a single variable, which restricts the ability to depict the multifaceted nature or the complex heterogeneity found within each racial and ethnic group [70]. Lastly, our study revealed important trends in the changes in sentiment of patient experience tweets among racial and ethnic minority groups, and these paralleled similar trends related to increasing insurance coverage and access to care among minority groups following the implementation of the full provisions of the ACA in 2014. It is promising that our findings from Twitter appear consistent with these real-world changes, though it is important to recognize that our study cannot reveal whether changing conversations on Twitter are the direct result of the widespread health policy changes.

Conclusion

Social media platforms like Twitter may provide a novel method of capturing patient experiences of racial and ethnic minority populations beyond traditional surveys. Based on our study findings on Twitter, disparities in sentiment of patient experience

across race and ethnicity appear to have declined from 2014 to 2016, which parallels similar positive changes in health care services for minority groups following the implementation of the ACA's full provisions in 2014. Continuing to monitor these patient experiences across diverse racial and ethnic groups on Twitter could be used to explore the long-term impacts of broader changes in health policies, such as the Affordable Care Act, to ensure that these disparities continue to shrink over time. Health care administrators, providers, and policy makers should consider the potential use of social media data as a method for augmenting existing measurement approaches and data sources for better understanding of patient experiences among underrepresented groups. These are particularly important steps that could inform efforts to improve the quality of care delivery among marginalized racial and ethnic minority groups.

Conflicts of Interest

None declared.

References

1. Purnell TS, Calhoun EA, Golden SH, Halladay JR, Krok-Schoen JL, Appelhans BM, et al. Achieving Health Equity: Closing The Gaps In Health Care Disparities, Interventions, And Research. Health Aff (Millwood) 2016 Aug 01;35(8):1410-1415. [doi: 10.1377/hlthaff.2016.0158] [Medline: 27503965]
2. Saha S, Arbelaez JJ, Cooper LA. Patient-physician relationships and racial disparities in the quality of health care. Am J Public Health 2003 Oct;93(10):1713-1719. [doi: 10.2105/ajph.93.10.1713] [Medline: 14534227]
3. Krieger N, Waterman PD, Chen JT, Soobader M, Subramanian S. Monitoring socioeconomic inequalities in sexually transmitted infections, tuberculosis, and violence: geocoding and choice of area-based socioeconomic measures—the public health disparities Geocoding Project (US). Public Health Reports 2003 May;118(3):240-260. [doi: 10.1016/s0033-3549(04)50245-5]
4. Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. The Lancet 2017 Apr;389(10077):1453-1463. [doi: 10.1016/s0140-6736(17)30569-x]
5. George S, Duran N, Norris K. A Systematic Review of Barriers and Facilitators to Minority Research Participation Among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health 2014 Feb;104(2):e16-e31. [doi: 10.2105/ajph.2013.301706]
6. Giordano LA, Elliott MN, Goldstein E, Lehrman WG, Spencer PA. Development, implementation, and public reporting of the HCAHPS survey. Med Care Res Rev 2010 Feb;67(1):27-37. [doi: 10.1177/1077558709341065] [Medline: 19638641]
7. Goldstein E, Elliott MN, Lehrman WG, Hambarsoomian K, Giordano LA. Racial/ethnic differences in patients' perceptions of inpatient care using the HCAHPS survey. Med Care Res Rev 2010 Feb;67(1):74-92. [doi: 10.1177/1077558709341066] [Medline: 19652150]
8. Elliott MN, Lehrman WG, Goldstein E, Hambarsoomian K, Beckett MK, Giordano LA. Do hospitals rank differently on HCAHPS for different patient subgroups? Med Care Res Rev 2010 Feb;67(1):56-73. [doi: 10.1177/1077558709339066] [Medline: 19605621]
9. Lasek RJ, Barkley W, Harper DL, Rosenthal GE. An evaluation of the impact of nonresponse bias on patient satisfaction surveys. Med Care 1997 Jun;35(6):646-652. [doi: 10.1097/00005650-199706000-00009] [Medline: 9191708]
10. Burroughs TE, Waterman BM, Gilin D, Adams D, McCollegan J, Cira J. Do On-Site Patient Satisfaction Surveys Bias Results? The Joint Commission Journal on Quality and Patient Safety 2005 Mar;31(3):158-166. [doi: 10.1016/s1553-7250(05)31021-x]
11. Perneger TV, Chamot E, Bovier PA. Nonresponse bias in a survey of patient perceptions of hospital care. Med Care 2005 Apr;43(4):374-380. [doi: 10.1097/01.mlr.0000156856.36901.40] [Medline: 15778640]
12. Richman WL, Kiesler S, Weisband S, Drasgow F. A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. Journal of Applied Psychology 1999 Oct;84(5):754-775. [doi: 10.1037/0021-9010.84.5.754]
13. Mazor KM, Clauser BE, Field T, Yood RA, Gurwitz JH. A demonstration of the impact of response bias on the results of patient satisfaction surveys. Health Serv Res 2002 Oct;37(5):1403-1417 [FREE Full text] [doi: 10.1111/1475-6773.11194] [Medline: 12479503]
14. van Ryn M, Fu SS. Paved with good intentions: do public health and human service providers contribute to racial/ethnic disparities in health? Am J Public Health 2003 Feb;93(2):248-255. [doi: 10.2105/ajph.93.2.248] [Medline: 12554578]
15. Greenfield S, Kaplan SH, Kahn R, Ninomiya J, Griffith JL. Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessment results. Ann Intern Med 2002 Jan 15;136(2):111-121. [doi: 10.7326/0003-4819-136-2-200201150-00008] [Medline: 11790062]
16. Saunders CL, Elliott MN, Lyratzopoulos G, Abel GA. Do Differential Response Rates to Patient Surveys Between Organizations Lead to Unfair Performance Comparisons? Medical Care 2016;54(1):45-54. [doi: 10.1097/mlr.0000000000000457]

17. Huynh KP, Jacho-Chávez DT, Tripathi G. The econometrics of complex survey data: Theory and applications. Bingley, United Kingdom: Emerald Group Publishing; 2019.
18. The Patient Protection and Affordable Care Act. 2010. URL: https://www.govinfo.gov/content/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf [accessed 2020-08-09]
19. Rosenbaum S. The Patient Protection and Affordable Care Act: implications for public health policy and practice. Public Health Rep 2011;126(1):130-135 [FREE Full text] [doi: 10.1177/003335491112600118] [Medline: 21337939]
20. Federal Subsidies for Health Insurance Coverage for People Under Age 65: 2016 to 2026.: Congressional Budget Office; 2016. URL: https://www.cbo.gov/sites/default/files/114th-congress-2015-2016/reports/51385-healthinsurancebaseline.pdf [accessed 2020-08-11]
21. Millenson M, Marci J. Will the Affordable Care Act move patient-centeredness to center stage? Urban Institute 2012 Mar:1-10. [doi: 10.1037/e552122012-001]
22. Epstein RM, Fiscella K, Lesser CS, Stange KC. Why the nation needs a policy push on patient-centered health care. Health Aff (Millwood) 2010 Aug;29(8):1489-1495. [doi: 10.1377/hlthaff.2009.0888] [Medline: 20679652]
23. Clement J. Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019.: Statistica; 2019 Aug 14. URL: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ [accessed 2017-04-16]
24. Twitter by the Numbers: Stats, Demographics & Fun Facts.: OmniCore Agency URL: https://www.omnicoreagency.com/twitter-statistics/ [accessed 2018-02-23]
25. Crannell WC, Clark E, Jones C, James TA, Moore J. A pattern-matched Twitter analysis of US cancer-patient sentiments. J Surg Res 2016 Dec;206(2):536-542. [doi: 10.1016/j.jss.2016.06.050] [Medline: 27523257]
26. Hswen Y, Naslund JA, Brownstein JS, Hawkins JB. Monitoring Online Discussions About Suicide Among Twitter Users With Schizophrenia: Exploratory Study. JMIR Ment Health 2018 Dec 13;5(4):e11483 [FREE Full text] [doi: 10.2196/11483] [Medline: 30545811]
27. Hswen Y, Naslund JA, Brownstein JS, Hawkins JB. Online Communication about Depression and Anxiety among Twitter Users with Schizophrenia: Preliminary Findings to Inform a Digital Phenotype Using Social Media. Psychiatr Q 2018 Sep;89(3):569-580 [FREE Full text] [doi: 10.1007/s11126-017-9559-y] [Medline: 29327218]
28. Hswen Y, Gopaluni A, Brownstein JS, Hawkins JB. Using Twitter to Detect Psychological Characteristics of Self-Identified Persons With Autism Spectrum Disorder: A Feasibility Study. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e12264 [FREE Full text] [doi: 10.2196/12264] [Medline: 30747718]
29. Du J, Xu J, Song H, Tao C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):69 [FREE Full text] [doi: 10.1186/s12911-017-0469-6] [Medline: 28699569]
30. Hswen Y, Naslund JA, Chandrashekar P, Siegel R, Brownstein JS, Hawkins JB. Exploring online communication about cigarette smoking among Twitter users who self-identify as having schizophrenia. Psychiatry Res 2017 Nov;257:479-484 [FREE Full text] [doi: 10.1016/j.psychres.2017.08.002] [Medline: 28841509]
31. Hswen Y, Sewalk KC, Alsentzer E, Tuli G, Brownstein JS, Hawkins JB. Investigating inequities in hospital care among lesbian, gay, bisexual, and transgender (LGBT) individuals using social media. Soc Sci Med 2018 Oct;215:92-97. [doi: 10.1016/j.socscimed.2018.08.031] [Medline: 30219749]
32. Hswen Y, Zhang A, Sewalk K, Tuli G, Brownstein JS, Hawkins JB. Use of social media to assess the impact of equitable state policies on LGBTQ patient experiences: An exploratory study. Healthc (Amst) 2020 Jun;8(2):100410. [doi: 10.1016/j.hjdsi.2020.100410] [Medline: 32241681]
33. Perrin A. Social Media Usage: 2005-2015.: Pew Research Center URL: https://www.pewresearch.org/internet/2015/10/08/social-networking-usage-2005-2015/ [accessed 2020-07-24]
34. Duggan M, Brenner J. The demographics of social media users. Pew Research Center's Internet & American Life Project Washington, DC 2012:1-3.
35. Sewalk KC, Tuli G, Hswen Y, Brownstein JS, Hawkins JB. Using Twitter to Examine Web-Based Patient Experience Sentiments in the United States: Longitudinal Study. J Med Internet Res 2018 Oct 12;20(10):e10043 [FREE Full text] [doi: 10.2196/10043] [Medline: 30314959]
36. Hutto CJ, Gilbert E. A parsimonious rule-based model for sentiment analysis of social media text. Vader 2014:216-225.
37. Gilbert CHE. A parsimonious rule-based model for sentiment analysis of social media text. Vader 2014:216-225.
38. Kralj Novak P, Smailović J, Sluban B, Mozetič I. Sentiment of Emojis. PLoS One 2015;10(12):e0144296 [FREE Full text] [doi: 10.1371/journal.pone.0144296] [Medline: 26641093]
39. Mateos P. A review of name-based ethnicity classification methods and their potential in population studies. Popul Space Place 2007 Jul;13(4):243-263. [doi: 10.1002/psp.457]
40. Piazza A, Rendine S, Zei G, Moroni A, Cavalli-Sforza LL. Migration rates of human populations from surname distributions. Nature 1987;329(6141):714-716. [doi: 10.1038/329714a0] [Medline: 3670373]
41. M'charek A. The Human Genome Diversity Project: an ethnography of scientific practice. Cambridge, UK: Cambridge University Press; 2005:1-208.
42. Comenetz J. Frequently occurring surnames in the 2010 census.: US Census Bureau; 2010. URL: https://www2.census.gov/topics/genealogy/2010surnames/surnames.pdf [accessed 2020-08-11]

43. Mislove A, Lehmann S, Ahn YY, Onnela JP, Rosenquist JN. Understanding the Demographics of Twitter Users. 2011 Presented at: 5th Annual AAAI Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Spain.
44. Rodriguez HP, von Glahn T, Grembowski DE, Rogers WH, Safran DG. Physician effects on racial and ethnic disparities in patients' experiences of primary care. J Gen Intern Med 2008 Oct;23(10):1666-1672 [FREE Full text] [doi: 10.1007/s11606-008-0732-8] [Medline: 18651194]
45. Zhu J, Weingart SN, Ritter GA, Tompkins CP, Garnick DW. Racial/Ethnic Disparities in Patient Experience With Communication in Hospitals. Medical Care 2015;53(5):446-454. [doi: 10.1097/mlr.0000000000000350]
46. Katz RJ. Addressing the health care needs of American Indians and Alaska Natives. Am J Public Health 2004 Jan;94(1):13-14. [doi: 10.2105/ajph.94.1.13] [Medline: 14713686]
47. Moghaddam JF, Momper SL, Fong T. Discrimination and participation in traditional healing for American Indians and Alaska Natives. J Community Health 2013 Dec;38(6):1115-1123. [doi: 10.1007/s10900-013-9721-x] [Medline: 23821254]
48. Sarche M, Spicer P. Poverty and health disparities for American Indian and Alaska Native children: current knowledge and future prospects. Ann N Y Acad Sci 2008;1136:126-136 [FREE Full text] [doi: 10.1196/annals.1425.017] [Medline: 18579879]
49. Jones DS. The Persistence of American Indian Health Disparities. Am J Public Health 2006 Dec;96(12):2122-2134. [doi: 10.2105/ajph.2004.054262]
50. Castor ML, Smyser MS, Taualii MM, Park AN, Lawson SA, Forquera RA. A Nationwide Population-Based Study Identifying Health Disparities Between American Indians/Alaska Natives and the General Populations Living in Select Urban Counties. Am J Public Health 2006 Aug;96(8):1478-1484. [doi: 10.2105/ajph.2004.053942]
51. Clark M. To tweet our own cause: A mixed-methods study of the online phenomenon "Black Twitter". Carolina Digital Repository 2014 Dec:1-174.
52. Sommers BD, Gunja MZ, Finegold K, Musco T. Changes in Self-reported Insurance Coverage, Access to Care, and Health Under the Affordable Care Act. JAMA 2015 Jul 28;314(4):366-374. [doi: 10.1001/jama.2015.8421] [Medline: 26219054]
53. Uberoi N, Finegold K, Gee E. Health insurance coverage and the Affordable Care Act 2010-2016. Washington, DC: Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation; 2010. URL: https://aspe.hhs.gov/system/files/pdf/187551/ACA2010-2016.pdf [accessed 2020-08-11]
54. Lillie-Blanton M, Hoffman C. The role of health insurance coverage in reducing racial/ethnic disparities in health care. Health Aff (Millwood) 2005;24(2):398-408. [doi: 10.1377/hlthaff.24.2.398] [Medline: 15757923]
55. Clemans-Cope L, Kenney GM, Buettgens M, Carroll C, Blavin F. The Affordable Care Act's coverage expansions will reduce differences in uninsurance rates by race and ethnicity. Health Aff (Millwood) 2012 May;31(5):920-930. [doi: 10.1377/hlthaff.2011.1086] [Medline: 22566430]
56. Chen J, Vargas-Bustamante A, Mortensen K, Ortega AN. Racial and Ethnic Disparities in Health Care Access and Utilization Under the Affordable Care Act. Medical Care 2016;54(2):140-146. [doi: 10.1097/mlr.0000000000000467]
57. Artiga S. Health coverage by race and ethnicity: the potential impact of the Affordable Care Act.: Kaiser Family Foundation; 2014. URL: https://www.kff.org/wp-content/uploads/2014/07/8423-health-coverage-by-race-and-ethnicity.pdf [accessed 2020-08-11]
58. Artiga S, Ubri P, Foutz J, Damico A. Health Coverage by Race and Ethnicity: Examining Changes Under the ACA and the Remaining Uninsured.: Kaiser Family Foundation; 2016. URL: http://files.kff.org/attachment/Issue-Brief-Health-Coverage-by-Race-and-Ethnicity-Examining-Changes-Under-the-ACA-and-the-Remaining-Uninsured [accessed 2020-08-11]
59. Fiscella K, Sanders MR. Racial and Ethnic Disparities in the Quality of Health Care. Annu Rev Public Health 2016;37:375-394. [doi: 10.1146/annurev-publhealth-032315-021439] [Medline: 26789384]
60. Agency for Healthcare Research and Quality. 2016 National Healthcare Quality and Disparities Report. 2016. URL: https://www.ahrq.gov/research/findings/nhqrdr/nhqdr16/index.html#:~:text=The%20report%20is%20based%20on,health%20care%20services%20and%20settings.&text=The%20report%20is%20produced%20with,Disparities%20Report%20Data%20and%20Tools [accessed 2020-08-11]
61. Burstin H, Leatherman S, Goldmann D. The evolution of healthcare quality measurement in the United States. J Intern Med 2016 Feb;279(2):154-159 [FREE Full text] [doi: 10.1111/joim.12471] [Medline: 26785953]
62. Fink C, Bos A, Perrone A, Liu E, Kopecky J. Twitter, Public Opinion, and the 2011 Nigerian Presidential Election. 2011 Presented at: 2013 International Conference on Social Computing (SocialCom); Sep 8-14, 2011; Alexandria, VA p. 311-320. [doi: 10.1109/SocialCom.2013.50]
63. Madden M, Lenhart A, Cortesi A, Gasser U, Duggan M, Smith A, et al. Teens, Social Media, and Privacy.: Pew Research Center; 2013. URL: https://www.pewresearch.org/internet/2013/05/21/teens-social-media-and-privacy/ [accessed 2020-08-11]
64. Social Media Fact Sheet.: Pew Research Center URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2018-11-10]
65. Oktay H, Firat A, Ertem Z. Demographic breakdown of twitter users: An analysis based on names. Academy of Science and Engineering (ASE) 2014:A.
66. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. Annu Rev Public Health 2006;27:1-28. [doi: 10.1146/annurev.publhealth.27.021405.102113] [Medline: 16533107]

67. Wendler D, Kington R, Madans J, Van Wye G, Christ-Schmidt H, Pratt LA, et al. Are racial and ethnic minorities less willing to participate in health research? PLoS Med 2006 Feb;3(2):e19 [FREE Full text] [doi: 10.1371/journal.pmed.0030019] [Medline: 16318411]
68. Mesch G, Mano R, Tsamir J. Minority status and health information search: a test of the social diversification hypothesis. Soc Sci Med 2012 Sep;75(5):854-858. [doi: 10.1016/j.socscimed.2012.03.024] [Medline: 22633160]
69. Gayet-Ageron A, Agoritsas T, Schiesari L, Kolly V, Perneger TV. Barriers to participation in a patient satisfaction survey: who are we missing? PLoS One 2011;6(10):e26852 [FREE Full text] [doi: 10.1371/journal.pone.0026852] [Medline: 22046382]
70. Connolly H, Gardener D. Who are the Other ethnic groups Social and welfare reports. Office for National Satistics, London 2006:1-2.

Abbreviations

ACA: Affordable Care Act
LGBTQ: lesbian, gay, bisexual, transgender, and queer
VADER: Valence Aware Dictionary for Sentiment Reasoner

Edited by G Eysenbach; submitted 14.11.19; peer-reviewed by T Bernardo, A Cyr; comments to author 31.12.19; revised version received 28.05.20; accepted 21.06.20; published 21.08.20.

Please cite as:
Hswen Y, Hawkins JB, Sewalk K, Tuli G, Williams DR, Viswanath K, Subramanian SV, Brownstein JS
Racial and Ethnic Disparities in Patient Experiences in the United States: 4-Year Content Analysis of Twitter
J Med Internet Res 2020;22(8):e17048
URL: http://www.jmir.org/2020/8/e17048/
doi:10.2196/17048
PMID:32821062

Original Paper

Global Infodemiology of COVID-19: Analysis of Google Web Searches and Instagram Hashtags

Alessandro Rovetta[1*]; Akshaya Srikanth Bhagavathula[2*], PharmD

[1]Research and Disclosure Division, Mensana srls, Brescia, Italy

[2]Institute of Public Health, College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, United Arab Emirates

[*]all authors contributed equally

Corresponding Author:
Alessandro Rovetta
Research and Disclosure Division
Mensana srls
Via Moro Aldo 5
Brescia, 25124
Italy
Phone: 39 3927112808
Email: rovetta.mresearch@gmail.com

Abstract

Background: Although "infodemiological" methods have been used in research on coronavirus disease (COVID-19), an examination of the extent of infodemic moniker (misinformation) use on the internet remains limited.

Objective: The aim of this paper is to investigate internet search behaviors related to COVID-19 and examine the circulation of infodemic monikers through two platforms—Google and Instagram—during the current global pandemic.

Methods: We have defined *infodemic moniker* as a term, query, hashtag, or phrase that generates or feeds fake news, misinterpretations, or discriminatory phenomena. Using Google Trends and Instagram hashtags, we explored internet search activities and behaviors related to the COVID-19 pandemic from February 20, 2020, to May 6, 2020. We investigated the names used to identify the virus, health and risk perception, life during the lockdown, and information related to the adoption of COVID-19 infodemic monikers. We computed the average peak volume with a 95% CI for the monikers.

Results: The top six COVID-19–related terms searched in Google were "coronavirus," "corona," "COVID," "virus," "corona virus," and "COVID-19." Countries with a higher number of COVID-19 cases had a higher number of COVID-19 queries on Google. The monikers "coronavirus ozone," "coronavirus laboratory," "coronavirus 5G," "coronavirus conspiracy," and "coronavirus bill gates" were widely circulated on the internet. Searches on "tips and cures" for COVID-19 spiked in relation to the US president speculating about a "miracle cure" and suggesting an injection of disinfectant to treat the virus. Around two thirds (n=48,700,000, 66.1%) of Instagram users used the hashtags "COVID-19" and "coronavirus" to disperse virus-related information.

Conclusions: Globally, there is a growing interest in COVID-19, and numerous infodemic monikers continue to circulate on the internet. Based on our findings, we hope to encourage mass media regulators and health organizers to be vigilant and diminish the use and circulation of these infodemic monikers to decrease the spread of misinformation.

(J Med Internet Res 2020;22(8):e20673) doi:10.2196/20673

KEYWORDS

COVID-19; coronavirus; Google; Instagram; infodemiology; infodemic; social media

Introduction

Globally, the internet is an extremely important platform for obtaining knowledge and information about the coronavirus disease (COVID-19) pandemic [1-3]. The Google Trends tool provides real-time insights into internet search behaviors on various topics, including COVID-19 [4]. Social media platforms such as Facebook, Twitter, and Instagram allow users to communicate their thoughts, feelings, and opinions by sharing short messages. A unique aspect of social media data from Instagram is that image-based posts are accessible, and the use of topic-related hashtags allows access to topic-related information for users [5]. In general, there is a growing interest

in examining social data to understand and monitor public behavior in real time [6,7].

Research on the internet and social data are called *infodemiology* or *infoveillance* studies [8]. Infodemiology is defined as "the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy" [9]. Although several studies have been conducted using infodemiological methods in COVID-19 research, a limited number of studies have examined the extent of COVID-19–related misinformation on the internet [10-14]. We are defining an *infodemic moniker* to be a term, query, hashtag, or phrase that generates or feeds the misinformation circulating on the internet. These monikers can profoundly affect public health communication, giving rise to errors in interpretation, misleading information, xenophobia, and fake news [12-17]. In this context, we aimed to investigate the internet search behaviors related to COVID-19 and the extent of infodemic monikers circulating on Google and Instagram during the current pandemic period.

Methods

We used Google Trends and Instagram hashtags to explore internet search activities and behaviors related to the COVID-19 pandemic from February 20, 2020, to May 06, 2020. We investigated the following: names used to identify the virus, health and risk perception, lifestyles during the lockdown, and information related to the adoption of infodemic monikers related to COVID-19. The complete list of terms used to identify the most frequently searched queries in Google and hashtag suggestions for Instagram are presented in Multimedia Appendix 1.

The obtained infodemic monikers are characterized as follows:

1. *Generic*: The moniker confuses, due to a lack of specificity.
2. *Misinformative*: The moniker associates a certain phenomenon with fake news.
3. *Discriminatory*: The moniker encourages the association of a problem with a specific ethnicity and/or geographical region.
4. *Deviant*: The moniker does not identify the requested phenomenon.
5. *Other specificities*: We kept two additional points for special cases that prove to be exceptionally serious.

To determine the severity of the various infodemic monikers circulating on the internet, each moniker was assigned 1 to 2 points on the infodemic scale (I-scale) ranging from 0 (minimum) to 10 (maximum). Based on the sum of the I-scale scores, the infodemic monikers were classified as follows: not infodemic (0), slightly infodemic (1), moderately infodemic (2-4), highly infodemic (5-8), and extremely infodemic (9-10).

To assign points, we have adopted the following procedure:

- *Generic*: 1 point is assigned if the keyword is a scientific term but gives rise to misunderstanding (eg, "COVID" instead of "COVID-19"); 2 points are assigned if the keyword is a combination of two scientific terms that can be confused with previously used terms (eg, "SARS-CoV" instead of "SARS-CoV-2" or "SARS-COVID" instead of "SARS-CoV-2" and "COVID-19").
- *Misinformative*: 1 point is assigned if the keyword can lead to both fake news pointing to individuals (eg, "coronavirus Bill Gates"); 2 points are assigned if the keyword is used to spread misinformation using unrelated or not officially confirmed sources (eg, "coronavirus laboratory").
- *Discriminatory*: 1 point is assigned if the keyword refers to a specific country and incites unfounded, racial fear (eg, "coronavirus China"); 2 points are assigned if the keywords explicitly target a specific ethnicity (eg, "Chinese coronavirus").
- *Deviant*: 1 point is assigned if the keyword expresses opinions to influence public opinion (eg, "ban china" or "china app"); 2 points are assigned if the keyword expresses a particular attitude to influence the public (eg, "china puppets" or "savage WHO").
- *Other specificities*: 1 additional point is assigned when the adoption of a certain moniker is associated with real facts but involves serious health or economic risks (eg, "uv coronavirus"); 2 additional points are assigned when the adoption of a certain moniker involves only health risks (eg, "no sew mask" or "anti-mask protest").

For each search keyword considered, Google Trends provided normalized data in the form of relative search volume (RSV) based on search popularity scale ranging from 0 (low) to 100 (highly popular). Using these RSV values, we computed the average peak volume (APC) with a 95% CI (for a Gaussian distribution) during the study period.

Instagram, a platform for image-based posts with hashtags, was also screened. We retrieved content based on hashtags through image classifiers every 3-4 days during the study period. All irrelevant content was excluded. The data collected included contents posted on Instagram and self-reported user demographic information. No personal information, such as emails, phone numbers, or addresses, was collected. The data from the Instagram hashtags were collected manually through the Instagram-suggested tags associated with specific countries.

All data used in the study were obtained from anonymous open sources. Thus, ethical approval was not required.

Results

The top five COVID-19–related infodemic and scientific terms used in Google searches were "coronavirus," "corona," "COVID," "virus," "corona virus," and "COVID-19" (Figure 1). The most frequently used keywords globally were "coronavirus" (APC=1378, 95% CI 1246-1537), followed by "corona" (APC=530, 95% CI 477-610) and "COVID" (APC=345, 95% CI 292-398). Several keywords related to COVID-19 were identified (Table 1); of the top 10, four had an I-scale value >4: "corona," "virus," "corona virus," and "coronavirus China."

Figure 1. Top scientific and infodemic names related to coronavirus disease (COVID-19) in Google searches globally. RSV: relative search volume.

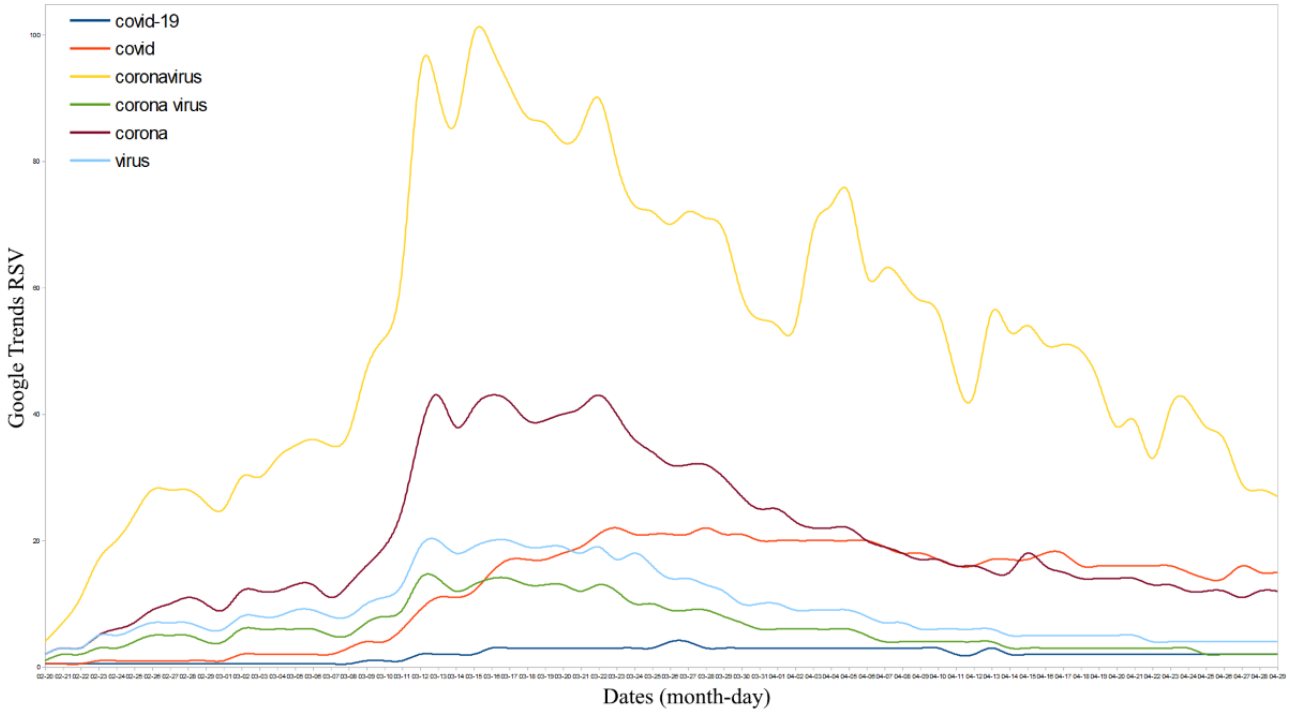


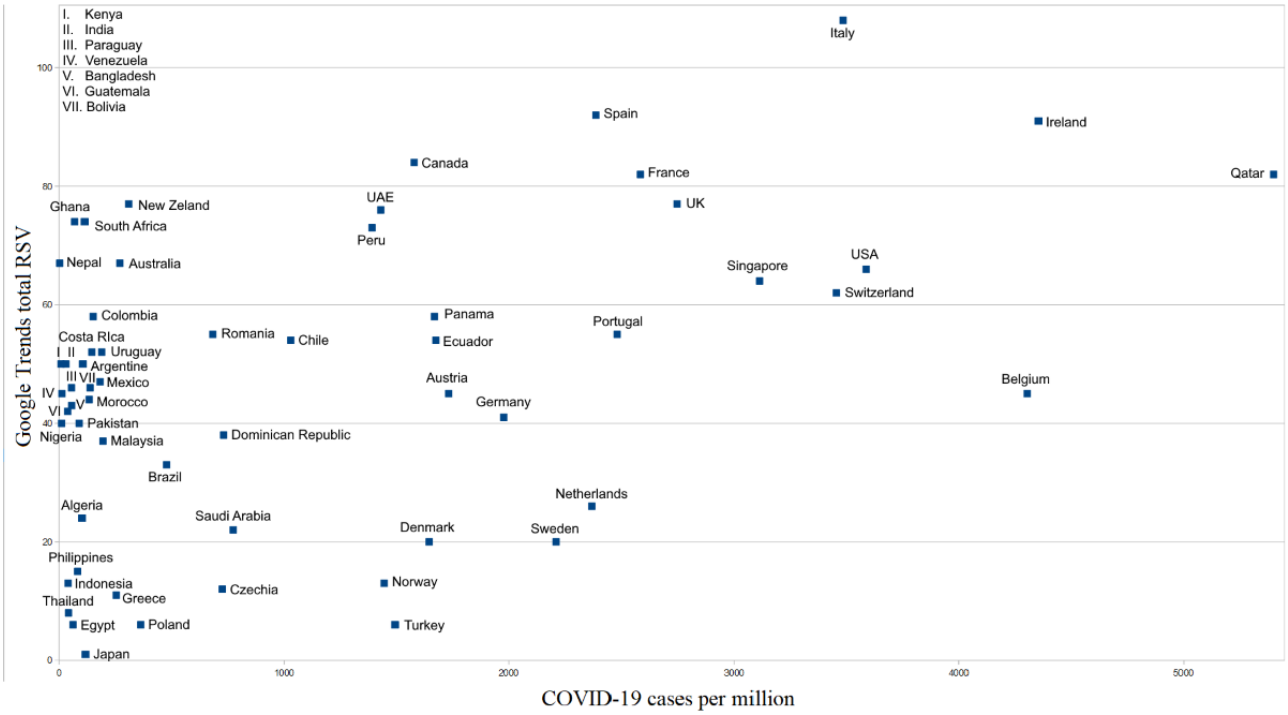
Table 1. Top infodemic and scientific Google searches related to coronavirus disease (COVID-19) globally.

Keyword	Total average peak volume	95% CI	I-scale[a] score
coronavirus	1378	1246-1537	4
corona	530	477-610	8
COVID	345	292-398	1
virus	239	212-292	6
corona virus	159	133-186	7
coronavirus Italy	54	45-62	4
COVID-19	53	45-60	0
coronavirus USA	32	29-36	4
coronavirus China	30	25-34	6
coronavirus Germany	23	20-27	4
corona Italy	13	12-14	7
corona Deutschland	12	10-14	7
SARS	9	8-10	5
corona China	9	7-11	9
corona Wuhan	1	0-2	9
SARS-CoV-2	1	0-1	0

[a]I-scale: infodemic scale ranging from 0-10.

The country-wise dispersion of the scientific and infodemic names of COVID-19 used in Google searches are shown in Figure 2. Countries with a higher number of COVID-19 cases per 1 million people had greater Google search interest related to COVID-19 (eg, Italy, Spain, Ireland, Canada, France, and Qatar). These COVID-19–related search queries were significantly correlated with the incidence of COVID-19 cases in these countries (Pearson R=0.45, P<.001).

Figure 2. Country-wise dispersion of the scientific and infodemic names of coronavirus disease (COVID-19). RSV: relative search volume.



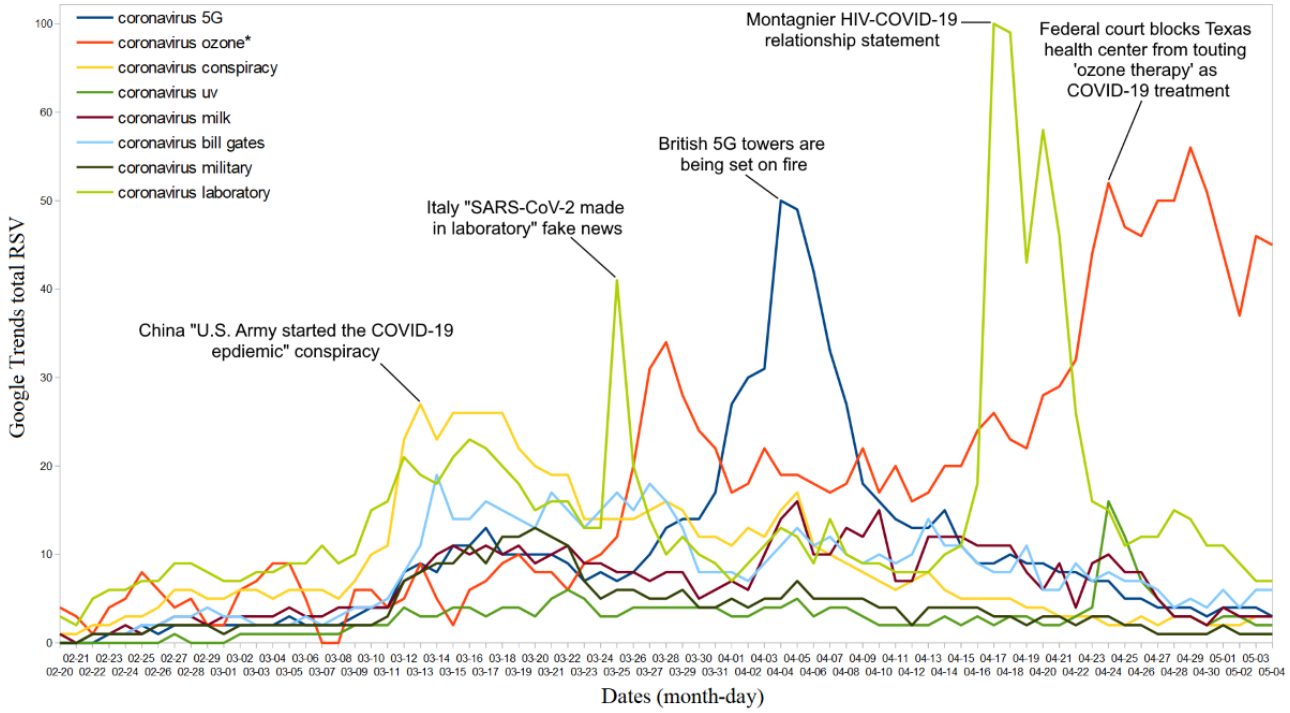
The top COVID-19 monikers related to fake news (eg, “coronavirus ozone,” “coronavirus laboratory,” and “coronavirus 5G”) that frequently circulated on the internet are presented in Table 2. Among these, “coronavirus conspiracy” and “coronavirus laboratory” had the highest I-scale scores globally (score=9). Additionally, the use of monikers with moderate-to-high infodemicity far exceeded the use of scientific names (Table 1)—52% of Google web searches were moderately infodemic (totalAPC=1487, 95% CI 1326-1647) and 34% highly infodemic (total APC=992, 95% CI 933-1050). The circulation of these infodemic monikers was further examined to understand the events associated with these searches. Infodemic monikers related to coronavirus origins, such as “SARS-CoV-2 made in the laboratory,” went viral (APC=41) when the National Associated Press Agency (Agenzia Nazionale Stampa Associata) of Italy posted a 2015 video about the origins of SARS-CoV-2 virus on March 25, 2020 [18]. In addition, the moniker reached breakout levels (RSV=100) on April 17, 2020, when the French Noble Prize winner Professor Luc Montagnier stated that the new coronavirus was the result of a laboratory accident in a high-security laboratory in Wuhan [19]. Detailed information on the different infodemic monikers and associated events are shown in Figure 3.

Table 2. Top global fake news–related Google searches on coronavirus disease (COVID-19).

Keyword	Total average peak volume	95% CI	I-scale[a] score
coronavirus ozone	19	15-22	5
coronavirus laboratory	16	12-19	9
coronavirus 5G	10	8-13	8
coronavirus conspiracy	9	8-11	9
coronavirus bill gates	8	7-10	6
coronavirus milk	7	6-8	8
coronavirus military	4	4-5	8
coronavirus uv	3	3-4	5

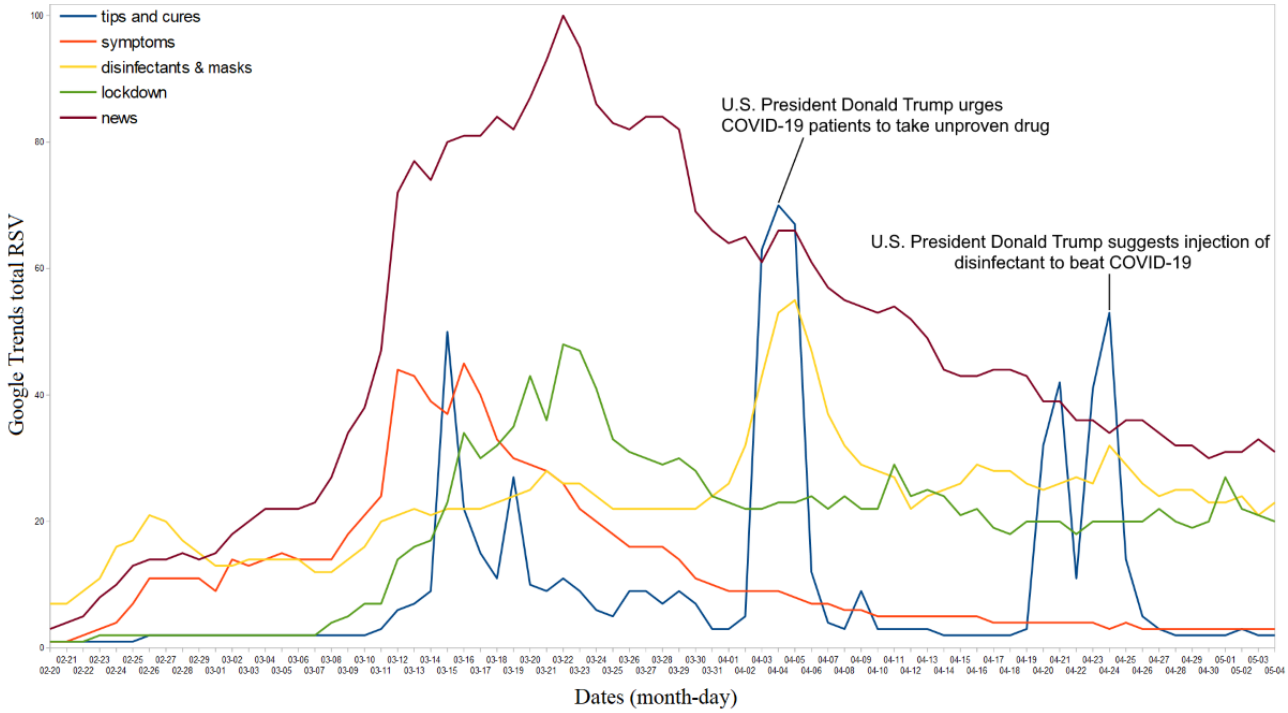
[a]I-scale: infodemic scale ranging from 0-10.

Figure 3. Top global web searches related to coronavirus disease (COVID-19), rated "high" or "extreme" on the infodemic scale, and associated events. RSV: relative search volume.



The top searches related to health, precautions, and COVID-19 news are presented in Figure 4. Google searches related to COVID-19 news remained at the top throughout the pandemic period. However, searches related to “tips and cures” for COVID-19 spiked multiple times when the US president suggested that hydroxychloroquine (an unproven drug) was a “miracle cure” on April 4, 2020 (RSV=70) [20]; he also suggested injecting disinfectant to treat COVID-19 on April 24, 2020 (RSV=53) [21]. Other searches related to the use of medical masks and disinfectants (APC=23, 95% CI 21-25), lockdown (APC=19, 95% CI 16-22), and COVID-19 symptoms (APC=12, 95% CI 10-15) were less frequently used in Google searches.

Figure 4. Top global web searches related to health, precautions, and coronavirus disease (COVID-19) news. RSV: relative search volume.



The top five nations most cited by global Instagram users in relation to COVID-19 were Italy (963,000 hashtags), Brazil (551,000 hashtags), Spain (376,000 hashtags), Indonesia (298,000 hashtags), and Turkey (244,000 hashtags) (Table 3). Over 2 million hashtags categorized as extremely infodemic, such as “corona,” “corona memes,” “coronado,” and “corona

time," have been in circulation. Nevertheless, the most globally used hashtags were those related to health and prevention, such as "stay home/safe" and "lockdown life" (>165 million). The contribution of the "covid-19" hashtag for COVID-19–related information was 35.6% (n=26,200,000), followed by "coronavirus" (n=22,500,000, 30.5%), "corona" (n=18,800,000, 25.6%), and "COVID" (n=5,900,000, 8%) (Table 4).

Table 3. Top 10 Instagram hashtags related to coronavirus disease (COVID-19), as of May 6, 2020.

Rank	Country	Monikers/hashtags, n[a]	Virus		Health	
			Monikers/hashtags	n[a]	Monikers/hashtags	n[a]
1	Italy	9.63	covid-19	306	health-stay home/safe	933
2	Brazil	5.51	coronavirus	267	lockdown life	718
3	Spain	3.76	corona	188	masks	135
4	Indonesia	2.98	covid	69	memes	25
5	Turkey	2.44	corona memes	14	gym/fitness	24
6	India	1.65	coronavirus Italy	9.63	art/hobbies	22
7	Malaysia	0.89	coronado	8.19	cooking	21
8	Dominican Republic	0.83	corona time	7.12	fashion	16
9	United States	0.75	coronavirus memes	6.41	hair/beard style	14
10	Argentina	0.74	coronavirus Brazil	5.51	fun/party	13

[a]Multiples in 100,000.

Table 4. Top Instagram hashtags related to coronavirus disease (COVID-19) scientific and infodemic names.

Hashtag	Count, n (%)
#2019nCOV	38,993 (0.1)
#SARSCoV2	49,011 (0.1)
#SARS	53,633 (0.1)
#COVID	5,890,625 (8.0)
#corona	18,849,998 (25.6)
#coronavirus	22,458,007 (30.5)
#COVID19	26,213,280 (35.6)

Discussion

Principal Findings

In light of the ongoing COVID-19 pandemic, we are the first to investigate the internet search behaviors of the public and the extent of infodemic monikers circulating on Google and Instagram globally. Our results suggest that (a) "coronavirus," "corona," "COVID," "virus," "corona virus," and "COVID-19" are the top five terms used in the Google searches; (b) countries (eg, Italy, Spain, Ireland, Canada, and France) with a high incidence of COVID-19 cases (per million) have greater Google search queries about COVID-19; (c) "coronavirus ozone," "coronavirus laboratory," "coronavirus 5G," "coronavirus conspiracy," and "coronavirus bill gates" are widely used infodemic monikers on the internet; (d) although COVID-19 news remains at the top, web searches related to "tips and cures" for COVID-19 spiked when the US president speculated about a "miracle cure" and the use of a disinfectant injection to treat COVID-19; (e) 66.1% (n=48,700,000) of Instagram users used "COVID-19" and "coronavirus" as a hashtag to disperse information related to COVID-19.

Exploring research using nontraditional data sources such as social media has several implications. First, our results demonstrated a potential application for the use of Instagram as a complementary tool to aid in understanding online search behavior; we also provided real-time tracking of infodemic monikers circulating on the internet. The strength of this study is the investigation of various infodemic monikers dispersed on the internet and correlating them with the events associated with that particular day. Although we used correlations to examine the possible linear association between search queries and the event, it should be noted that use of a search engine is voluntary and self-initiated search queries represent the users who are truly curious or worried about a situation. Thus, we believe that the unobtrusive search behavior of netizens may have resulted in an increase in search volume. By characterizing and classifying various infodemic monikers based on the degree of infodemicity (ie, via the I-scale), researchers can foster new methods of using social media data to monitor the monikers' outcomes. The analysis and methods used in this study could aid public health and communication agencies in identifying and diminishing infodemic monikers circulating on the internet.

Findings from this study validate and extend previously published works that used Google keywords [1,12,13]. We also demonstrate the potential for the use of Instagram hashtags to monitor and predict both the cyber behavior and relaying of misinformation on the internet [22-24]. In 2017, Guidry et al [22] studied Ebola-related risk perception in Instagram users and identified that a significant proportion of posts had rampant misinformation about the Ebola disease during the outbreak. In addition, the percentage of Instagram posts and tweets posted by health organizations (eg, Centers for Disease Control and Prevention, World Health Organization, Médecins Sans Frontières [Doctors Without Borders]) that correct misinformation is less than 5% [22]. In general, negative information posted on the internet tends to receive a greater weight among netizens. Thus, this should be counter-balanced with evidence-based content from health organizations, particularly in the current pandemic situation. For example, when the US president suggested injecting disinfectant to treat COVID-19, the number of Google searches considering it as a cure sharply increased (APC=53) and resulted in 30 cases of disinfectant poisoning within 18 hours in New York City [25]. Health authorities should be vigilant and provide more positive and informative messages to combat the circulation of infodemic monikers on social media. Future studies will need to investigate the influence of infodemic monikers on individual cyber behavior.

Limitations

Our study used Google Trends, which only provides the search behavior of people using the Google search engine. Furthermore, our study focused on Google and Instagram for data retrieval. Future studies should consider studying the same topic on other social media platforms to capture a more diverse population of users. Instagram searches were conducted manually, introducing a potential for human error. Going forward, the use of an automated program can improve the accuracy of the data collected and analyzed. Lastly, Google Trends did not provide any information about the methodology used to generate search data and algorithms.

Conclusion

Using Google Trends and Instagram hashtags, the present study identified that there is a growing interest in COVID-19 globally and in countries with a higher incidence of the virus. Searches related to “COVID-19 news” are quite frequent and two thirds of Instagram users have used “COVID-19” and “coronavirus” as hashtags to disperse information related to the virus. Several infodemic monikers are circulating on the internet, with “coronavirus conspiracy” and “coronavirus laboratory” identified as the most dangerous (I-scale score=9). Given the prevalence of infodemic monikers, mass media regulators and health organizers should monitor and diminish the impact of these monikers. These governing bodies should also be encouraged to take serious actions against those spreading misinformation in social media.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Terms used to identify the most frequently searched queries in Google and the hashtag suggestions for Instagram.
[DOCX File , 14 KB - jmir_v22i8e20673_app1.docx]

References

1. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn Y, Simon K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. Proc Natl Acad Sci U S A 2020 May 26;117(21):11220-11222 [FREE Full text] [doi: 10.1073/pnas.2005335117] [Medline: 32366658]
2. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P. Association of the COVID-19 pandemic with Internet Search Volumes: A Google Trends Analysis. Int J Infect Dis 2020 Jun;95:192-197 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.033] [Medline: 32305520]
3. Lin Y, Liu C, Chiu Y. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun 2020 Jul;87:30-32 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.020] [Medline: 32283286]
4. COVID-19 Information & Resources. Google. 2020. URL: https://www.google.com/covid19/ [accessed 2020-05-20]
5. Giannoulakis S, Tsapatsoulis N. Evaluating the descriptive power of Instagram hashtags. Journal of Innovation in Digital Ecosystems 2016 Dec;3(2):114-129. [doi: 10.1016/j.jides.2016.10.001]
6. Salathé M. Digital epidemiology: what is it, and where is it going? Life Sci Soc Policy 2018 Jan 04;14(1):1 [FREE Full text] [doi: 10.1186/s40504-017-0065-7] [Medline: 29302758]
7. Chaffey D. Global social media research summary July 2020. Smart Insights. 2020 Aug. URL: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/ [accessed 2020-08-14]
8. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
9. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]

10. Hernández-García I, Giménez-Júlvez T. Assessment of Health Information About COVID-19 Prevention on the Internet: Infodemiological Study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717 [FREE Full text] [doi: 10.2196/18717] [Medline: 32217507]
11. Park HW, Park S, Chong M. Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea. J Med Internet Res 2020 May 05;22(5):e18897 [FREE Full text] [doi: 10.2196/18897] [Medline: 32325426]
12. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the Internet: Infodemiology Study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]
13. Rovetta A, Bhagavathula AS. COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. JMIR Public Health Surveill 2020 May 05;6(2):e19374 [FREE Full text] [doi: 10.2196/19374] [Medline: 32338613]
14. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
15. Shimizu K. 2019-nCoV, fake news, and racism. The Lancet 2020 Feb 29;395(10225):685-686. [doi: 10.1016/S0140-6736(20)30357-3] [Medline: 32059801]
16. Chung RY, Li MM. Anti-Chinese sentiment during the 2019-nCoV outbreak. The Lancet 2020 Feb 29;395(10225):686-687 [FREE Full text] [doi: 10.1016/S0140-6736(20)30358-5] [Medline: 32122469]
17. Haynes S. As Coronavirus Spreads, So Does Xenophobia and Anti-Asian Racism. Time. 2020 Feb. URL: https://time.com/5797836/coronavirus-racism-stereotypes-attacks/ [accessed 2020-04-09]
18. Coronavirus: Il caso del video del Tgr Leonardo 2015 sul supervirus creato in Cina. ANSA. 2020 Mar. URL: https://www.ansa.it/sito/notizie/politica/2020/03/25/coronavirus-il-caso-del-video-del-tgr-leonardo-2015-sul-supervirus-creato-in-cina_7adf8316-6ca5-42cd-96de-c18f7fb53595.html [accessed 2020-04-30]
19. COVID-19: la théorie d'un virus fabriqué vivement contestée. La Presse. 2020 Apr. URL: https://www.lapresse.ca/actualites/sciences/202004/17/01-5269764-covid-19-la-theorie-dun-virus-fabrique-vivement-contestee.php [accessed 2020-04-30]
20. Grady D, Kannapell A. Trump Urges Coronavirus Patients to Take Unproven Drug. The New York Times. 2020 Apr. URL: https://www.nytimes.com/2020/04/04/health/coronavirus-drug-trump-hydroxycholoroquine.html [accessed 2020-04-30]
21. Clark D. Trump suggests 'injection' of disinfectant to beat coronavirus and 'clean' the lungs. NBC News. 2020 Apr. URL: https://www.nbcnews.com/politics/donald-trump/trump-suggests-injection-disinfectant-beat-coronavirus-clean-lungs-n1191216 [accessed 2020-04-30]
22. Guidry J, Jin Y, Orr C, Messner M, Meganck S. Ebola on Instagram and Twitter: How health organizations address the health crisis in their social media engagement. Public Relations Review 2017 Sep;43(3):477-486. [doi: 10.1016/j.pubrev.2017.04.009]
23. Zarei K, Farahbakhsh R, Crespi N, Tyson G. A first Instagram dataset on COVID-19. arXiv 2020:e [FREE Full text]
24. Gupta R, Ariefdjohan M. Mental illness on Instagram: a mixed method study to characterize public content, sentiments, and trends of antidepressant use. J Ment Health 2020 Apr 23:1-8. [doi: 10.1080/09638237.2020.1755021] [Medline: 32325006]
25. Glatter R. Calls to poison centers spike after the president comments about using disinfectants to treat coronavirus. Forbes. 2020 Apr. URL: https://www.forbes.com/sites/robertglatter/2020/04/25/calls-to-poison-centers-spike--after-the-presidents-comments-about-using-disinfectants-to-treat-coronavirus.html [accessed 2020-05-15]

Abbreviations

APC: average peak volume
COVID-19: coronavirus disease
I-scale: infodemic scale
RSV: relative search volume

Edited by G Eysenbach, G Fagherazzi; submitted 25.05.20; peer-reviewed by W Aldhaleei, A Amresh, A Zink; comments to author 13.07.20; revised version received 02.08.20; accepted 03.08.20; published 25.08.20.

Please cite as:
Rovetta A, Bhagavathula AS
Global Infodemiology of COVID-19: Analysis of Google Web Searches and Instagram Hashtags
J Med Internet Res 2020;22(8):e20673
URL: http://www.jmir.org/2020/8/e20673/
doi:10.2196/20673
PMID:32748790

Original Paper

Assessing Public Opinion on CRISPR-Cas9: Combining Crowdsourcing and Deep Learning

Martin Müller[1*], MSc; Manuel Schneider[2*], MSc; Marcel Salathé[1], PhD; Effy Vayena[2], PhD

[1]Digital Epidemiology Lab, School of Life Sciences, School of Computer and Communication Sciences, EPFL, Geneva, Switzerland

[2]Health Ethics and Policy Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

[*]these authors contributed equally

Corresponding Author:
Manuel Schneider, MSc
Health Ethics and Policy Lab
Department of Health Sciences and Technology
ETH Zurich
Hottingerstrasse 10
Zurich, 8092
Switzerland
Phone: 41 44 632 26 16
Email: manuel.schneider@hest.ethz.ch

Abstract

Background: The discovery of the CRISPR-Cas9–based gene editing method has opened unprecedented new potential for biological and medical engineering, sparking a growing public debate on both the potential and dangers of CRISPR applications. Given the speed of technology development and the almost instantaneous global spread of news, it is important to follow evolving debates without much delay and in sufficient detail, as certain events may have a major long-term impact on public opinion and later influence policy decisions.

Objective: Social media networks such as Twitter have shown to be major drivers of news dissemination and public discourse. They provide a vast amount of semistructured data in almost real-time and give direct access to the content of the conversations. We can now mine and analyze such data quickly because of recent developments in machine learning and natural language processing.

Methods: Here, we used Bidirectional Encoder Representations from Transformers (BERT), an attention-based transformer model, in combination with statistical methods to analyze the entirety of all tweets ever published on CRISPR since the publication of the first gene editing application in 2013.

Results: We show that the mean sentiment of tweets was initially very positive, but began to decrease over time, and that this decline was driven by rare peaks of strong negative sentiments. Due to the high temporal resolution of the data, we were able to associate these peaks with specific events and to observe how trending topics changed over time.

Conclusions: Overall, this type of analysis can provide valuable and complementary insights into ongoing public debates, extending the traditional empirical bioethics toolset.

(J Med Internet Res 2020;22(8):e17830) doi:10.2196/17830

KEYWORDS

CRISPR; natural language processing; sentiment analysis; digital methods; infodemiology; infoveillance; empirical bioethics; social media

Introduction

Genome editing has many potential applications, ranging from gene therapy [1] to crop enhancement [2] and production of biomolecules [3,4]. While it has been possible to modify the genomes of eukaryotic cells since the 1980s, traditional methods have proven to be rather impractical, inaccurate, or impossible to use at scale [5-8]. Accurately targeted gene editing has only become possible within the last decade [9,10] using a CRISPR-Cas9–based method. In 2013, the method was further developed to be used on human cells [11,12], which allowed for the first successful experiment to alter the human germline

DNA of non-viable embryos in April 2015 [13]. The experiment, conducted by a group of Chinese scientists, raised ethical concerns among researchers and the general public about the potential far-reaching consequences of introducing germline modifications [14,15]. Such ethical concerns include unexpected side effects on the evolution of humans, as well as cultural and religious arguments. In November 2018, Jiankui He announced the genetic editing of two viable human embryos with the goal of introducing HIV resistance [16]. The work came to be known to a global public under the term “CRISPR babies” and was condemned by the scientific community as unethical, unnecessary, and harmful to the two babies [17,18].

As the costs of the technology drop further and usage becomes more widespread, governments and policy makers are faced with the challenging task of posing adequate ethical restrictions to prevent misuse. To gain time to introduce appropriate ethical frameworks, some scientists have called for a moratorium on genetically editing the human germline [19-21]. Previous studies on opinion towards GMO plants highlight how certain events or scandals (eg, with respect to food safety) may have a major long-term impact on public opinion and later drive policy decisions [22-25]. Understanding the public attitudes towards topics such as CRISPR is therefore of paramount importance for policy making [26,27].

Several surveys have been conducted with the goal of evaluating the public’s perception of CRISPR and genetic engineering in general [28-32]. Such surveys have found that participants are largely in favor of the technology used for somatic purposes (eg, in the context of treatment) but less so for germline editing, especially if this is not for clearly medical purposes. Additionally, the studies underline certain demographic correlations (eg, that women, people belonging to ethnic minorities, and religious communities are more critical about the potential applications of CRISPR [28,30]). Somewhat unsurprisingly, the surveys also show that public views are not always aligned with expert opinions [32]. A recent study that explored coverage of news articles on CRISPR in North America between 2012 and 2017 found CRISPR to be overwhelmingly portrayed as positive and potentially overhyped in news media compared to the public’s views [33].

Social media platforms allow people to discuss a topic online with other people around the globe, creating an abundance of semistructured conversational data. Sentiment analysis provides a way to study people’s perception of a topic, based on personal statements, and to process large volumes of such data in an automated way. Sentiment analysis has been used in the past to analyze different features such as emotions and polarity in several different contexts [34]. While traditional methods are based on linguistic expert knowledge (eg, rule-based methods), newer methods leverage machine learning, can be trained for specific contexts, and dominate traditional methods on polarity classification tasks [35]. Additionally, the supervised machine learning approaches have the advantage that the performance of a model for the specific context can be evaluated. The adaption to a specific context is particularly useful for tweets, which have a very specific, informal language [36]. Accordingly, machine learning methods have been successfully used for Twitter sentiment analysis [37,38]. Most classical supervised machine learning algorithms for text classification (such as Naive Bayes or support vector machines [SVMs]) rely on manual feature extraction. Recently, a type of semisupervised machine learning model called Bidirectional Encoder Representations from Transformers (BERT) has been introduced to natural language processing [39]. BERT models are pretrained on large corpuses of raw text and can be adapted to a target task in a process called transfer learning. BERT models are based on the transformer, a neural network architecture that has been shown to outperform previously mentioned models in most natural language processing tasks, including text classification and sentiment analysis [40,41]. BERT has also been used in top-ranking submissions in the SemEval2019 challenges on detection of hate speech and offensive language in social media data [42,43].

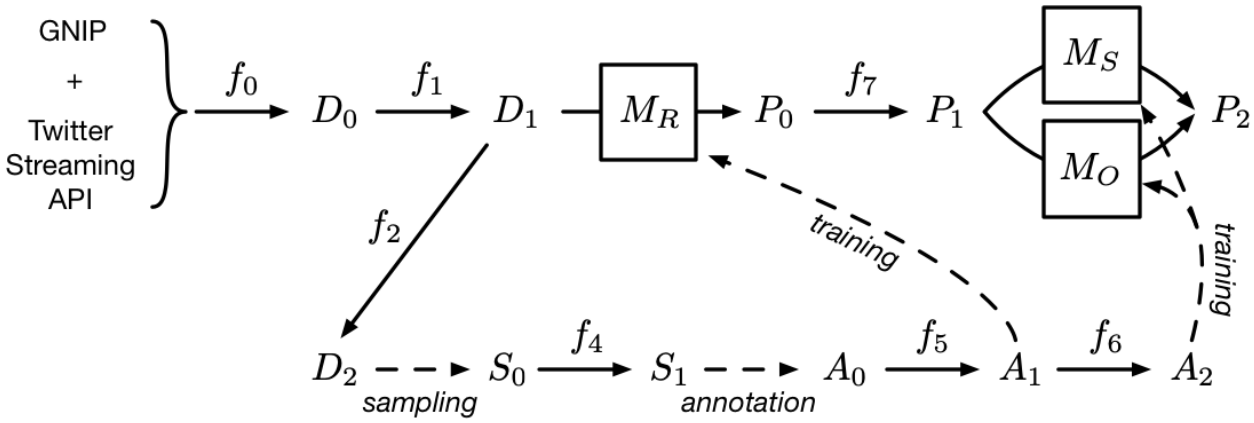
In this study, we conducted the first analysis of a complete dataset of all tweets about CRISPR published over a 6.5-year period. The analyzed timespan includes the first experiment of CRISPR on human cells in 2013 but also recent events, such as the first genetic editing of viable human babies in November 2018. Furthermore, we make use of recent advances in text classification models, such as BERT [39], which use semisupervised machine learning to generate a high-resolution temporal signal of the sentiment towards CRISPR over the observed timespan. By combining multiple text classification methods, we obtain results that can also be linked back to previous studies conducted with traditional methods, such as surveys.

Methods

Overview

Our analysis consisted of 4 different explorative approaches, all of which build upon the sentiments of the tweets. Therefore, sentiment analysis represents the core of our analysis. In order to determine the sentiment for the entirety of tweets published over the last 6.5 years, we trained a predictive model on a previously manually annotated subset of the data. The process can be divided into 5 main tasks, which we describe in the following sections (see Figure 1 for an overview of the process): data collection, preparation, annotation, training, and analysis.

Figure 1. Overview of the data processing pipeline. Labels f_{0-7} denote filtering steps, D_{0-2} datasets, S_{0-1} samples, A_{0-2} annotation sets, and P_{0-2} predictions. M_R, M_S, and M_O represent machine learning models. API: application programming interface.



Data Collection

The data set (denoted as D_0 in Figure 1) for our analysis consists of all tweets (including retweets, quoted tweets, replies, and mentions) that match the character sequence CRISPR (in any capitalization), have been detected to be in English language, and were published between January 1, 2013 and May 31, 2019. We retrieved these data either through the Twitter Streaming API or through GNIP, a Twitter subsidiary that allows access to historical data that were not retrievable through the Twitter Streaming API. The 3 aforementioned filtering conditions were used as parameters in the retrieval through Twitter APIs (denoted as f_0) as well as for the requested data from GNIP.

The number of tweets varied greatly over time, ranging from 4818 in 2013 to 445,744 in 2018, totaling 1,508,044 tweets by 348,502 distinct users (also refer to Multimedia Appendix 1). Since the focus was on the overall evolution of the discourse provided by aggregated information, this study considered only the text in the tweet objects and ignored user-related information (such as location) or media content (such as photos or videos). In addition, any occurrences of Twitter handles and URLs in the text were anonymized (replaced by @<user> and <url>, respectively) to protect individuals.

Preparation

In a preparatory step, tweets suitable for annotation were selected from D_0. As an inclusion criterion, only tweets with ≥3 English words (after removal of stop words) were considered (f_1). Although a tweet with <3 non-stop words may express a sentiment, we chose this threshold to ensure that the annotators had at least a minimal context to determine if the tweet was in fact relevant to the topic and what sentiment it expressed. The word count was determined by the help of NLTK's (Natural Language Toolkit, a python library for natural language processing) TweetTokenizer and English word and stop word corpora [44]. The filtering and subsequent dataset operations and analysis were carried out using pandas, a python package for data analysis [45]. The resulting dataset D_1 (n=1,334,114) was used as the basis for the subsequent analysis. To avoid the annotation of duplicates, all retweets, quoted tweets, and other duplicates of tweets with the same text were removed, leading to dataset D_2 (n=433,930).

Next, we selected a random sample S_0 (n=29,238), so that we obtained a more or less evenly distributed number of tweets over the observed timespan. This was achieved by binning the data by all 77 months and selecting a constant number of tweets from each monthly bin. In contrast to a fully random sample, our sampling scheme contained no oversampling bias with regard to very recent content. Therefore, the generated sample was more representative of the whole observation period and accounted for the possibility that the nature of the tweets changed notably over time.

Annotation

After generating the sample, the selected tweets were annotated through the Crowdbreaks platform [38,46], which uses crowdsourcing to annotate social media data. The platform allows for the creation of a question sequence that is then submitted in combination with a tweet as a task to MTurk (Amazon Mechanical Turk) [47]. The question sequence contained 3 questions for each task. The first question was on the relevance of the tweet to the topic of CRISPR-Cas9, allowing "relevant" and "not relevant" as possible answers. The second question was on the sentiment (positive, negative, or neutral), and the third question was on the organism (humans, human embryos, animals [other than human], plants, bacteria, multiple, not specified).

Before submitting the task to MTurk, the availability of the tweet was automatically checked. This was done in order to respect the user's right to either delete their content or set it to private after the time of data collection. Filtering by tweets that were still available yielded the sample S_1 (n=22,513), which was subsequently annotated with regard to the 3 questions mentioned earlier. This resulted in annotation set A_0. To detect workers with questionable performance, the annotators' raw agreement was calculated, which denotes the fraction of the number of actual agreements over the number of possible agreements an annotator had with other annotators. An annotator was considered an outlier if this value was larger than 3 standard deviations from the mean, the annotator had less than 20 possible agreements with other annotators, or the annotator was involved

in less than 3 separate tasks. All annotations by outlier annotators were subsequently removed. The resulting Fleiss' kappa agreement scores [48] were 0.81 and 0.28 for the questions of relevance and sentiment, respectively. Tweets for which a unanimous consensus of at least 3 independent annotators could be found were merged into dataset A_1. For the questions on sentiment and organism, only tweets that were labelled as relevant were considered and exported to A_2. This resulted in 3 cleaned datasets with annotated tweets for relevance (n=16,421), sentiment (n=4718), and organism (n=1196), which we used to train 3 classifiers.

Training

In order to classify the data with regard to relevance, sentiment, and organism, we constructed 3 classifiers: M_R, M_S, and M_O, respectively. The classifiers tried to predict the respective labels from the text of the tweet alone. In the process, we analyzed the performance of 4 different classifier models: Bag of Words (BoW), Sent2Vec sentence embeddings [49] coupled with SVMs [50], FastText [51], and BERT [39]. The tokenization process was different for each model class. In order to evaluate the models, the cleaned annotation data were shuffled and split into training (80%) and test sets (20%).

For the BoW, SVM, and FastText models, we used supervised learning to train the 3 classifiers for sentiment, relevance, and organisms. A limited search of model parameters was conducted. In the case of BERT, we started from the pretrained (unsupervised) English BERT-large-uncased model provided by the Huggingface library [52] and conducted an additional step of unsupervised, domain-specific pretraining on our raw body of tweets. This model then served as the basis for the final, supervised training step (ie, fine-tuning the general model with classifier-specific labelled data). For this fine-tuning step, a learning rate of 1e-05 and 2 epochs of training were used. This work was conducted using PyTorch [53] and the Huggingface library [52].

After the training phase, we selected the classifiers for relevance, sentiment, and organism (M_R, M_S, and M_O in Figure 1) by evaluating the performance of the models on the test set (see Multimedia Appendix 2 for different model performances). The fine-tuned BERT model was the best performing sentiment classifier (M_S), with a macro-averaged F1 score of 0.727 ($F1_{positive}$=0.827, $F1_{neutral}$=0.715, $F1_{negative}$=0.639). The fine-tuned BERT model was also found to be the best performing model for the relevance (M_R) and organism (M_O) classifiers with macro-averaged F1 scores of 0.91 ($F1_{related}$=0.997, $F1_{unrelated}$=0.823) and 0.89 ($F1_{humans}$=0.873, $F1_{embryos}$=0.762, $F1_{animals}$=1, $F1_{plants}$=0.889, $F1_{bacteria}$=0.909, $F1_{unspecified}$=0.902), respectively.

Prediction

For the analysis, the best performing model (fine-tuned BERT) for relevance M_R was used to predict dataset D_1 and yield the predicted dataset P_0 (n=1,334,114) of the same length containing a label for relevance. Next, all tweets predicted as not relevant were removed from P_0, yielding the dataset P_1 (n=1,311,544). This dataset was then used to predict sentiment and organism using the models M_S and M_O, resulting in the final dataset P_2.

Analysis

In our analysis, we used the sentiments in relation to tweet activity (number of tweets), topics of the tweets (hashtags), organisms the tweets were talking about (predicted), and themes identified from previous studies on CRISPR mentioned earlier (through regular expressions) to gain different kinds of insights. Wherever we used sentiments for numerical calculations, we used +1 for positive, 0 for neutral, and −1 for negative sentiment. Further, we extrapolated the numbers for 2019 where applicable for better comparison since we only had data until May 31, 2019. The different parts of the analysis are explained in more detail in the following paragraphs.

The first part of the analysis was concerned with the development of the sentiment in relation to the number of tweets over time. The detection of a temporary deviation from the general sentiment was of particular interest. While we included all tweets for the analysis of activity, we excluded tweets with neutral sentiment for the analysis of sentiment to make deviations more visible. We aggregated activity and sentiments on a daily basis. For the sentiments, however, the sentiment value of a specific day was determined by taking the mean value of all positive and negative sentiments within a sliding 7-day window centered around that day (±3 days). Further, we tested whether the yearly means based on the positive and negative tweet sentiments were significantly different from each other with the Welch's *t*-test [54,55] using scipy's statistics module [56]. We then used scipy's module for peak detection [56] to detect events of interest, using a relative prominence cut-off of 0.2. In order to identify potential sources for the change in sentiment, we manually identified major events that relate to CRISPR.

In the second part, we used the predictions of the model M_O and the sentiments to compare the development of the sentiment for different organisms. We calculated the mean sentiments over a month and excluded all months that did not have at least 100 tweets for the respective organism. Further, we used the same test as we did for the yearly means to compare the organism class means based on the individual tweet sentiments (positive, negative, and neutral).

Third, we analyzed hashtags as a proxy for the topics a user was talking about in his or her tweet. The hashtag #CRISPR was excluded from the analysis since CRISPR was the overarching topic all tweets had in common. We counted the occurrences of every hashtag per year. We used the exact hashtags and did not group similar hashtags. For example, the hashtags #crisprbaby and #crisprbabies were treated as different hashtags. We did this due to the difficulty of automatically matching similar hashtags, since they can be a composition of multiple words that made strategies like stemming not straightforward. For each hashtag and year, we then calculated the mean sentiment and selected the 15 most common hashtags for each year for further analysis. We then manually compared how these top 15 topics per year increased and decreased in popularity

throughout the years, as well as how the sentiments for these topics changed.

In the fourth and last part of our analysis, we based our analysis on the earlier conducted studies. We conducted a literature search in scientific databases according to a predefined search strategy (see Multimedia Appendix 3). The search was conducted in the fall of 2017. We reviewed the resulting studies and identified the reasons why people had a positive or negative attitude towards CRISPR and issues that concerned them. In the process, we summarized these reasons and concerns for each study and compiled a list with a short description for each of them. Since there was thematic overlap across the studies, we inductively determined the themes of these summaries and compiled a regular expression representing each theme based on the summary text. Additionally, we added themes and corresponding regular expressions based on publications and events that occurred between the fall of 2017 and the summer of 2019. The regular expressions then allowed us to automatically check for matches on the entire Twitter dataset as a proxy for the presence of the themes that occurred in the studies. See Multimedia Appendix 4 for the themes and regular expressions.

Results

Overview

Our analysis includes over 1,300,000 tweets (dataset P_1, n=1,311,544) over the time period from January 1, 2013 until May 31, 2019. The predicted sentiments of the tweets were predominantly positive (685,578/1,311,544, 52.3%) or neutral (528,196/1,311,544, 40.3%). Only a minor fraction was predicted as negative (97,770/1,311,544, 7.5%). In the following sections, we report our results focusing on different aspects.

Temporal Development

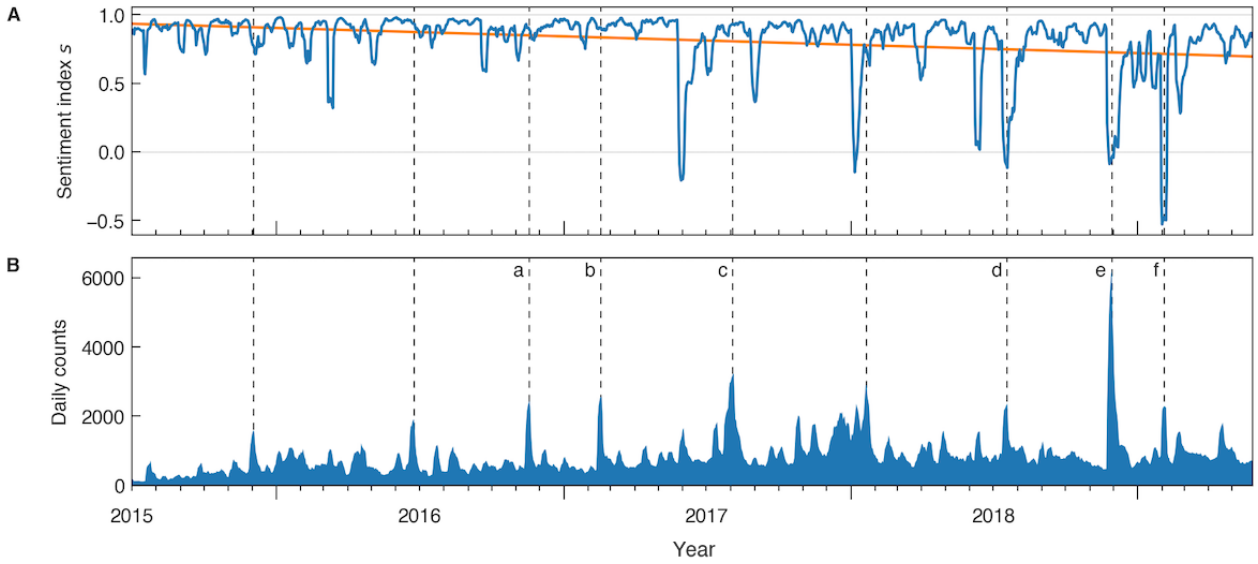
Figure 2 shows a temporal analysis of the predicted sentiments in relation to key historical events surrounding CRISPR. A sentiment of zero indicates an equal portion of positive and negative tweets, and the values 1 and −1 indicate a signal with only positive or negative tweets, respectively. Figure 2A shows the sentiments between July 2015 and June 2019. The time period before July 2015 was excluded, as activity was too low for a high-resolution sentiment signal. The sentiment remained mostly positive, with an average of 85% positive tweets and only 15% negative tweets. Especially over the initial time period until March 2017, the sentiment shows little variation. After that, the sentiment reveals a series of sharp negative spikes, on multiple occasions dropping below zero. Over the observed time period, the sentiment shows a slight negative trend (slope of $-0.061\ y^{-1}$, standard error $0.005\ y^{-1}$), as indicated by the linear trend line in orange. The differences between the yearly means of the tweet sentiments were all significant ($P<.001$; see Multimedia Appendix 5 for all means, standard deviations, and test statistics).

We then compared the sentiment curve to the observed activity surrounding CRISPR in the same time span, as shown in Figure 2B. Shown are the mean daily counts of the sample P_1 over a sliding window of 7 days. Activity varied considerably, with an average baseline of about 1000 tweets per day and peaks of up to roughly 6000 tweets per day.

We detected 9 peaks of interest. They are marked with dashed lines in Figure 2. When comparing peaks of high activity to the sentiment, it can be seen that peaks of high activity before mid-2018 did not result in a negative sentiment response. Peaks of strong negative sentiment started to appear in 2017 but it was not accompanied by the same level of activity until after 2018.

In a second step, major news events were manually mapped to coinciding peaks (for a full list, see Multimedia Appendix 6). A subset of these peaks was marked with letters a-f in Figure 2B for illustrative purposes. In all cases, the most retweeted tweet within days of the peak was linking a news article describing the event. The events include the first use of CRISPR in humans by a group of Chinese scientists in November 2016 (peak a) and the US Patent Office deciding in favor of the Broad Institute (peak b). Both of these events did not lead to a significant change in sentiment. Peak c coincides with the publication of a study that reported the correction of a mutation in human embryos [57], causing widespread media attention and, as before, did not cause a drop in sentiment. However, in July 2018, a study by the Wellcome Sanger Institute [58] warned about serious side effects, such as cancer, that CRISPR could have when used in humans (peak d). This peak led to a clear negative response in the sentiment index and marks the first negative peak with high media attention. When researcher He Jiankui revealed creating the world's first genetically edited babies in November 2018 [16] (peak e), the highest activity was recorded. Although He's revelation caused a strong negative signal, the strongest negative sentiment was recorded shortly after, in February 2019 (peak f). This event coincides with the re-emergence of a news story from August 2017 when biohackers managed to encode a malware program into a strand of DNA [59].

Figure 2. A) Predicted sentiment towards CRISPR between July 2015 and June 2019. The blue curve denotes the sentiment s, which is calculated as the mean of the weighted counts of positive and negative tweets over a centered rolling window of 7 days. The orange curve denotes a linear fit of the sentiment s. B) Daily counts of all analyzed tweets. The blue area shows the daily sum of positive, negative, and neutral tweets as the mean within a 7-day centered rolling window. All peaks above a relative prominence of 0.2 are marked with dashed lines; a-f denote peaks that coincide with certain events.

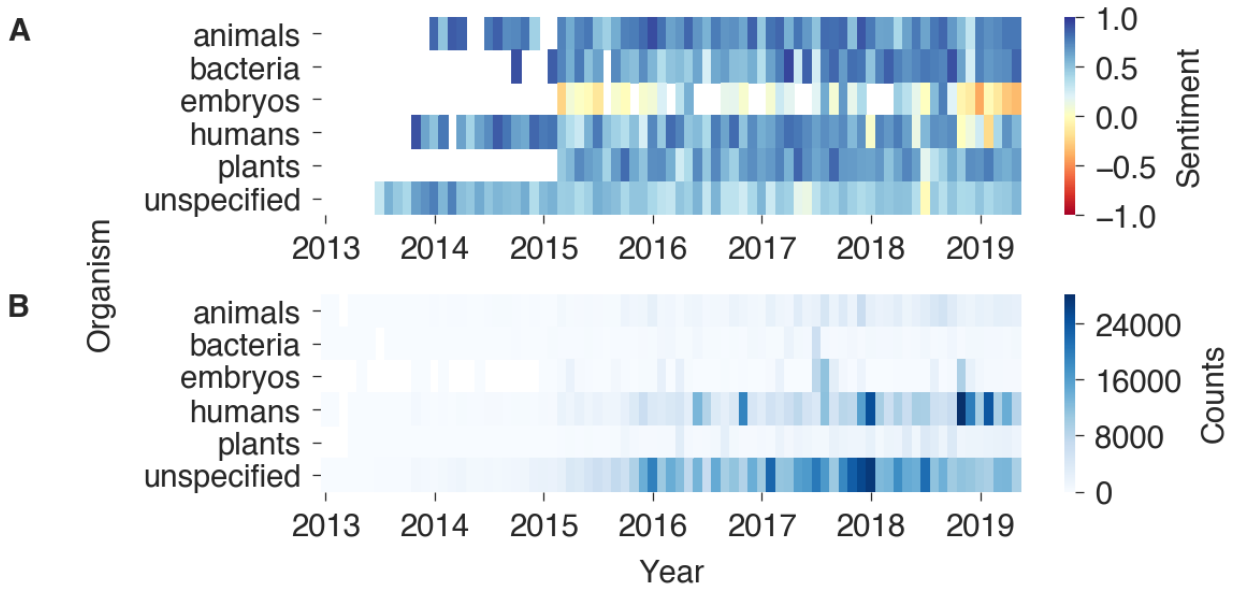


Organisms

In order to improve our understanding of the sentiment signal, the data were predicted with respect to which organism each tweet was about (see the Methods section). We predicted the organism of the tweets in the dataset P_1 (n=1,311,544) resulting in the classes animals (7.6%), bacteria (2.4%), embryos (4.3%), humans (30.3%), plants (4.9%), and unspecified (50.6%). It is noteworthy that more than half of all tweets do not specifically refer to an organism in the context of CRISPR. After unspecified, the class humans is the second largest group, followed with some margin by animals (eg, mice for animal testing), plants, and embryos. The classes humans and embryos combined account for a little more than one-third of all tweets. Tweets specifically mentioning CRISPR in the context of bacteria were rather rare.

Figure 3A shows the monthly sentiment for each organism class, which are based on the monthly counts shown in Figure 3B (all monthly means and standard deviations can be found in Multimedia Appendix 7). Of all classes, embryos exhibited the most negative-leaning sentiment (mean sentiment 0.14 over all monthly means) and was also the class with the strongest variations between months (SD 0.27). Further, a relatively high sentiment was measured for the classes animals (mean 0.70, SD 0.14), bacteria (mean 0.65, SD 0.18), and plants (mean 0.61, SD 0.14), followed by the class humans (mean 0.58, SD 0.23), which showed a dip in the sentiment in the months following November 2018. The class unspecified had a slightly lower sentiment (mean 0.45, SD 0.13) compared with the other classes. In addition to this monthly breakdown, the differences between the organism class means based on the individual tweets were all significant (P<.001), except for the difference between the class means of bacteria and plants with a 3.8% probability of occurring by chance (P=.038; see Multimedia Appendix 5 for all test statistics).

Figure 3. A) Heatmap of monthly sentiments by predicted organism. The sentiments were calculated as the mean of the weighted counts by sentiment (the weights included −1, 0, and 1 for negative, neutral, and positive tweets, respectively) for each month and organism class. Blue and red colors indicate positive and negative sentiment values, respectively. The sentiments of heatmap cells with <100 tweets of that month and organism are transparent. B) Monthly counts by predicted organism.

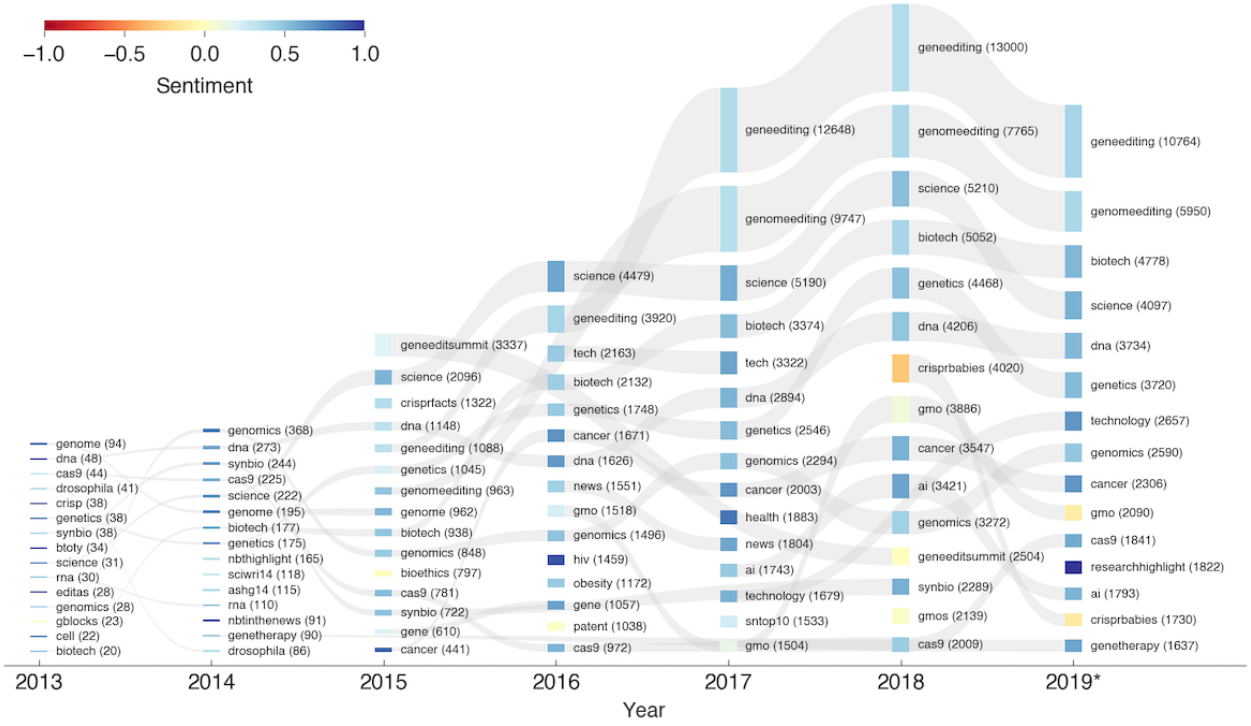


Hashtags

The most frequently used hashtags of every year revealed the topics of highest interest and how they evolved over time (see Figure 4). Naturally, the occurrences of individual hashtags increased over the years along with the total number of tweets. Certain very common hashtags, such as #dna, #science, #biotech, or #geneediting and #genomeediting, appeared as top hashtags in multiple years. When relating the hashtags with the sentiment of the text they appeared in, we can see that most of these common hashtags were used in the context of a positive or very positive sentiment. The 3 hashtags with the most positive sentiments and that were used at least 100 times were #cancer (mean sentiment 0.85, SD 0.36) in 2015, #hiv (mean 0.90, SD 0.34) in 2016, and #researchhighlight (mean 1.00, SD 0.06) in 2019. It is also notable that #science was among the 5 most common hashtags in every year except for 2013 and was consistently related to a positive sentiment, with means between 0.52 (in 2018) and 0.74 (in 2013).

Only a few hashtags were related to negative sentiments. The most prominent one was #crisprbabies, with mean sentiments of −0.30 (SD 0.65) in 2018 and −0.13 (SD 0.63) in 2019, followed by #gmo (mean −0.11, SD 0.76) in 2019, #bioethics (mean −0.02, SD 0.45) in 2015, and #geneeditsummit (mean −0.01, SD 0.46) in 2018. It is worth noting that the hashtag #geneeditsummit only appeared in 2015 and 2018 and that its associated sentiment dropped from 0.20 to −0.01. The hashtag refers to the two summits on human genome editing, which were held in Washington D.C. in 2015 and in Hong Kong in November 2018, coinciding with the first gene editing of viable human embryos. Similarly, the hashtag #gmo became slightly more negative in 2018, with a mean sentiment of 0.09 compared to 2016 (mean 0.24) and 2017 (mean 0.14) and even dropped to −0.11 in 2019. The hashtag #bioethics only appeared in 2015 and was associated with a relatively low sentiment of −0.02. This may highlight the various ethical concerns raised during the 2015 Human Gene Editing summit. See Multimedia Appendix 8 for the full list of the counts, sentiments, and standard deviations of the most used hashtags by year.

Figure 4. Visualization of the sentiment associated with the most frequently used hashtags every year. For every year, the 15 hashtags with the highest counts for that year are included (the hashtag #crispr was excluded). The hashtags are sorted by yearly counts (indicated by the bar height), where the hashtag with the highest count is at the top. The color represents the average sentiment for the respective hashtag, with blue representing a very positive sentiment and red representing a very negative sentiment. If a hashtag is listed in multiple years, the occurrences are linked with a gray band. The number of tweets with the hashtag is indicated in parentheses next to the respective hashtag. For the year 2019, the counts were extrapolated from the months before June to the full year.

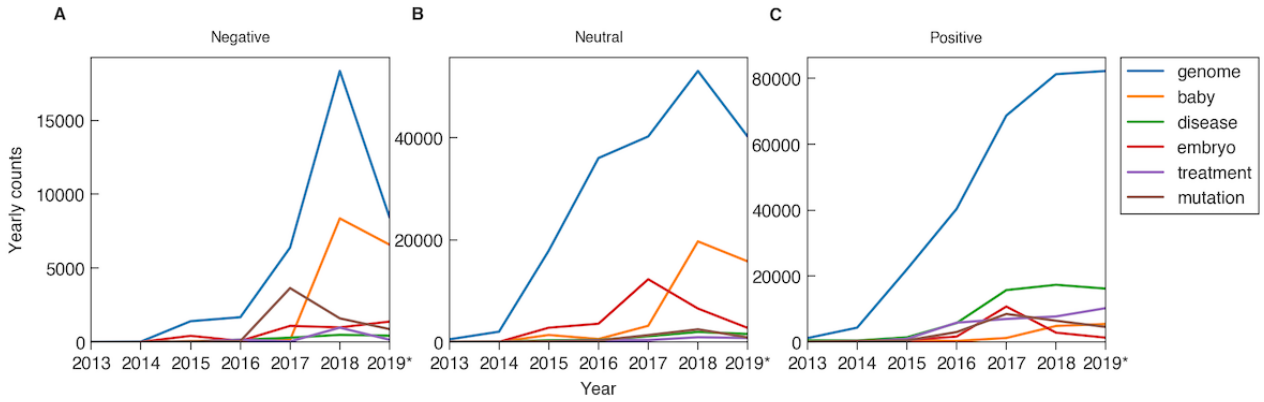


Themes

In comparison to the hashtags, the themes derived from previous studies can relate the Twitter discussion to known themes of interest to the public (see the Methods section for a description of the analysis). The 6 themes that were matched most are presented in Figure 5 and grouped by positive, neutral, and negative sentiments. The themes include genome (with a total count of 526,612 [extrapolated for 2019]), baby (68,269), disease (64,181), embryo (49,084), treatment (35,865), and mutation (34,884). Unsurprisingly, the theme “genome” was matched most frequently, occurring in 34% of the tweets.

The reported themes show distinct occurrence patterns depending on sentiment, yielding an aggregated picture of the discussion surrounding CRISPR throughout the years. Spikes are evident in certain years (see Multimedia Appendix 9 for the counts per year of the top 6 themes), and the most significant change in occurrences happened for the theme “baby,” which increased substantially from 2017 to 2018, likely associated with the “CRISPR babies” scandal in November 2018. While a spike could be observed for all 3 sentiments, the increase was far more pronounced in the neutral and negative classes (see Figure 5). The theme “mutation” shows a negative peak in 2017, when risks about potential side effects of CRISPR surfaced. Relative to other themes, the themes “disease” and “treatment” were major themes in a discussion associated with a positive sentiment.

Figure 5. Yearly occurrences of themes. Multiple themes with distinct regex patterns were matched to the text of tweets, and the 6 most frequent themes were selected. Panels A, B, and C show the yearly counts of themes when grouped by negative, neutral, and positive sentiment, respectively. For the year 2019, the counts were extrapolated from the months before June to a full year.



Discussion

Principal Findings

We have generated the first high-resolution temporal signal for sentiments towards CRISPR on Twitter, spanning a duration of more than 6 years. Our results suggest that, overall, the CRISPR technology was discussed in a positive light, which aligns well with a previous study that considered the coverage of CRISPR in the press [33]. However, more recently, the sentiment reveals a series of strong negative dips, pointing to a more critical view. The frequency and magnitude of these dips have increased since 2017, which is underlined by the overall declining sentiment. It is noteworthy that the dips usually coincide with high activity, suggesting that many people are only exposed to the topic of CRISPR when it is presented in an unfavorable way.

Further, we could tie the most prominent peaks in tweeting activity to real world events. The last 3 peaks, which coincide with the release of possibly concerning news (side effects, CRISPR babies, malware), also align with strong dips in the tweet sentiment. Together, this indicates that there is at least a partial connection between tweets and the discourse off Twitter and that the sentiment changes are not only the result of a self-contained discussion on the social media platform. Even more so, the peak detection potentially allows the timely identification of significant incidents that can shape public discourse and opinion.

As shown in the breakdown of sentiment by organism, the negative sentiment was stronger in the embryo and human classes but stayed mostly positive towards other organisms. The data therefore suggest that the many ethical issues related to human germline editing are reflected in the tweets. However, criticism may not be targeted at the use of CRISPR in humans per se: Hashtags such as #hiv or #genetherapy were connected to very positive sentiments, which suggests a positive attitude towards developing CRISPR for use in medical treatment. This aspect is further strengthened when considering the sentiment of themes such as “treatment” or “disease.” These observations are in line with several surveys in which participants demonstrated strong support of CRISPR for use in medical treatment but were critical regarding modifications of human germline cells [28-32].

The dataset that includes continuous observations over a long period of time allows for conclusions to be drawn about the public perception of CRISPR both on short and long time scales. For example, when the article on biohacking re-emerged in 2019 (peak f), shortly after the discussions around CRISPR babies, it was discussed in significantly more negative terms than at the time of its publication in 2017. Therefore, the intermediate developments seem to have had a negative influence on the perception of the event. This is in line with the overall negative trend. The presence and absence of themes observed in the data hint at the influence that key events might have on the discussion. While the theme “mutation” was discussed intensely in 2017, its occurrence in tweets dropped in the following year, 2018, in which “baby” became the most occurring theme except for “genome”.

Our results support the use of Twitter and similar platforms for the study of public discourse. Discussion about a subject matter can be investigated in real-time, in depth at the level of individual statements, and on the basis of existing data. The insights gained through such studies can bring new issues to light, indicate which topics need extra attention with respect to ethical considerations and policy making, and allow a quicker response to technological advancements. In addition, the presented method offers a novel approach to promote public engagement, especially in the areas of biotechnologies and health care, as argued by the Nuffield Council on Bioethics [60].

Limitations

Although the predicted sentiment index seems to overlap well with survey results, it cannot be directly used as a substitute for an opinion poll. Polling allows for the collection of answers to specific questions of interest instead of inferring them from public statements. Furthermore, the Twitter community is not necessarily representative of the whole population of a country. However, sentiment analysis avoids the disadvantages of traditional methods such as response bias and provides more detailed insights through access to granular data of online discussions.

We cannot exclude the possibility that the gradual decrease over time was influenced or caused by a general shift in the sentiment of the scientific Twitter community. Our analysis relies only

on Twitter, and we did not validate the findings on another social media platform. Also, we cannot directly tie the sentiment in tweets to the conversation off Twitter. Nonetheless, our results show that there is a connection between tweets, findings in earlier studies, and real-world events and that insights can be gained from this type of analysis on Twitter that are not accessible through other methods.

Further, we acknowledge that most people's opinions might not fit into the positive, neutral, and negative classes presented in this study. We therefore tried to counteract this problem by categorizing the data not only by sentiment but also by relevance and organism, allowing for a better understanding of the measured sentiment. Furthermore, we recognize the challenging nature of deducing someone's true opinion based on a short message alone and the fact that it is only possible within a statistical margin of error. This error is slightly larger for the negative class, as the F1 score of this class was relatively low compared to the other classes due to a strong label imbalance. We believe, however, that our method is nevertheless suitable to capture certain trends on a larger scale.

Conclusions and Future Direction

We demonstrated that the sentiment analysis of tweets provides a high-resolution picture of the ongoing debate on CRISPR, allowing us to study the evolution of the discourse while extending the capacity of traditional methods. Further, the presence of the same themes that have been identified in existing studies confirms the validity of our signal with respect to content. The existence of events that match the activity peaks also indicates the sensitivity of the signal towards off-Twitter incidents. Therefore, our approach offers an additional method to surveys and that can be deployed to get richer information, a larger sample size, and higher temporal resolution.

Future work can go beyond the deduction of sentiments and shed more light on the nature of discussions and arguments raised and how they influence each other, giving a better idea of the reasoning behind people's opinions. Furthermore, specific topics, such as the discussion surrounding a potential moratorium of CRISPR, may be analyzed in more detail and provide actionable outcomes.

Since the presented analysis can automatically process a large amount of data in almost real-time, it extends the traditional toolset of empirical methods for discourse analysis. It may therefore help analyze public opinion and support policy and decision making.

Data and Code Availability

The data, machine learning models used, and source code for this analysis can be found in our public repository: https://gitlab.ethz.ch/digitalbioethics/crispr-sentiment-analysis

Acknowledgments

MM and MSch designed the experiment, performed the analysis, and wrote the paper in equal parts. MS and EV initiated the work, guided the experimental design, and made corrections to the paper. MS wrote the abstract. We thank Agata Ferretti for the support during the initial literature review and Ellen Lapper for proofreading.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Yearly counts. Number of tweets per year since January 1, 2013, until May 31, 2019. A steady increase in volume can be observed. In parentheses is the extrapolated number for 2019 (from the first five months).
[PDF File (Adobe PDF File), 28 KB - jmir_v22i8e17830_app1.pdf]

Multimedia Appendix 2
Model performance. Classification scores for selected models. Subfigures A, B and C correspond to three different classifiers trained for sentiment, relevance and organism, respectively. The y-axis shows the best corresponding model for a specific model type after hyperparameter search was performed. The model types are random (pick a class at random), majority (always pick the most frequent class), bag of words, fastText, BERT and a fine-tuned version of BERT-large (denoted as BERT ft). The x-axis denotes the test performance scores of accuracy (green), and macro-averaged precision (blue), recall (orange) and F1 scores (red). The fine-tuned BERT model was the best performing model for all three classification problems irrespective of the metric used.
[PDF File (Adobe PDF File), 112 KB - jmir_v22i8e17830_app2.pdf]

Multimedia Appendix 3
Preliminary literature review search strategy and databases.
[PDF File (Adobe PDF File), 50 KB - jmir_v22i8e17830_app3.pdf]

Multimedia Appendix 4
Themes and regex patterns. Derived themes and corresponding regex patterns from preliminary literature review.
[PDF File (Adobe PDF File), 37 KB - jmir_v22i8e17830_app4.pdf]

Multimedia Appendix 5
Mean sentiments and test statistics for years and organism classes. On the left, the table shows the mean sentiment (Sent) for each year and organism class based on the sentiments of the individual tweets. Further, the standard deviation (SD) and number (Count) of tweets for each group are reported. On the right, the p-values or the significance level if significant (alpha = 0.001) and the t-values from Welch's t-test among the years and organism class means are shown. A value refers to the comparison between the classes given by its row and column labels. For example, the p-value for Welch's t-test for the difference between the mean of the bacteria class and the plants class is 0.038.
[PDF File (Adobe PDF File), 75 KB - jmir_v22i8e17830_app5.pdf]

Multimedia Appendix 6
Identified events. Selected events with a peak prominence above 0.2. The marks correspond to the selected events in Figure 2 of the article. Peak times have been automatically detected as described in the methods section. The corresponding events have been inferred from visual inspection of the data.
[PDF File (Adobe PDF File), 30 KB - jmir_v22i8e17830_app6.pdf]

Multimedia Appendix 7
Monthly mean sentiments and standard deviations per organism. The table shows the mean sentiments (Sent) and their standard deviations (SD) for every month and organism. A dash (–) indicates that less than 100 tweets were in the respective organism class for that month and that we did not calculate the mean sentiment. Months with empty rows had no tweets in that class. The mean values of this table were used in Figure 3.
[PDF File (Adobe PDF File), 61 KB - jmir_v22i8e17830_app7.pdf]

Multimedia Appendix 8
Top hashtags' counts and sentiments. List of top 15 hashtags, corresponding counts (Count), sentiments (Sent) and standard deviations (SD) by year. The extrapolated hashtag counts for 2019 are shown under 2019*, the original counts for the first five months under 2019. The mean values of this table were used in Figure 4.
[PDF File (Adobe PDF File), 49 KB - jmir_v22i8e17830_app8.pdf]

Multimedia Appendix 9
Top themes found in tweets. List of top 6 themes with highest overall occurrence across sentiment. The table shows the number of occurrences in tweets for every sentiment and year. The year 2019 was extrapolated to determine the top themes, indicated by the star (*), based on the first five months of 2019. These counts were used in Figure 5.
[PDF File (Adobe PDF File), 32 KB - jmir_v22i8e17830_app9.pdf]

References

1. Rosenberg SA, Aebersold P, Cornetta K, Kasid A, Morgan RA, Moen R, et al. Gene Transfer into Humans — Immunotherapy of Patients with Advanced Melanoma, Using Tumor-Infiltrating Lymphocytes Modified by Retroviral Gene Transduction. N Engl J Med 1990 Aug 30;323(9):570-578. [doi: 10.1056/nejm199008303230904]
2. Fraley RT, Rogers SG, Horsch RB, Sanders PR, Flick JS, Adams SP, et al. Expression of bacterial genes in plant cells. Proc Natl Acad Sci U S A 1983 Aug 01;80(15):4803-4807 [FREE Full text] [doi: 10.1073/pnas.80.15.4803] [Medline: 6308651]
3. Johnson I. Human insulin from recombinant DNA technology. Science 1983 Feb 11;219(4585):632-637. [doi: 10.1126/science.6337396] [Medline: 6337396]
4. Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, et al. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 2000 Jan 14;287(5451):303-305 [FREE Full text] [doi: 10.1126/science.287.5451.303] [Medline: 10634784]
5. Thomas KR, Folger KR, Capecchi MR. High frequency targeting of genes to specific sites in the mammalian genome. Cell 1986 Feb 14;44(3):419-428. [doi: 10.1016/0092-8674(86)90463-0] [Medline: 3002636]
6. Choulika A, Perrin A, Dujon B, Nicolas JF. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. Mol. Cell. Biol 1995 Apr 01;15(4):1968-1973. [doi: 10.1128/mcb.15.4.1968]
7. Jasin M. Genetic manipulation of genomes with rare-cutting endonucleases. Trends in Genetics 1996 Jun;12(6):224-228. [doi: 10.1016/0168-9525(96)10019-6]
8. Porteus MH, Baltimore D. Chimeric nucleases stimulate gene targeting in human cells. Science 2003 May 02;300(5620):763 [FREE Full text] [doi: 10.1126/science.1078395] [Medline: 12730593]
9. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science 2007 Mar 23;315(5819):1709-1712. [doi: 10.1126/science.1138140]
10. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 2012 Jun 28;337(6096):816-821. [doi: 10.1126/science.1225829]

11. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 2013 Jan 03;339(6121):819-823. [doi: 10.1126/science.1231143]
12. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-Guided Human Genome Engineering via Cas9. Science 2013 Jan 03;339(6121):823-826. [doi: 10.1126/science.1232033]
13. Liang P, Xu Y, Zhang X, Ding C, Huang R, Zhang Z, et al. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. Protein Cell 2015 Apr 18;6(5):363-372. [doi: 10.1007/s13238-015-0153-5]
14. Caplan AL, Parent B, Shen M, Plunkett C. No time to waste--the ethical challenges created by CRISPR: CRISPR/Cas, being an efficient, simple, and cheap technology to edit the genome of any organism, raises many ethical and regulatory issues beyond the use to manipulate human germ line cells. EMBO Rep 2015 Nov;16(11):1421-1426 [FREE Full text] [doi: 10.15252/embr.201541337] [Medline: 26450575]
15. Bosley KS, Botchan M, Bredenoord AL, Carroll D, Charo RA, Charpentier E, et al. CRISPR germline engineering—the community speaks. Nat Biotechnol 2015 May 12;33(5):478-486. [doi: 10.1038/nbt.3227]
16. Regalado A. Chinese scientists are creating CRISPR babies.: MIT Technology Review; 2018 Nov 25. URL: https://www.technologyreview.com/s/612458/exclusive-chinese-scientists-are-creating-crispr-babies/ [accessed 2020-07-29]
17. Cyranoski D, Ledford H. Genome-edited baby claim provokes international outcry. Nature 2018 Nov 26;563(7733):607-608. [doi: 10.1038/d41586-018-07545-0] [Medline: 30482929]
18. Normile D. Shock greets claim of CRISPR-edited babies. Science 2018 Nov 30;362(6418):978-979. [doi: 10.1126/science.362.6418.978] [Medline: 30498103]
19. Baltimore D, Berg P, Botchan M, Carroll D, Charo RA, Church G, et al. Biotechnology. A prudent path forward for genomic engineering and germline gene modification. Science 2015 Apr 03;348(6230):36-38 [FREE Full text] [doi: 10.1126/science.aab1028] [Medline: 25791083]
20. Lander ES. Brave New Genome. N Engl J Med 2015 Jul 02;373(1):5-8. [doi: 10.1056/nejmp1506446]
21. Lander ES, Baylis F, Zhang F, Charpentier E, Berg P, Bourgain C, et al. Adopt a moratorium on heritable genome editing. Nature 2019 Mar 13;567(7747):165-168. [doi: 10.1038/d41586-019-00726-5]
22. Legge Jr JS, Durant RF. Public opinion, risk assessment, and biotechnology: Lessons from attitudes toward genetically modified foods in the european union. Review of Policy Research 2010;27(1):59-76. [doi: 10.1111/j.1541-1338.2009.00427.x]
23. Ferretti MP. Why Public Participation in Risk Regulation? The Case of Authorizing GMO Products in the European Union. Science as Culture 2007 Dec;16(4):377-395. [doi: 10.1080/09505430701706723]
24. Marris C. Public views on GMOs: deconstructing the myths. Stakeholders in the GMO debate often describe public opinion as irrational. But do they really understand the public? EMBO Rep 2001 Jul;2(7):545-548 [FREE Full text] [doi: 10.1093/embo-reports/kve142] [Medline: 11463731]
25. Geller G, Bernhardt BA, Holtzman NA. The Media and Public Reaction to Genetic Research. JAMA 2002 Feb 13;287(6):773. [doi: 10.1001/jama.287.6.773-jms0213-3-1]
26. Travis J. Inside the summit on human gene editing: A reporter's notebook.: Science; 2015 Dec 04. URL: https://www.sciencemag.org/news/2015/12/inside-summit-human-gene-editing-reporter-s-notebook [accessed 2020-07-29]
27. National Academies of Sciences, Engineering, and Medicine. Human Genome Editing: Science, Ethics, and Governance. Washington, DC: The National Academies Press; 2017.
28. Weisberg SM, Badgio D, Chatterjee A. A CRISPR New World: Attitudes in the Public toward Innovations in Human Genetic Modification. Front Public Health 2017 May 22;5:117 [FREE Full text] [doi: 10.3389/fpubh.2017.00117] [Medline: 28589120]
29. McCaughey T, Sanfilippo P, Gooden G, Budden D, Fan L, Fenwick E, et al. A Global Social Media Survey of Attitudes to Human Genome Editing. Cell Stem Cell 2016 May;18(5):569-572. [doi: 10.1016/j.stem.2016.04.011]
30. Gaskell G, Bard I, Allansdottir A, da Cunha RV, Eduard P, Hampel J, et al. Public views on gene editing and its uses. Nat Biotechnol 2017 Nov 9;35(11):1021-1023. [doi: 10.1038/nbt.3958]
31. Hendriks S, Giesbertz NAA, Bredenoord AL, Repping S. Reasons for being in favour of or against genome modification: a survey of the Dutch general public. Hum Reprod Open 2018;2018(3):hoy008 [FREE Full text] [doi: 10.1093/hropen/hoy008] [Medline: 30895249]
32. McCaughey T, Budden DM, Sanfilippo PG, Gooden GE, Fan L, Fenwick E, et al. A Need for Better Understanding Is the Major Determinant for Public Perceptions of Human Gene Editing. Human Gene Therapy 2019 Jan;30(1):36-43. [doi: 10.1089/hum.2018.033]
33. Marcon A, Master Z, Ravitsky V, Caulfield T. CRISPR in the North American popular press. Genet Med 2019 Oct;21(10):2184-2189 [FREE Full text] [doi: 10.1038/s41436-019-0482-5] [Medline: 30976097]
34. Liu B. Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies 2012 May 23;5(1):1-167. [doi: 10.2200/s00416ed1v01y201204hlt016]
35. Ravi K, Ravi V. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems 2015 Nov;89:14-46. [doi: 10.1016/j.knosys.2015.06.015]
36. Kouloumpis E, Wilson T, Moore JD. Twitter Sentiment Analysis: The Good the Bad and the OMG!. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; Jul 17-21, 2011; Barcelona, Catalonia, Spain p. 538-541.

37. Severyn A, Moschitti A. Twitter Sentiment Analysis with Deep Convolutional Neural Networks. In: SIGIR '15: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY, United States: Association for Computing Machinery; 2015 Presented at: 38th International ACM SIGIR Conference on Research and Development in Information Retrieval; Aug 9-13, 2015; Santiago, Chile p. 959-962. [doi: 10.1145/2766462.2767830]
38. Müller MM, Salathé M. Crowdbreaks: Tracking Health Trends Using Public Social Media Data and Crowdsourcing. Front Public Health 2019 Jul 27;7(5):81-551 [FREE Full text] [doi: 10.3389/fpubh.2019.00081] [Medline: 31037238]
39. Devlin J, Chang MW, Lee K, Touanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers).: Association for Computational Linguistics; 2019 Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2-7, 2019; Minneapolis, MN p. 4171-4186. [doi: 10.18653/v1/N19-1423]
40. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Jordan MI, LeCun Y, Solla SA, editors. Advances in neural information processing systems: Proceedings of the First 12 Conferences. Cambridge, MA: The MIT Press; 2017:5998-6008.
41. Sun C, Huang L, Qiu X. Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Stroudsburg, PA, USA: Association for Computational Linguistics; 2019 Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2-7, 2019; Minneapolis, MN p. 380-385. [doi: 10.18653/v1/N19-1035]
42. Basile V, Bosco C, Fersini E, Nozza D, Patti V, Pardo FMR, et al. SemEval-2019 Task 5: Multilingual Detection of Hate Speech Against Immigrants and Women in Twitter. In: Proceedings of the 13th International Workshop on Semantic Evaluation. Stroudsburg, PA, USA: Association for Computational Linguistics; 2019:54-63.
43. Zampieri M, Malmasi S, Nakov P, Rosenthal S, Farra N, Kumar R. SemEval-2019 Task 6: Identifying and Categorizing Offensive Language in Social Media (OffensEval). In: Proceedings of the 13th International Workshop on Semantic Evaluation. Stroudsburg, PA, USA: Association for Computational Linguistics; 2019:75-86.
44. Bird S, Klein E, Loper E. Natural Language Processing with Python. Sebastopol, CA: O'Reilly Media Inc; Jun 2009.
45. McKinney W. Data Structures for Statistical Computing in Python. In: Proceedings of the 9th Python in Science Conference. 2010 Presented at: 9th Python in Science Conference; June 28 - July 3, 2010; Austin, TX p. 56-61. [doi: 10.25080/majora-92bf1922-00a]
46. Crowdbreaks. URL: https://www.crowdbreaks.org [accessed 2020-07-29]
47. Amazon Mechanical Turk. URL: https://www.mturk.com [accessed 2020-07-29]
48. Fleiss JL, Cohen J. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. Educational and Psychological Measurement 1973 Oct 01;33(3):613-619. [doi: 10.1177/001316447303300309]
49. Pagliardini M, Gupta P, Jaggi M. Unsupervised Learning of Sentence Embeddings Using Compositional n-Gram Features. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). Stroudsburg, PA, USA: Association for Computational Linguistics; 2018 Presented at: 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; Jun 1-6, 2018; New Orleans, LA p. 528-540. [doi: 10.18653/v1/n18-1049]
50. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995 Sep;20(3):273-297. [doi: 10.1007/BF00994018]
51. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of tricks for efficient text classification. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. Stroudsburg, PA, USA: Association for Computational Linguistics; 2017 Presented at: 15th Conference of the European Chapter of the Association for Computational Linguistics; April 3-7, 2017; Valencia, Spain p. 427-431. [doi: 10.18653/v1/e17-2068]
52. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. HuggingFace's Transformers: State-of-the-art Natural Language Processing. arXiv. 2019. URL: https://arxiv.org/abs/1910.03771v3 [accessed 2020-07-29]
53. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Advances in Neural Information Processing Systems 32.: Curran Associates, Inc; 2019 Presented at: Neural Information Processing Systems 2019; Dec 8-14, 2019; Vancouver, Canada p. 8026-8037.
54. Ruxton GD. The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U Test. Behavioral Ecology 2006;17(4):688-690. [doi: 10.1093/beheco/ark016]
55. Alhabash S, VanDam C, Tan P, Smith SW, Viken G, Kanver D, et al. 140 Characters of Intoxication: Exploring the Prevalence of Alcohol-Related Tweets and Predicting Their Virality. SAGE Open 2018 Oct 04;8(4):215824401880313. [doi: 10.1177/2158244018803137]
56. Jones E, Oliphant T, Peterson P, et al. SciPy: Open Source Scientific Tools for Python. 2001. URL: https://www.scipy.org/ [accessed 2020-07-29]
57. Ma H, Marti-Gutierrez N, Park S, Wu J, Lee Y, Suzuki K, et al. Correction of a pathogenic gene mutation in human embryos. Nature 2017 Aug 2;548(7668):413-419. [doi: 10.1038/nature23305]

58. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol 2018 Jul 16;36(8):765-771. [doi: 10.1038/nbt.4192]
59. Greenberg A. Biohackers Encoded Malware in a Strand of DNA. Wired. 2017. URL: https://www.wired.com/story/malware-dna-hack [accessed 2020-07-29]
60. Nuffield Council on Bioethics. Emerging biotechnologies: technology, choice and the public good. London, UK: Nufield Council on Bioethics; 2012.

Abbreviations

API: application programming interface
BERT: Bidirectional Encoder Representations from Transformers
BoW: Bag of Words
SVM: support vector machine

Edited by G Eysenbach; submitted 20.01.20; peer-reviewed by J Coquet, V Osadchiy; comments to author 22.03.20; revised version received 08.05.20; accepted 03.06.20; published 31.08.20.

Please cite as:
Müller M, Schneider M, Salathé M, Vayena E
Assessing Public Opinion on CRISPR-Cas9: Combining Crowdsourcing and Deep Learning
J Med Internet Res 2020;22(8):e17830
URL: http://www.jmir.org/2020/8/e17830/
doi:10.2196/17830
PMID:32865499

Original Paper

Landscape of Participant-Centric Initiatives for Medical Research in the United States, the United Kingdom, and Japan: Scoping Review

Nao Hamakawa[1], MPH; Rumiko Nakano[2], MA; Atsushi Kogetsu[1], MD; Victoria Coathup[3], PhD; Jane Kaye[3,4], PhD; Beverley Anne Yamamoto[2], PhD; Kazuto Kato[1], PhD

[1]Department of Biomedical Ethics and Public Policy, Graduate School of Medicine, Osaka University, Suita, Osaka, Japan

[2]Graduate School of Human Sciences, Osaka University, Suita, Osaka, Japan

[3]Centre for Health, Law and Emerging Technologies, Faculty of Law, University of Oxford, Oxford, United Kingdom

[4]Centre for Health, Law and Emerging Technologies, Melbourne Law School, University of Melbourne, Melbourne, Australia

Corresponding Author:
Kazuto Kato, PhD
Department of Biomedical Ethics and Public Policy
Graduate School of Medicine
Osaka University
2-2 Yamadaoka
Suita, Osaka, 5650871
Japan
Phone: 81 6 6879 3687
Email: kato@eth.med.osaka-u.ac.jp

Abstract

Background: Information and communication technology (ICT) has made remarkable progress in recent years and is being increasingly applied to medical research. This technology has the potential to facilitate the active involvement of research participants. Digital platforms that enable participants to be involved in the research process are called participant-centric initiatives (PCIs). Several PCIs have been reported in the literature, but no scoping reviews have been carried out. Moreover, detailed methods and features to aid in developing a clear definition of PCIs have not been sufficiently elucidated to date.

Objective: The objective of this scoping review is to describe the recent trends in, and features of, PCIs across the United States, the United Kingdom, and Japan.

Methods: We applied a methodology suggested by Levac et al to conduct this scoping review. We searched electronic databases—MEDLINE (Medical Literature Analysis and Retrieval System Online), Embase (Excerpta Medica Database), CINAHL (Cumulative Index of Nursing and Allied Health Literature), PsycINFO, and Ichushi-Web—and sources of grey literature, as well as internet search engines—Google and Bing. We hand-searched through key journals and reference lists of the relevant articles. Medical research using ICT was eligible for inclusion if there was a description of the active involvement of the participants.

Results: Ultimately, 21 PCIs were identified that have implemented practical methods and modes of various communication activities, such as patient forums and use of social media, in the field of medical research. Various methods of decision making that enable participants to become involved in setting the agenda were also evident.

Conclusions: This scoping review is the first study to analyze the detailed features of PCIs and how they are being implemented. By clarifying the modes and methods of various forms of communication and decision making with patients, this review contributes to a better understanding of patient-centric involvement, which can be facilitated by PCIs.

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.7407

(J Med Internet Res 2020;22(8):e16441) doi:10.2196/16441

KEYWORDS

participant-centric initiatives; patient involvement; patient engagement; participatory research; participatory medicine; information and communication technology; patient participation

Introduction

The use of information and communication technology (ICT) is increasing in all aspects of health care delivery and medical research, enabling vast amounts of data to be accumulated and analyzed at an unprecedented rate. Furthermore, a growing number of people are participating in research using smart devices, such as smartphones, tablets, and wearable devices [1]. Participation in medical research using ICT is expected to increase in the future. At the same time, patient-reported outcomes, where the patient reports directly on his or her condition, such as pain, fatigue, and quality of life, are recognized as important for both clinical and research settings [2,3]. This new trend in medicine also facilitates the use of ICT in the medical research fields.

In parallel, attitudes toward medical research are also shifting to more active involvement of research participants. This emerging model of research, in which researchers and participants collaborate through the research process, is gaining momentum internationally. Several research funders, such as INVOLVE in the United Kingdom, which was established by the National Institute for Health Research (NIHR) to support public involvement in the National Health Service (NHS), and the Patient-Centered Outcomes Research Institute (PCORI) in the United States, are known as leading organizations promoting the involvement of research participants [4].

The definition of patient and public involvement and engagement has not yet been established. There is no designated terminology to describe the active involvement of participants in research. According to INVOLVE, *involvement* is defined as "research that is being carried out 'with' or 'by' members of the public, rather than 'to,' 'about,' or 'for' them," while *engagement* indicates "information and knowledge about research is provided and disseminated" [5]. On the other hand, PCORI in the United States uses the term *engagement* in the following way: "The meaningful involvement of patients, caregivers, clinicians, and other healthcare stakeholders throughout the entire research process" [6]. Another example of the fluidity of the terminology can be seen in a consensus statement on patient and public involvement with data-intensive health research, which was developed by an international group of experts; they use the term *public involvement and engagement* [7]. While the terminology may change, a shared sentiment in all these documents is that they define patients or the public as experts in their personal knowledge and experiences. Collaborating with patients and the public is expected to improve research quality and relevance, ensure transparency and accountability, and foster innovation and research [5-9].

In this review, we use the definition from INVOLVE [5], where *involve* or *involvement* refers to a status where participants play an active role in medical research, beyond the level of simply being a participant to take part in research or inputting their data [10]. For instance, we hoped to find examples where participants collaborate with researchers and have influence over research design, analysis, management, and/or dissemination. While recognizing the diversity of approaches to involvement, here, we are not including activities such as the raising of awareness of research by patients or the creation of events, such as workshops and festivals to engage with the public [11].

Participant-centric initiatives (PCIs) are new initiatives that employ ICT for facilitating active involvement in research and are defined as "digital tools, platforms, or projects that have been developed to help participants become more actively involved in the research process" [12]. PCIs have the potential to provide a number of benefits to both participants and researchers, including facilitating participant recruitment and retention, providing the basis for long-term partnerships, and sustaining public confidence in research [13]. In addition to this, the interactive interface facilitates communication between research participants and researchers throughout the research process and allows participants to be placed at the center of the decision-making process [13]. Furthermore, it is believed that PCIs can address issues toward protecting individual interests by mediating participants' control and choice within diverse research contexts [12].

While examples and features of PCIs have been reported by Anderson et al [12] and Kaye et al [13], in 2012, the definition of PCIs and methods of involvement that are promoted through PCIs had not yet been established. Moreover, the status and characteristics of PCIs since they were first outlined in these papers have not been reported. We believed that it was important to capture the current landscape of PCIs and describe detailed characteristics and methods of participant involvement.

By applying a scoping review methodology, we examined PCIs that have been implemented in the United States, the United Kingdom, and Japan, and we systematically analyze their detailed functions and features. The reason for selecting these countries for the study was that the United States and the United Kingdom have been actively advancing patient-centric approaches in medical research, and the majority of reported PCIs in previous research was located in these two countries [12,13]. Our aim was to build on the existing research by updating the PCI landscape for the United Kingdom and the United States and by further contributing to the literature by adding Japan. Japan is a leading country in health research and practice, and we were well positioned to access the literature. While a traditional model of medical research remains pervasive in Japan [14], there is a clear shift toward exploring or prioritizing patient and public involvement, as seen in recent statements made by the Japan Agency for Medical Research and Development, a major medical funding agency [15]. Given this new focus, it was timely to search and explore the current landscape of PCIs available in Japan. We felt that focusing our attention on these three countries would make an important contribution to the literature on PCIs.

Therefore, the aims of this study were (1) to identify existing PCIs used for medical research in the United States, the United Kingdom, and Japan, (2) to describe recent trends and features of PCIs, and (3) to highlight the methods of participant involvement facilitated by PCIs.

Methods

Overview

The methodology followed here is based on the previously published study protocol [16]. The scoping review was considered the most appropriate method to address the aims of the study for the following reasons. Firstly, the scoping review is recommended, as it is particularly relevant to disciplines with emerging evidence [17]. Unlike a systematic review, we are not trying to answer a specific question, but rather "examine the extent, range, and nature" of PCIs [18]. Secondly, a scoping review provides comprehensive search methods to incorporate a range of study designs in both published and grey literature. The scoping review methodological framework described by Levac et al [17] was applied to this study.

Identifying the Research Question

According to the features of PCIs described by Anderson et al [12], it is not necessary to designate the field of research upfront, but we decided to limit our study to PCIs generating data for medical research. In addition, citizen science, which we define here as a research activity directed and conducted by citizens without collaboration with researchers, was considered beyond our remit [19]; therefore, we focused on interaction between participants and researchers using PCIs.

In this study, *medical research* refers to any research involving human subjects aimed at improving clinical outcomes, including prevention; understanding the etiology of diseases and/or effect of treatments; and improving the quality of life of patients.

Identifying Relevant Studies

We first conducted a literature search in June 2017 in the following databases: MEDLINE (Medical Literature Analysis and Retrieval System Online), Embase (Excerpta Medica Database), CINAHL (Cumulative Index of Nursing and Allied Health Literature), PsycINFO, and Ichushi-Web. We conducted our search by using subject headings and keywords. Search terms included a combination of keywords and subject headings such as Medical Subject Headings (MeSH). Keywords were comprised of the following terms: "participant" AND "centric" OR "centered" OR "centred" OR "engage" OR "involve" OR "collaborat" OR "partner" OR "led" OR "driven" OR "initiat" OR "oriented." Subject headings for "participation," "technology," and "research" were searched and adjusted to best meet the requirements of each database. The detailed search strategies and history are shown in Multimedia Appendix 1. A subject librarian was consulted and provided guidance on the search strategy. We also conducted a cited literature search using the Web of Science and by hand-searching of key journals—*Digital Health* and *The Journal of mHealth*—in August 2017.

A grey literature search was also conducted using Open Grey in December 2017, and a website search was conducted using Google and Bing from April to June 2018. The grey literature and website searches were conducted by using the same search keywords as in the literature database search. Some searches showed a large number of items, for example, more than 100,000 hits. However, as it was practically challenging to identify relevant websites by screening all items, we took a pragmatic approach and screened websites that had appeared within the first 50 results for each keyword.

Study Selection

The relevant articles and websites were screened based on the inclusion criteria described in Textbox 1. These criteria were formulated based on the description of the PCIs [12,13], and some criteria were added by the research team in the process of screening. The articles obtained by the database search were screened by two independent reviewers (NH and RN), and PCIs were identified from this process. The items identified from the grey literature and website searches were screened by NH, and the selection was confirmed by the research team after several rounds of discussion on the screening results.

Textbox 1. Selection criteria for relevant articles and websites.

Inclusion criteria:

- Research enables participants to become actively involved in the research design
- Complies with participant-centric initiative (PCI) features described by Anderson et al [12]: (1) digital device or tool, computer program, or digital platform, and (2) projects that empower participants to engage in the research process
- Articles, documents, or websites published in English or Japanese
- Adult population (ie, over 18 years of age)
- Focuses on medical research purposes
- Available to participants in the United States, the United Kingdom, or Japan

Exclusion criteria:

- Platforms that enable patients to connect and communicate with other patients only
- Platforms that use data for research, but there is no interaction between participants and researchers
- Research activity directed and conducted by citizens without the support of scientists
- Research intended to improve the efficiency of clinical practices or to develop tools for health care services
- Medical research that aims to engage with participants without using a digital platform

Data Extraction

The features of PCIs were extracted and mapped from the relevant articles and information on the websites of each PCI. The data extraction was completed by NH, and the preliminary results were reviewed by research study members to ensure validity. Characteristics to extract included the following: characteristics of PCI websites, type of medical research, and method of involvement. The list of data elements that were extracted are shown in Multimedia Appendix 2.

Collating, Summarizing, and Reporting the Results

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart template [20] was applied to report on the search process. The key characteristics of the included PCIs were summarized in results tables and charts. The key findings were described in a summary report before disseminating to expert panel members.

Consultation With Expert Panel

A consultation is recommended as an optional stage in conducting a scoping review [17]. We carried out a consultation in January 2019 with a small number of experts to receive feedback about the obtained results, including additional information and perspectives. Using our research network, the expert panel included four researchers and a patient representative with expertise in patient and public involvement, health care and digital technologies, and clinical cohort studies from the United States, the United Kingdom, Australia, and Japan.

Results

Identifying Relevant Studies and Study Selection

The search results and screening process are shown in Figure 1. The screening process can be divided into two flowcharts: the database search and the grey literature search.

In the database search, 10 relevant articles [1,12,13,21-27] remained after the title, abstract, and full-text screening, and the descriptions of 13 potential PCIs were found in those papers. Within these 13 potential PCIs, 3 had been implemented in nontargeted countries (ie, Italy, the Netherlands, and Iceland), and 3 websites were not accessible, possibly due to the termination of the projects. Only 1 website focused on releasing personal genetic data openly to the public; we considered this beyond the inclusion criteria of this scoping review because all participants were scientists, and they were not aiming at collaborating with patients or citizens. As a result, 6 PCIs satisfying the inclusion criteria were identified in the database search.

The grey literature and website searches were conducted in an iterative process reflecting the results of the database search that had been performed earlier. In total, 3622 documents and websites were screened, but none described the implementation of PCIs for medical research. However, 16 additional PCIs were identified by a supplementary search of 2 excluded articles [28,29] from the database search. Both articles were related to the PCORI. Another 2 potential PCIs were discovered while searching for information on other PCIs that had been found in the database search. In total, 16 PCIs were identified in the website search. After removing duplicates, 21 PCIs [30-50] were ultimately identified in the obtained results (see Table 1). The detailed characteristics of these 21 PCIs are also shown in Multimedia Appendix 3.

Figure 1. Flowchart for study selection. MEDLINE: Medical Literature Analysis and Retrieval System Online; Embase: Excerpta Medica Database; CINAHL: Cumulative Index of Nursing and Allied Health Literature; PCI: participant-centric initiative.

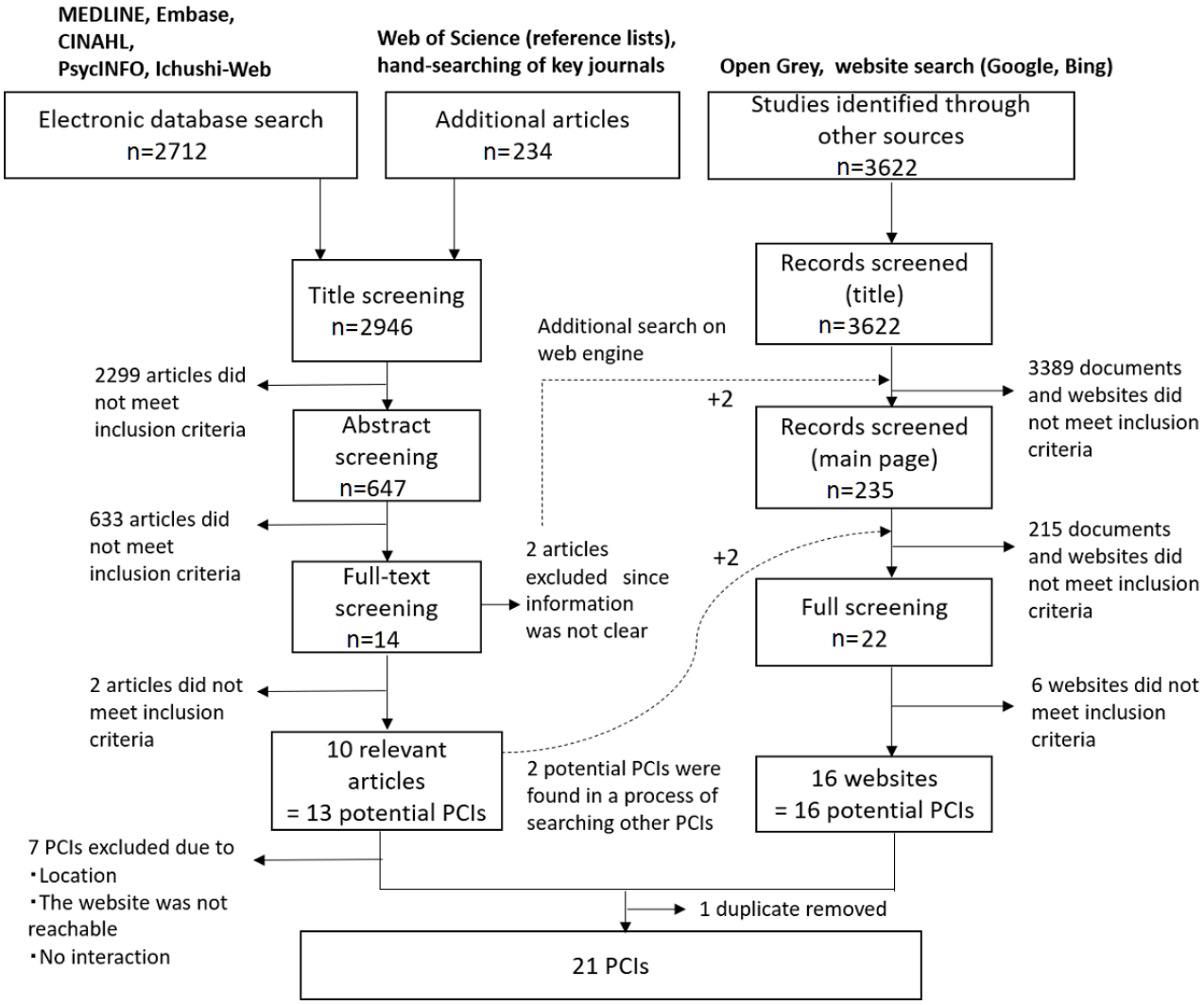


Table 1. Participant-centric initiatives (PCIs) identified by our scoping review.

No.	Name of PCI	Location	Type of organization	Areas of focus	Launch year	Number of users[a]
1	23andMe [30]	United States	Industry	Diverse (more than 230)	2006	>1,200,000
2	PatientsLikeMe [31]	United States	Industry	Diverse (more than 2800)	2006	>600,000
3	PEER[b] [32]	United States	NPO[c]	Diverse (about 50)	2014	>15,000
4	GenomeConnect [33]	United States	Research institute	Diverse (genetic disorders)	2014	1400
5	RUDY[d] [34]	United Kingdom	University	Fibrous dysplasia, vasculitis, osteogenesis imperfect, etc	2014	993
6	MoodNetwork [35]	United States	Research hospital	Mood disorders	2015	Unknown
7	mPower [36]	United States	NPO	Parkinson disease	2015	15,000
8	J-RARE [37]	Japan	NPO	Distal myopathy, relapsing polychondritis, Marfan syndrome, etc	2013	≥47
9	ABOUT Network [38]	United States	NPO and university	Hereditary breast cancer	2016	10,500
10	Arthritis Power [39]	United States	NPO	Rheumatoid arthritis, fibromyalgia, inflammatory bowel disease (IBD), etc	2014	15,365
11	IBD[e] Partners [40]	United States	NPO and university	Crohn disease	2011	15,680
12	Rare Epilepsy Network (REN) [41]	United States	NPO	Rare epilepsy	2014	1392
13	COPD PPRN[f] [42]	United States	NPO	Chronic obstructive pulmonary disease and asthma	2014	75,000
14	Health eHeart [43]	United States	University	Cardiovascular diseases	2013	75,000
15	IAN[g] [44]	United States	NPO	Developmental disorder	2006	>20,000
16	iConquerMS (multiple sclerosis) [45]	United States	NPO	Multiple sclerosis	2015	≥3100
17	AD-PCPRN[h] [46]	United States	Research hospital	Alzheimer disease and dementia	2014	57,000
18	NephCure Kidney Network Patient Registry [47]	United States	NPO	Primary nephrotic syndrome	2014	666
19	PI[i] CONNECT [48]	United States	NPO	Primary immunodeficiency	Unknown	5040
20	V-PPRN[j] [49]	United States	University	Behçet disease, vasculitis, polyarteritis nodosa, etc	Unknown	Unknown
21	MyApnea [50]	United States	Research hospital	Sleep apnea	2013	12,677

[a]The number of registrants (ie, users) is based on information publicly available in July 2018.

[b]PEER: Promise for Engaging Everyone Responsibly.

[c]NPO: nonprofit organization; includes patient organizations and research organizations.

[d]RUDY: Rare and Undiagnosed Diseases Study.

[e]IBD: Inflammatory Bowel Disease.

[f]COPD PPRN: Chronic Obstructive Pulmonary Disease Patient-Powered Research Network.

[g]IAN: Interactive Autism Network.

[h]AD-PCPRN: Alzheimer's Disease Patient- and Caregiver-Powered Network.

[i]PI: Primary Immunodeficiency.

[j]V-PPRN: Vasculitis Patient-Powered Research Network.

Difficulties in Searching for Relevant Studies

During the database search, it became clear that there were no designated subject headings to describe features of the PCIs within this emerging field; for instance, the existing MeSH do not have terms to include "participant-centric," "engagement," or "involvement"; alternatively, "patient participation" was suggested as the only MeSH option. The subject heading is typically used as an effective search tool in a database search. In this case, however, we had to combine keywords to search for articles relevant to PCIs. The lack of effective search terms may reflect the novelty of the field. For the same reason, the results of grey literature and website searches using Google and Bing were also limited.

Many of the documents that were excluded in the screening process were related to patients' decision making or participation in clinical practice or medical interventions, including test screening, rather than medical research. Another major category of excluded items concerned the digitization of research and health care with the aim of developing tools for services, with no description of the involvement of participants.

Results of the Expert Panel

After the consultation, all members of the expert panel commented on the obtained results. Overall, they considered the results "important and well worthy of publication" and the search strategy as "sensible." In addition, the experts gave feedback on the analysis and discussion sections of the paper. The feedback was incorporated into our analysis of results and the discussion.

There were also suggestions on potential PCIs we had not identified through the literature search. A total of 5 potential platforms were suggested by two members of the panel—one from Japan and one from the United Kingdom. Moreover, one member suggested that disease registries were underrepresented in the obtained results. The main reasons that we did not pick up these registries with our literature search is that any scientific papers generated by them appeared after our cutoff point or they did not indicate active involvement of patients in existing papers.

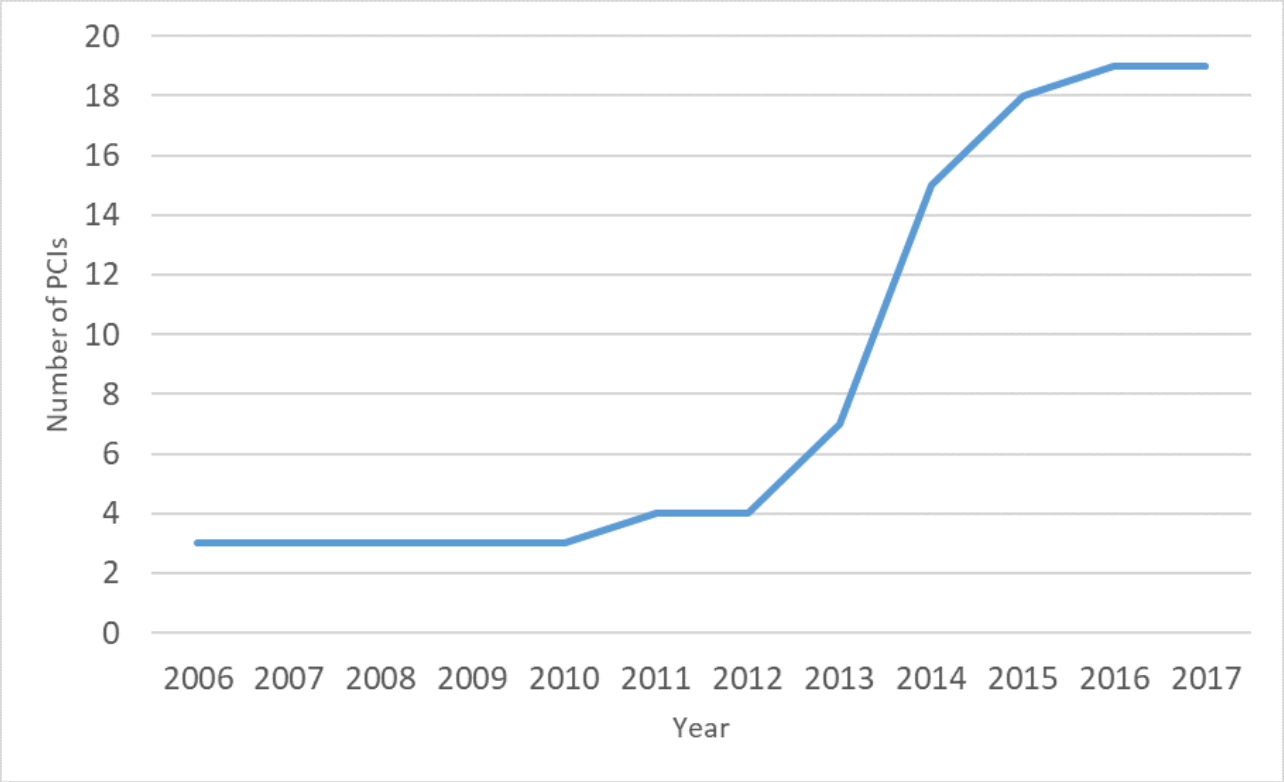
In response, we screened the registries listed on the National Institutes of Health (NIH) website. A total of 62 registries were screened by NH and AK. As a result, we identified 5 PCIs in the United States and Japan, which had not come out in our original literature search. Moreover, an anonymous reviewer suggested an additional PCI during the peer-review process. As we had decided that the results of the literature searches would be our endpoint for this study, these additional PCIs are offered here as supplementary data (see Multimedia Appendix 4). It is noteworthy, however, that the features of these PCIs do not change the overall landscape we describe as the main endpoint of the study.

Recent Trends in PCIs

PCIs have been implemented for research focusing on various diseases, including rare diseases, mood disorders, heart diseases, and dementia. Of the 21 kinds of PCIs identified in this scoping review, 4 (19%)—23andMe, PatientsLikeMe, PEER (Promise for Engaging Everyone Responsibly), and GenomeConnect—had features that included a variety of disease areas and had cross-cutting registry functions ranging from dozens to thousands of diseases. One of them was a direct-to-consumer genetic testing company (ie, 23andMe). In these PCIs, multiple kinds of medical research were being conducted. Furthermore, these 4 PCIs aimed to match the genomic information of research participants with their disease phenotype. On the contrary, the main registration objectives of the other 17 PCIs for medical research concerned specific disease areas; some of them had gradually expanded their disease areas of focus, such as RUDY (Rare and Undiagnosed Diseases Study) and V-PPRN (Vasculitis Patient-Powered Research Network).

Examining the year in which each PCI was established, our results show a dramatic upward trend from 2013 onwards (see Figure 2).

Figure 2. The trend in the number of participant-centric initiatives (PCIs) identified in the scoping review, cumulative total by year. NB: For 2 PCIs, the launch year was unknown.



Medical Research Facilitated by PCIs

The medical research conducted using digital tools, such as personal computers and smartphones, had one or both of the following purposes:

1. Understanding the symptoms: changes in the presentation of symptoms over time, genetic information and disease phenotype matching, daily changes of conditions, and health conditions associated with everyday life, including exercise and diet.
2. Comparison of treatment effectiveness: a case-control study comparing the effects of interventions, such as comparing medication records with symptoms or the impact of an online exercise class.

We found that at least 277 articles had already been published by several PCI research groups by January 2018, including some in major scientific journals [33-36,40,50-52]. Among them, PatientsLikeMe and 23andMe are promoting research covering various disease areas and have published more than 100 papers [51,52]. All PCIs were used to facilitate participation in substudies or provide opportunities to join upcoming research.

Table 2 shows the specific data collected by each PCI. Overall, the majority of PCIs had analyzed questionnaires on symptoms and quality of life. Functions to upload the results of DNA testing and electronic health record data were also implemented in 7 PCIs, corresponding to one-third of the total. Other functions included tracking personal data, such as body temperature, blood pressure, and weight. A total of 3 PCIs had an interface to collect data using new functions on smartphones and wearable devices, such as those relating to movement and voice.

Table 2. Types of data collected by participant-centric initiatives (PCIs).

Types of data	Number of PCIs (N=21), n (%)
DNA test result	7 (33)
Closed-ended questionnaire	17 (81)
Open-ended questionnaire	5 (24)
Treatment and medication	16 (76)
Motion and voice	3 (14)
Self-reported measurement	4 (19)
Electronic health record	5 (24)[a]

[a]In addition to the 5 PCIs that collected electronic health record data, 2 were in preparation.

Model of Informed Consent

Table 3 shows the result of consent models that were indicated on each of the websites. A *specific consent model*—a participant's consent is requested each time they participate in a new study—was used for approximately half the PCIs (12/21, 57%). On the contrary, a *broad consent model* was used in 1 PCI (5%) that was developed by a for-profit organization (ie, PatientsLikeMe).

In total, 3 PCIs out of 21 (14%)—PEER, RUDY, and J-RARE—implemented a *dynamic consent model*. Dynamic consent is a new model that enables participants to change their consent status over time online [53]. Furthermore, participants can also select the range of data sharing and methods of communication. There were 5 PCIs out of 21 (24%) where the consent process was unclear.

Table 3. Model of consent implemented in participant-centric initiatives (PCIs).

Types of data	Number of PCIs (N=21), n (%)
Specific consent	12 (57)
Broad consent	1 (5)
Dynamic consent	3 (14)
Treatment and medication	16 (76)
Unknown	3 (14)

Various Communication Activities

By examining the activities undertaken by the 21 PCIs, it became clear that various modes of communication had been used, including patient forums, webinars, and dialogue (ie, patients and researchers exchanging messages online) (see Table 4). Moreover, a majority of the PCIs had used multiple social media accounts. Some PCIs had disseminated information through blogs and newsletters. Others had their own digital interface that allowed participants to take part in networking with researchers and other participants. Out of 21 PCIs, 2 (10%) had been conducting face-to-face forums or seminars.

Table 4. Modes and methods of communication for the participant-centric initiatives (PCIs).

No.	Name of PCI	Patient forum[a]	Webinar[b]	Dialogue[c]	Use of social media for communication				Other modes of communication
					Facebook	Twitter	YouTube	Other	
1	23andMe		✓		✓	✓	✓	✓	News
2	PatientsLikeMe	✓		✓	✓	✓	✓	✓	News
3	PEER[d]	✓			✓	✓	✓		MOSAIC[e]
4	GenomeConnect			Unclear	✓	✓	✓	✓	Newsletter and mailing list
5	RUDY[f]	✓			✓	✓			N/A[g]
6	MoodNetwork	✓		✓	✓		✓		N/A
7	mPower		✓				✓	✓	Patient satisfaction questionnaire
8	J-RARE								Questionnaire
9	ABOUT Network		✓		✓	✓			GAP360[h]
10	Arthritis Power			✓	✓	✓	✓		N/A
11	IBD[i] Partners		✓	✓	✓	✓	✓	✓	Blog and dashboard for research ideas
12	Rare Epilepsy Network (REN)	✓	✓		✓	✓			Dashboard
13	COPD PPRN[j]			✓					COPD360°
14	Health eHeart			✓	✓	✓		✓	Health eHeart community
15	IAN[k]		✓		✓	✓	✓		N/A
16	iConquerMS (multiple sclerosis)	✓		✓	✓	✓	✓	✓	Newsletter and iConquerMS community
17	AD-PCPRN[l]				✓	✓	✓		
18	NephCure Kidney Network Patient Registry		✓		✓				Patient story and regional volunteer community
19	PI[m] CONNECT	✓	✓		✓	✓	✓		N/A
20	V-PPRN[n]		✓		✓	✓		✓	N/A
21	MyApnea	✓		✓	✓	✓			Online bulletin board, blog, and personalized report

[a]Included community day, leadership summit, and research forum.

[b]Included content for general use.

[c]Included the sharing of experiences, thoughts, and information with researchers and other patients; networking.

[d]PEER: Promise for Engaging Everyone Responsibly.

[e]MOSAIC: Model of Observational Screening for the Analysis of Interaction and Communication.

[f]RUDY: Rare and Undiagnosed Diseases Study.

[g]N/A: not applicable.

[h]GAP: Generate, Assess, Prioritize, Plan, Perform, and Publish.

[i]IBD: Inflammatory Bowel Disease.

[j]COPD PPRN: Chronic Obstructive Pulmonary Disease Patient-Powered Research Network.

[k]IAN: Interactive Autism Network.

[l]AD-PCPRN: Alzheimer's Disease Patient- and Caregiver-Powered Network.

[m]PI: Primary Immunodeficiency.

[n]V-PPRN: Vasculitis Patient-Powered Research Network.

Participant Decision Making in Various Research Processes

It became clear that there were various phases in decision making by research participants. First, the 3 PCIs that implemented the dynamic consent model allowed each participant to control the range of data sharing. Most PCIs used an interface that enabled inputs of participant feedback on research and operations or agenda setting by allowing participants to propose new research questions. Furthermore, more than half the PCIs (14/21, 67%) had a governance structure that included participant representatives in the decision-making process of the research design and conduct of the research (see Table 5).

Table 5. Decision-making process implemented in participant-centric initiatives (PCIs).

No.	Name of PCI	Data-sharing control	Individual feedback and suggesting research questions[a]	Research design and governance[b]
1	23andMe		✓	
2	PatientsLikeMe		✓	✓
3	PEER[c]	✓	✓	✓
4	GenomeConnect		✓	✓
5	RUDY[d]	✓	✓	✓
6	MoodNetwork		✓	✓
7	mPower		✓	
8	J-RARE	✓		✓
9	ABOUT Network		✓	✓
10	Arthritis Power		✓	✓
11	IBD[e] Partners		✓	
12	Rare Epilepsy Network (REN)		✓	
13	COPD PPRN[f]			✓
14	Health eHeart		✓	✓
15	IAN[g]		✓	
16	iConquerMS (multiple sclerosis)		✓	✓
17	AD-PCPRN[h]			✓
18	NephCure Kidney Network Patient Registry		✓	✓
19	PI[i] CONNECT		✓	
20	V-PPRN[j]		✓	
21	MyApnea		✓	✓

[a]Individual comments, suggestions of research questions, decisions of priority, etc.

[b]The main purpose is to determine the overall policy as representative of research participants, such as an advisory board, a steering committee, and a governor group.

[c]PEER: Promise for Engaging Everyone Responsibly.

[d]RUDY: Rare and Undiagnosed Diseases Study.

[e]IBD: Inflammatory Bowel Disease.

[f]COPD PPRN: Chronic Obstructive Pulmonary Disease Patient-Powered Research Network.

[g]IAN: Interactive Autism Network.

[h]AD-PCPRN: Alzheimer's Disease Patient- and Caregiver-Powered Network.

[i]PI: Primary Immunodeficiency.

[j]V-PPRN: Vasculitis Patient-Powered Research Network.

Discussion

Principal Findings

In this study, we conducted a scoping review to capture the recent trend and features of PCIs for medical research, in particular, by focusing on active involvement. A total of 21 PCIs were identified by the scoping review. After analyzing the detailed functions and characteristics of each, the landscape of PCIs became clearer. To our knowledge, this is the first scoping review conducted in this emerging area to map the extent and range of PCIs currently available across the United States, the United Kingdom, and Japan.

Recent Trends in PCIs

The PCIs identified in this study were utilized in medical research in various fields. The number of participants registered for each PCI ranged from approximately 50 to 100,000 (see Table 1). Moreover, it was revealed that there are two types of PCIs in terms of target population and research purpose. A total of 4 PCIs targeted dozens or more diseases, aimed at understanding the relationship between genomic information and clinical phenotype. On the contrary, the other 17 PCIs targeted specific disease areas with the aim of deepening clinical understanding of the symptoms. In addition, PCIs connect participants and researchers with opportunities to join in different medical research studies. Regarding the significant number of participants and the broad targets of research, it is evident that ICT interfaces provide effective ways to promote research without the limitations of conducting research in real time and geographic area.

Another noteworthy trend was the sharp increase in the number of PCIs after 2013. The reasons for the increase may be improvements in technology and greater understanding of the benefits of using digital technologies in health care, as well as an increasing awareness of the importance of engaging with patients. As 12 of the 21 PCIs (57%) were funded by PCORI in the United States, the trend may also have been influenced by the year of its establishment. PCORI was established in 2010 with the aim of funding comparative clinical effectiveness research for patients and those who care for them to make better-informed health decisions [54]. Moreover, PCORI promotes patient engagement to ensure that results are relevant and useful to stakeholders [55,56]. Meanwhile, various factors, such as an increased research focus on patient-reported outcomes [57] and more medical research using smartphones and wearable devices [58], are also thought to be related to this recent trend in PCIs.

There are at least 270 scientific papers published that used data collected by PCIs. Out of 21 PCIs, 2 (10%)—23andMe and PatientsLikeMe—are notable because they have published nearly 100 papers in the area between them; both are organized by for-profit organizations. It is necessary to analyze them further to understand the kinds of research papers that are published by these PCIs and why these for-profit organizations seem to have a greater reach than others.

Features of PCIs

The features of PCIs have been described in a prior paper [13] as "placing participants in control; using social media technology; promoting active participation; facilitating communication; appealing to public good." With the results of this scoping review, these features have become clearer.

Dynamic consent was implemented in 3 out of 21 PCIs (14%)—PEER, RUDY, and J-RARE—while 12 PCIs (57%) implemented the specific consent model. This result indicates that not all PCIs embedded individual control into the design of the interface.

The use of social media technology, such as information dissemination, was observed in most of the 21 PCIs. This allows participants to see the progress in research and the results. PCIs also generally provide user-friendly platforms that facilitate two-way interaction between participants and researchers and, in some PCIs, between the participants. In more than half of the 21 PCIs, participants were involved in suggesting research questions. These results indicate that PCIs enable participants to play an active role, not only in terms of controlling their own data-sharing settings but also by way of contributing to decision making in research design. They encourage participants to become involved in the agenda setting of the entire research community. A follow-up study is needed to examine the method and evaluation of these involvement activities.

Definition of PCIs

The definition of PCIs was suggested in prior studies in 2012 [12,13], as described in our introduction. However, while conducting this scoping review, we initially found it difficult to distinguish PCIs from other related digital platforms. By analyzing the obtained results, the essential concept and features of PCIs became clearer. Therefore, in this study, we reconsidered the definition of PCIs with clearer and more rigorous criteria. Our proposed updated definition of PCIs is as follows: (1) research activities use ICT, (2) participants are actively involved in the agenda setting of research, and they play roles beyond those of research subjects or assistants, (3) participants can communicate with researchers interactively throughout the research process, and (4) participants can choose the level of their involvement.

PCIs cultivate an environment to establish collaborative partnerships between participants and researchers. As this *participant-centric* model is still emerging, the number of PCIs that are currently in practice is small, compared with other kinds of medical research that uses ICT. Employing a rigorous criteria of patient involvement, we excluded researcher-led initiatives that aim at collecting data, such as many patient registries and disease registries, because they did not indicate active involvement by patients either on their website or in their published literature. We also excluded patient registries created with the active support of patient organizations if that involvement was limited to recruiting patients to provide data. Moreover, research projects that limit patient involvement to communicating with participants to disseminate information or to respond to inquiries were also excluded, as we did not consider this to be interactive communication. Clarifying

exclusions and inclusions as part of the screening process also allowed us to clearly identify the distinctive levels of involvement and the features of PCIs.

Comparison Between PCIs in the United States, the United Kingdom, and Japan

Of the 21 PCIs identified, 19 (90%) were initiated in the United States. There may be a number of possible reasons for this, some of which suggest avenues for future investigation. Firstly, it is possible that the culture of involving patients and members of the public in research is more established in the United States.

Our results indicated that many of the PCIs were funded by PCORI. The Patient-Powered Research Network is one of the research networks that is operated and governed by patient groups, and it has launched a number of projects using online platforms to collect self-reported data [55,56]. This may suggest that differences in policies of funding organizations have influenced the number of PCIs among the three countries. In the United Kingdom, research funding bodies, such as the Medical Research Council (MRC) and the NIHR, require all applicants to state how they will involve patients and members of the public in the design of the research study, which suggests the culture for patient involvement in medical research is also reasonably well established in the United Kingdom. Nevertheless, this has not yet translated into the establishment of PCIs. Indeed, we found only one funded research project that was using a digital platform.

Furthermore, the differences in the way that health care is organized may be an additional explanatory factor, but this would need further exploration. The US system is rather unique in the way it organizes health care and this may be tied into the more populous PCI landscape in the country. We are aware that patients in the United States are often engaged around social entrepreneurship in health care and play key roles in initiating new research projects [59]. There may also be different attitudes between the three countries toward the utilization of ICT and direct-to-consumer services. Another consideration is the population size. Although the explanatory power is probably somewhat weak, the larger number of PCIs in the United States may, in part, reflect the much larger population size compared to Japan and the United Kingdom.

Future research is needed to disentangle the factors impacting the establishment and sustainability of PCIs in order to better understand the underlying reasons for the differences between these countries.

Limitations and Future Challenges

In this paper, we investigated PCIs, which enable the active involvement of patients, by conducting a scoping review to understand this new field that has not yet been conceptually established. Therefore, there are a number of limitations to this work.

First, due to limitations in the search method, some PCIs might have been overlooked in our search. For example, any PCI-related documents that do not contain our keywords indicating active involvement of patients in the titles or abstracts would have been extremely difficult to detect. This would also hold for the website search results. We suggested that consistent definition and terminology of participant involvement should be established to overcome these methodological limitations. Second, any PCIs that do not publish their research in scientific articles, or those whose websites do not appear in any search engine results, would not have been found. Further, research articles and websites that were published after our search period would also have not been included in our findings.

For future research, PCIs implemented in countries other than those focused on in this study also deserve attention. For instance, other English-speaking countries, such as Canada, Australia, and New Zealand, where there is a push toward patient and public involvement in medical research could be prioritized in future research. Furthermore, it may be desirable to seek a method to evaluate PCIs by a variety of means, such as careful observation, surveys of participant opinions, and exploration of theoretical considerations. Such results will offer insights to further improve ongoing PCIs and aid the establishment of new ones in the future.

We can also propose possible ideas for moving forward the area of PCIs. In this review, we have documented the important activities of PCIs; at the same time, we have noted that there are still only a small number of projects that meet the criteria we have set for patient involvement. Therefore, the future challenge in the field is how to expand the number of PCIs and also the level of patient involvement that these platforms enable. Ideas for improvement can include the following: (1) wider dissemination of information about the value of PCIs for patients and patient organizations, (2) encouragement of networking between PCIs to facilitate the adoption of good practices, and (3) increase of support by government bodies for PCIs.

Conclusions

A scoping review was sufficient to capture recent trends in PCIs designed to facilitate medical research. We identified 21 PCIs currently operating in the United States, the United Kingdom, and Japan. This review contributes to a better understanding of the concept of *active involvement* by patients that can be facilitated by PCIs, by clarifying the various modes of communication and decision making. Although it is an emerging initiative in medical research, PCIs have the potential to facilitate fruitful collaboration between participants and researchers.

Acknowledgments

We thank Mr Toshiyuki Suwa for helping develop our search strategy. We are also grateful to Mr Anthony Castaldo from the US Hereditary Angioedema Association (US HAEA) and Hereditary Angioedema International (HAEi), Mr Simon Denegri from the NIHR, Dr Tiffany Boughtwood from the Murdoch Children's Research Institute, Dr Mizuki Morita from Okayama University, and Dr Soichi Ogishima from Tohoku University for consultations on the obtained results as the expert panel members. We thank

Dr Jessica Bell and Dr Teresa Finlay from the University of Oxford for contributing to the initial study design. We thank Dr Saad Anis from Edanz Group for editing a draft of this manuscript. This work is supported by JSPS (Japan Society for the Promotion of Science) Grants-in-Aid for Scientific Research (KAKENHI) (grants No. JP15K15167 and No. JP17K19812), Osaka University International Joint Research Promotion Program, and the Interdisciplinary Program for Biomedical Sciences (IPBS).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The search history of the literature database search.
[DOCX File , 20 KB - jmir_v22i8e16441_app1.docx]

Multimedia Appendix 2
The list of data elements extracted from relevant articles and information on websites of each participant-centric initiative (PCI).
[DOCX File , 15 KB - jmir_v22i8e16441_app2.docx]

Multimedia Appendix 3
Detailed characteristics of identified participant-centric initiatives (PCIs).
[XLSX File (Microsoft Excel File), 15 KB - jmir_v22i8e16441_app3.xlsx]

Multimedia Appendix 4
The list of participant-centric initiatives (PCIs) that had not come out of our original literature search.
[DOCX File , 16 KB - jmir_v22i8e16441_app4.docx]

References

1. Swan M. Crowdsourced health research studies: An important emerging complement to clinical trials in the public health research ecosystem. J Med Internet Res 2012 Mar 07;14(2):e46 [FREE Full text] [doi: 10.2196/jmir.1988] [Medline: 22397809]
2. Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. Qual Life Res 2000;9(8):887-900. [doi: 10.1023/a:1008996223999] [Medline: 11284208]
3. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: The CONSORT PRO extension. JAMA 2013 Feb 27;309(8):814-822. [doi: 10.1001/jama.2013.879] [Medline: 23443445]
4. Domecq JP, Prutsky G, Elraiyah T, Wang Z, Nabhan M, Shippee N, et al. Patient engagement in research: A systematic review. BMC Health Serv Res 2014 Feb 26;14:89 [FREE Full text] [doi: 10.1186/1472-6963-14-89] [Medline: 24568690]
5. What is public involvement in research? NIHR INVOLVE. URL: https://www.invo.org.uk/find-out-more/what-is-public-involvement-in-research-2/ [accessed 2019-09-13]
6. The value of engagement. Patient-Centered Outcomes Research Institute (PCORI). 2018 Oct 30. URL: https://www.pcori.org/about-us/our-programs/engagement/public-and-patient-engagement/value-engagement [accessed 2019-09-13]
7. Aitken M, Tully MP, Porteous C, Denegri S, Cunningham-Burley S, Banner N, With input from the four public panels of the Farr Institute of Health Informatics Research (2013-2018). Consensus statement on public involvement and engagement with data-intensive health research. Int J Popul Data Sci 2019;4(1):1-6 [FREE Full text] [doi: 10.23889/ijpds.v3i4.837]
8. UK standards for public involvement. National Institute for Health Research (NIHR). URL: https://sites.google.com/nihr.ac.uk/pi-standards/standards [accessed 2020-02-03]
9. Hayes H, Buckland S, Tarpey M. Briefing Notes for Researchers: Public Involvement in NHS, Public Health and Social Care Research. Eastleigh, UK: INVOLVE; 2012 Feb. URL: http://www.invo.org.uk/wp-content/uploads/2014/11/9938_INVOLVE_Briefing_Notes_WEB.pdf [accessed 2020-02-03]
10. Manafò E, Petermann L, Vandall-Walker V, Mason-Lai P. Patient and public engagement in priority setting: A systematic rapid review of the literature. PLoS One 2018;13(3):e0193579 [FREE Full text] [doi: 10.1371/journal.pone.0193579] [Medline: 29499043]
11. Concordat for Engaging the Public with Research. Swindon, UK: UK Research and Innovation (UKRI) URL: https://www.ukri.org/files/legacy/scisoc/concordatforengagingthepublicwithresearch-pdf/ [accessed 2020-02-03]
12. Anderson N, Bragg C, Hartzler A, Edwards K. Participant-centric initiatives: Tools to facilitate engagement in research. Appl Transl Genom 2012 Dec 01;1:25-29 [FREE Full text] [doi: 10.1016/j.atg.2012.07.001] [Medline: 24772384]

13. Kaye J, Curren L, Anderson N, Edwards K, Fullerton SM, Kanellopoulou N, et al. From patients to partners: Participant-centric initiatives in biomedical research. Nat Rev Genet 2012 Apr 03;13(5):371-376 [FREE Full text] [doi: 10.1038/nrg3218] [Medline: 22473380]
14. Slingsby BT. Decision-making models in Japanese psychiatry: Transitions from passive to active patterns. Soc Sci Med 2004 Jul;59(1):83-91. [doi: 10.1016/j.socscimed.2003.10.006] [Medline: 15087145]
15. Japan Agency for Medical Research and Development. URL: https://www.amed.go.jp/en/index.html [accessed 2020-02-13]
16. Coathup V, Hamakawa N, Finlay T, Bell J, Kaye J, Kato K. Participant-centric initiatives and medical research: Scoping review protocol. JMIR Res Protoc 2017 Dec 12;6(12):e245 [FREE Full text] [doi: 10.2196/resprot.7407] [Medline: 29233800]
17. Levac D, Colquhoun H, O'Brien KK. Scoping studies: Advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
18. Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. Int J Soc Res Methodol 2005 Feb 23;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
19. Woolley JP, McGowan ML, Teare HJA, Coathup V, Fishman JR, Settersten RA, et al. Citizen science or scientific citizenship? Disentangling the uses of public engagement rhetoric in national research initiatives. BMC Med Ethics 2016 Jun 04;17(1):33 [FREE Full text] [doi: 10.1186/s12910-016-0117-1] [Medline: 27260081]
20. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
21. Lambertson KF, Damiani SA, Might M, Shelton R, Terry SF. Participant-driven matchmaking in the genomic era. Hum Mutat 2015 Oct;36(10):965-973. [doi: 10.1002/humu.22852] [Medline: 26252162]
22. Kirkpatrick BE, Riggs ER, Azzariti DR, Miller VR, Ledbetter DH, Miller DT, ClinGen Resource. GenomeConnect: Matchmaking between patients, clinical laboratories, and researchers to improve genomic knowledge. Hum Mutat 2015 Oct;36(10):974-978 [FREE Full text] [doi: 10.1002/humu.22838] [Medline: 26178529]
23. Frost JH, Massagli MP, Wicks P, Heywood J. How the social web supports patient experimentation with a new therapy: The demand for patient-controlled and patient-centered informatics. AMIA Annu Symp Proc 2008 Nov 06:217-221 [FREE Full text] [Medline: 18999176]
24. Nishimura Y, Nishimura K, Ogishima S, J-RARE Patient Group Consortium. Initiatives in the patient-centered patient registry J-RARE.net. In: Proceedings of the 2nd Annual Meeting of The Japanese Society of Medical Networking for Intractable Diseases. 2014 Presented at: The Japanese Society of Medical Networking for Intractable Diseases; November 14-15, 2014; Kagoshima, Japan p. 68.
25. Morita M, Nishimura K, Ogishima S, Oda Y, Inoi K, Nagamatsu K, et al. Patient-Driven Patient Registry Study: Research Report on Health Labour Sciences Research Grant Research on Measures for Intractable Diseases [report in Japanese]. 2013. URL: http://www.nanbyo.jp/kenkyu/hokoku/H24siryo/51-63.pdf [accessed 2020-07-09]
26. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
27. Teare HJA, Hogg J, Kaye J, Luqmani R, Rush E, Turner A, et al. The RUDY study: Using digital technologies to enable a research partnership. Eur J Hum Genet 2017 Jun;25(7):816-822 [FREE Full text] [doi: 10.1038/ejhg.2017.57] [Medline: 28443622]
28. Fleurence RL, Beal AC, Sheridan SE, Johnson LB, Selby JV. Patient-powered research networks aim to improve patient care and health research. Health Aff (Millwood) 2014 Jul;33(7):1212-1219. [doi: 10.1377/hlthaff.2014.0113] [Medline: 25006148]
29. PCORnet PPRN Consortium, Daugherty SE, Wahba S, Fleurence R. Patient-powered research networks: Building capacity for conducting patient-centered clinical outcomes research. J Am Med Inform Assoc 2014;21(4):583-586 [FREE Full text] [doi: 10.1136/amiajnl-2014-002758] [Medline: 24821741]
30. 23andMe. URL: https://www.23andme.com/en-int/?myg02=true [accessed 2018-01-14]
31. PatientsLikeMe. URL: https://www.patientslikeme.com/ [accessed 2018-01-14]
32. Promise for Engaging Everyone Responsibly (PEER). Genetic Alliance. URL: http://www.geneticalliance.org/programs/biotrust/peer [accessed 2018-01-14]
33. GenomeConnect. URL: https://www.genomeconnect.org/ [accessed 2018-01-14]
34. RUDY Study. URL: https://research.ndorms.ox.ac.uk/rudy/ [accessed 2018-01-14]
35. MoodNetwork. URL: https://moodnetwork.org/ [accessed 2018-01-14]
36. mPower. URL: https://parkinsonmpower.org/ [accessed 2018-01-14]
37. J-RARE. URL: https://j-rare.net/ [accessed 2018-01-14]
38. ABOUT Network. URL: https://www.aboutnetwork.org/ [accessed 2018-04-16]
39. ArthritisPower. URL: https://arthritispower.creakyjoints.org/ [accessed 2018-04-16]
40. IBD Partners. URL: https://ccfa.med.unc.edu/ [accessed 2018-04-18]
41. Rare Epilepsy Network (REN). Epilepsy Foundation. URL: https://www.epilepsy.com/make-difference/research-and-new-therapies/engagement/rare-epilepsy-network-ren [accessed 2018-03-17]
42. COPD PPRN. URL: https://www.copdpprn.org/portal/Login?ReturnUrl=%2Fportal%2F#/ [accessed 2018-03-17]

43. Health eHeart Study. URL: https://www.health-eheartstudy.org/ [accessed 2018-03-18]
44. Interactive Autism Network (IAN). Kennedy Krieger Institute. URL: https://www.kennedykrieger.org/research/centers-labs-cores/the-interactive-autism-network-ian [accessed 2018-03-12]
45. iConquerMS. URL: https://www.iconquerms.org/ [accessed 2018-03-24]
46. National Alzheimer's & Dementia Patient- & Caregiver-Powered Research Network. Patient-Centered Outcomes Research Institute (PCORI). URL: https://www.pcori.org/research-results/2015/national-alzheimers-dementia-patient-caregiver-powered-research-network [accessed 2020-07-08]
47. The NephCure Kidney Network Patient Registry (NKN). NephCure Kidney International. 2015 Dec 15. URL: https://nephcure.org/2015/12/the-nephcure-kidney-network-patient-registry-nkn/ [accessed 2018-06-14]
48. IDF PI CONNECT. Immune Deficiency Foundation (IDF). URL: https://primaryimmune.org/pi-connect [accessed 2018-06-28]
49. Vasculitis Patient-Powered Research Network. URL: https://www.vpprn.org/ [accessed 2018-06-28]
50. MyApnea. URL: https://myapnea.org/ [accessed 2018-06-28]
51. Publications. 23andMe. URL: https://www.23andme.com/publications/for-scientists/ [accessed 2019-07-12]
52. Research Manuscripts Bibliography: The Complete Collection of PatientsLikeMe Research Publications. Cambridge, MA: PatientsLikeMe; 2019 Jan. URL: https://patientslikeme-bibliography.s3.amazonaws.com/PLM%20Research%20Manuscripts%20Bibliography.pdf [accessed 2019-08-06]
53. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: A patient interface for twenty-first century research networks. Eur J Hum Genet 2015 Feb;23(2):141-146 [FREE Full text] [doi: 10.1038/ejhg.2014.71] [Medline: 24801761]
54. Cukor D, Cohen LM, Cope EL, Ghahramani N, Hedayati SS, Hynes DM, et al. Patient and other stakeholder engagement in Patient-Centered Outcomes Research Institute funded studies of patients with kidney diseases. Clin J Am Soc Nephrol 2016 Sep 07;11(9):1703-1712 [FREE Full text] [doi: 10.2215/CJN.09780915] [Medline: 27197911]
55. Cooperative Agreement Funding Announcement: Improving Infrastructure for Conducting Patient-Centered Outcomes Research. The National Patient-Centered Clinical Research Network: Patient-Powered Research Networks (PPRN)—Phase One. Washington, DC: Patient-Centered Outcomes Research Institute (PCORI); 2013 Apr 23. URL: https://www.pcori.org/assets/PCORI-PPRN-Funding-Announcement-042313.pdf [accessed 2020-02-12]
56. Pletcher MJ, Forrest CB, Carton TW. PCORnet's Collaborative Research Groups. Patient Relat Outcome Meas 2018;9:91-95 [FREE Full text] [doi: 10.2147/PROM.S141630] [Medline: 29467584]
57. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: A new era in clinical research. Perspect Clin Res 2011 Oct;2(4):137-144 [FREE Full text] [doi: 10.4103/2229-3485.86879] [Medline: 22145124]
58. Haghi M, Thurow K, Stoll R. Wearable devices in medical internet of things: Scientific research and commercially available devices. Healthc Inform Res 2017 Jan;23(1):4-15 [FREE Full text] [doi: 10.4258/hir.2017.23.1.4] [Medline: 28261526]
59. Terry SF, Terry PF, Rauen KA, Uitto J, Bercovitch LG. Advocacy groups as research organizations: The PXE International example. Nat Rev Genet 2007 Feb;8(2):157-164. [doi: 10.1038/nrg1991] [Medline: 17230202]

Abbreviations

CINAHL: Cumulative Index of Nursing and Allied Health Literature
Embase: Excerpta Medica Database
HAEi: Hereditary Angioedema International
ICT: internet and communication technology
IPBS: Interdisciplinary Program for Biomedical Sciences
JSPS: Japan Society for the Promotion of Science
KAKENHI: Grants-in-Aid for Scientific Research
MEDLINE: Medical Literature Analysis and Retrieval System Online
MeSH: Medical Subject Headings
MRC: Medical Research Council
NHS: National Health Service
NIH: National Institutes of Health
NIHR: National Institute for Health Research
PCI: participant-centric initiative
PCORI: Patient-Centered Outcomes Research Institute
PEER: Promise for Engaging Everyone Responsibly
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RUDY: Rare and Undiagnosed Diseases Study
US HAEA: US Hereditary Angioedema Association
V-PPRN: Vasculitis Patient-Powered Research Network

Edited by G Eysenbach; submitted 30.09.19; peer-reviewed by K Jones, K Fuji; comments to author 18.11.19; revised version received 13.02.20; accepted 03.06.20; published 04.08.20.

Please cite as:
Hamakawa N, Nakano R, Kogetsu A, Coathup V, Kaye J, Yamamoto BA, Kato K
Landscape of Participant-Centric Initiatives for Medical Research in the United States, the United Kingdom, and Japan: Scoping Review
J Med Internet Res 2020;22(8):e16441
URL: https://www.jmir.org/2020/8/e16441
doi:10.2196/16441
PMID:32749228

Original Paper

Telemedicine in Germany During the COVID-19 Pandemic: Multi-Professional National Survey

Arne Peine[1*], MHBA, MD; Pia Paffenholz[2*], MD; Lukas Martin[1], MHBA, MD, PhD; Sandra Dohmen[1], MD; Gernot Marx[1], Prof Dr, FRCA; Sven H Loosen[3,4], MHBA, MD

[1]Department of Intensive Care and Intermediate Care, University Hospital Rheinisch Westfaelische Technische Hochschule Aachen, Aachen, Germany

[2]Department of Urology, Uro-Oncology, Robot Assisted and Reconstructive Urologic Surgery, University Hospital Cologne, Cologne, Germany

[3]Department of Medicine III, University Hospital Aachen, Aachen, Germany

[4]Clinic for Gastroenterology, Hepatology and Infectious Diseases, University Hospital Düsseldorf, Medical Faculty of Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[*]these authors contributed equally

Corresponding Author:
Arne Peine, MHBA, MD
Department of Intensive Care and Intermediate Care
University Hospital Rheinisch Westfaelische Technische Hochschule Aachen
Pauwelsstraße 30
Aachen, 52072
Germany
Phone: 49 2418038009
Email: apeine@ukaachen.de

Abstract

Background: In an effort to contain the effects of the coronavirus disease (COVID-19) pandemic, health care systems worldwide implemented telemedical solutions to overcome staffing, technical, and infrastructural limitations. In Germany, a multitude of telemedical systems are already being used, while new approaches are rapidly being developed in response to the crisis. However, the extent of the current implementation within different health care settings, the user's acceptance and perception, as well as the hindering technical and regulatory obstacles remain unclear.

Objective: The aim of this paper is to assess the current status quo of the availability and routine use of telemedical solutions, user acceptance, and the subjectively perceived burdens on telemedical approaches. Furthermore, we seek to assess the perception of public information quality among professional groups and their preferred communication channels.

Methods: A national online survey was conducted on 14 consecutive days in March and April 2020, and distributed to doctors, nurses, and other medical professionals in the German language.

Results: A total of 2827 medical professionals participated in the study. Doctors accounted for 65.6% (n=1855) of the professionals, 29.5% (n=833) were nursing staff, and 4.9% (n=139) were identified as others such as therapeutic staff. A majority of participants rated the significance of telemedicine within the crisis as high (1065/2730, 39%) or neutral (n=720, 26.4%); however, there were significant differences between doctors and nurses (P=.01) as well as between the stationary sector compared to the ambulatory sector (P<.001). Telemedicine was already in routine use for 19.6% (532/2711) of German health care providers and in partial use for 40.2% (n=1090). Participants working in private practices (239/594, 40.2%) or private clinics (23/59, 39.0%) experienced less regulatory or technical obstacles compared to university hospitals (586/1190, 49.2%). A majority of doctors rated the public information quality on COVID-19 as good (942/1855, 50.8%) or very good (213/1855, 11.5%); nurses rated the quality of public information significantly lower (P<.001). Participant's age negatively correlated with the perception of telemedicine's significance (ρ=–0.23; P<.001).

Conclusions: Telemedicine has a broad acceptance among German medical professionals. However, to establish telemedical structures within routine care, technical and regulatory burdens must be overcome.

(J Med Internet Res 2020;22(8):e19745) doi:10.2196/19745

KEYWORDS

telemedicine; coronavirus; COVID-19; telehealth; SARS-CoV-2; pandemic; survey; medical professional; availability; acceptance; burden

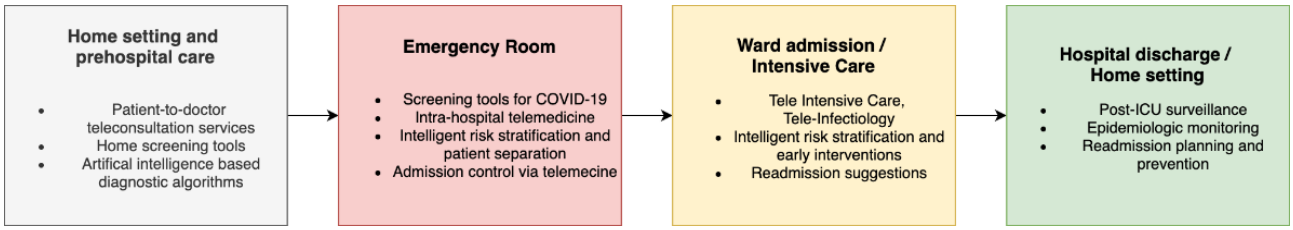
Introduction

Background

The global pandemic caused by the severe acute respiratory syndrome coronavirus 2 is creating a historic challenge for health care providers, patients, and societies throughout the world. Hospitals are drastically increasing intensive care capacities in an effort to contain the effects of the pandemic. However, staffing, technical, and infrastructural limitations are impeding progress in this regard. With an estimated intensive care unit (ICU) hospitalization rate of 5%, the pandemic quickly surpassed the global hospital care capacities [1]. In response to the crisis, new approaches are urgently needed to avoid a medical crisis.

To bring specialist care to the patients, diverse telemedical approaches have been implemented into patient care routine worldwide in both the ambulatory and hospital sectors (from the home setting to admission, treatment, and discharge), and adaptations have been developed for each use case (Figure 1).

Figure 1. Levels of telemedical interventions during the COVID-19 crisis. COVID-19: coronavirus disease; ICU: intensive care unit.



Telemedicine in Acute Emergency and Intensive Care

In the area of acute care medicine, specifically intensive care medicine, telemedicine has proven to be a success story. For example, the introduction of a remote intensivist program in two US tertiary care hospitals has led to a significant reduction of mortality (9.4% vs 12.9%; relative risk 0.73; 95% CI 0.55-0.95) and has proven to be cost-effective [2].

As a result of the worldwide shortage of medical professionals, not all patients are treated under the supervision of a specialized doctor. Although, it is estimated that, if specialized ICU physician staffing was implemented in nonrural US hospitals, approximately 53,000 lives and US $5.4 billion would be saved annually; as of 2010, only 10%-15% of the US ICUs were able to provide intensivist care, clearly a resource urgently needed in response to the coronavirus disease (COVID-19) [3]. Multiple studies have shown that providing a dedicated intensivist at an ICU leads to a significant reduction in mortality and reduces the length of stay [4,5]. Worldwide, telemedical tools have been rapidly adopted to the intensive care setting, providing specialist telemedical guidance to remote hospitals.

The advantages in the reduction of distance barriers between patients and physicians are also used to improve access to high-level intensive care in otherwise medically underserved areas. In adult intensive care wards, the introduction of telemedical surveillance by a specialized intensivist reduced severity-adjusted mortality by 33% and 30%, and the incidence of ICU complications by 44% and 50% in two intervention periods in an observational time series cohort study [6]. A recently published large retrospective study in the United States showed similar results as well for adult step-down or progressive care units, where patients in the telemedical intervention group had a survival benefit of 20% and had a significantly lower length of stay [7]. Consequently, telemedical solutions are also used for “in-house screening” of patients with COVID-19 (eg, by distributing tablet computers in emergency departments), minimizing the time of direct patient contact and, thus, cutting down the infection risk [8,9]. These findings were supported by a systematic review and meta-analysis of 13 studies involving 35 ICUs using a pre-post-design. The authors concluded that there was a reduction in ICU mortality (pooled odds ratio 0.80, 95% CI 0.66-0.97; *P*=.02) and a reduction in length of stay but stated that in-hospital mortality was not proven to be significantly reduced (pooled odds ratio 0.82, 95% CI 0.65-1.03; *P*=.08) [10].

Numerous worldwide experiences have shown that the formation of “telemedical excellence centers” is an efficient and fast way to provide telemedical specialist care to large populations. This is especially true for the reaction to global crises such as the coronavirus pandemic, as these telemedical centers can be created rapidly, concentrate specialist care locally, and deliver the highest quality care within large regions without travel restrictions or risk of infection for medical staff [11]. For example, in reaction to the pandemic and under support of the federal government of North-Rhine Westphalia in Germany, the University Hospitals of Aachen and Münster have built up a “virtual hospital” structure within weeks that provides a day-and-night availability of specialist intensivist and infectologist care for over 200 regional hospitals [12-14]. In China, the National Telemedicine Center of China in Zhengzhou has established a telemedicine-enabled outbreak alert and response network, connecting over 120 smaller hospitals [15]. In rapid response to the COVID-19 pandemic, a telemedical network was created in the Sichuan Province in Western China [11]. The first results of the retrospective success analysis showed telemedicine to be a “feasible, acceptable, and effective” way to provide health care and “allowed for significant improvements in health care outcomes.” Furthermore, entirely new, data-driven disease containment strategies using contact

tracing–based mobile sensors were rapidly developed and established to track and impede chains of infection [16].

Telemedicine in the Ambulatory Sector

In addition, within the ambulatory sector (eg, in home care and outpatient clinics), telemedical solutions are on the rise worldwide in response to the coronavirus crisis [17]. Over 50 US hospitals established or reinforced their telemedical health systems to allow clinicians to see patients who are at home without the risk of infection for medical professionals [8]. This might also be true for non–infection-related consultations (eg, in orthopedics [18] or chronic conditions [19]), reducing the need for repeated physical patient-physician contact and, consequently, the risk of cross-infections within a practice visit. A tool for the initial medical assessment of COVID-19 has been made available to German citizens by the German Central Institute for Statutory Health Insurance Physicians. The "COVID-Guide" is intended to enable patients to make an initial assessment of their own health situation in the event of possible complaints and uncertainties in connection with the coronavirus, which serves to accompany patients at home with telemedical means and, thus, recognize the occurrence of specific alarm symptoms at an early stage [20]. In contrast to other countries (eg, the Netherlands, the United Kingdom), German patients have generally free access to all medical specialties. A referral is not mandatory from the patient's point of view. A patient can consult a specialist immediately and does not have to take a detour via a general practitioner. In the case of so-called family doctor–centered care, members of the statutory health insurance are bound to contract doctors of the health insurance companies when choosing a (telemedical) doctor. Telemedical consultations are mostly performed through direct patient-doctor contact.

To establish telemedicine in routine care, user acceptance by medical professionals is of the utmost importance to make effective use of telemedical resources. Furthermore, infrastructural problems (eg, broadband internet connectivity, organizational structures) are still thought to be a relevant factor hindering effective implementation of telemedical care [21]. To our knowledge, no structured, large-scale assessment of the perception of telemedical services within the crisis has been carried out accounting for differences between levels of care and medical professional groups. In this study, we examine the current status quo of the availability and routine use of telemedical solutions, the user acceptance, and the subjectively perceived burdens on telemedical approaches. Telemedical providers often serve as an additional source for medical knowledge. This is particularly important in crises, where the distribution of reliable information is key. Consequently, we assess the perception of the quality of public information among professional groups and their preferred information channels to evaluate efficient communication with the aforementioned providers.

Methods

Data Collection and Recruitment

Survey data collection took place on all days of the evaluated time frame between March 27, 2020, and April 11, 2020. Participants were acquired via numerous communication channels, taking into account the heterogeneous access and technical skills of different medical professional groups. Access (weblink) to the survey system was distributed through official communication channels and mailing lists of numerous medical societies and social media groups. Furthermore, several large- and medium-sized hospitals shared the weblink to the survey within their internal communication systems (eg, intranet platforms). The telemedical survey was part of a larger survey on coronavirus conducted within the time frame.

Data acquisition took place via a publicly accessible, web-based survey system (LimeSurvey, version 3.22.10; LimeSurvey GmbH) based on the programming language Hypertext Preprocessor (version 7.1.33, The PHP Group) and JavaScript (version 262, June 2017, ECMA International). The survey was accessible through all common web browsers as well as mobile phones.

All computational infrastructure was hosted on an Apache server (The Apache Software Foundation), while the system was physically hosted in Nuremberg, Germany to comply with all European data protection laws. No technical failure or server downtime was observed during the acquisition period. Data was stored using a My Structured Query Language (MySQL) database (Oracle Corporation).

Statistical Analysis

All statistical preprocessing and analysis was carried out using the R statistical programming language [22] (Version 3.6.3, R Foundation for Statistical Computing) and the statistical software Jamovi (Version 1.2.16.0, The Jamovi Project) [23].

Appropriate sample size was calculated as described in Multimedia Appendix 1. A confidence level of 99% and an acceptable error margin of 3% were assumed [24].

The descriptive statistical data is laid out in total numbers as well as in relative percentages.

Assumption of normal distribution was tested using the Shapiro-Wilk test.

We used a nonparametic one-way analysis of variance to address potential differences between the professional groups. Post-hoc analysis was performed using Dwass-Steel-Chritchlow-Figner pairwise comparisons, when appropriate. Correlation analysis for ordinal variables was carried out using Spearman rho.

Results

Study Demographics and Survey Size

A total of 2927 participants took part in the survey within the observed time frame between March 27, 2020, and April 11, 2020. We calculated the ideal sample size for the survey to reach significance as described in the methods section. We assumed there were a total of 5.6 million medical professionals employed within the German health care domain, as reported by the German Federal Office of Statistics [25]. On a confidence level of 99% and within an acceptable error margin of 3%, the required sample rate was determined to be 1849 participants.

The median observed age was 43.0 (SD 11.8) years. Of the 2827 participants, 51.1% (n=1446) of participants identified as

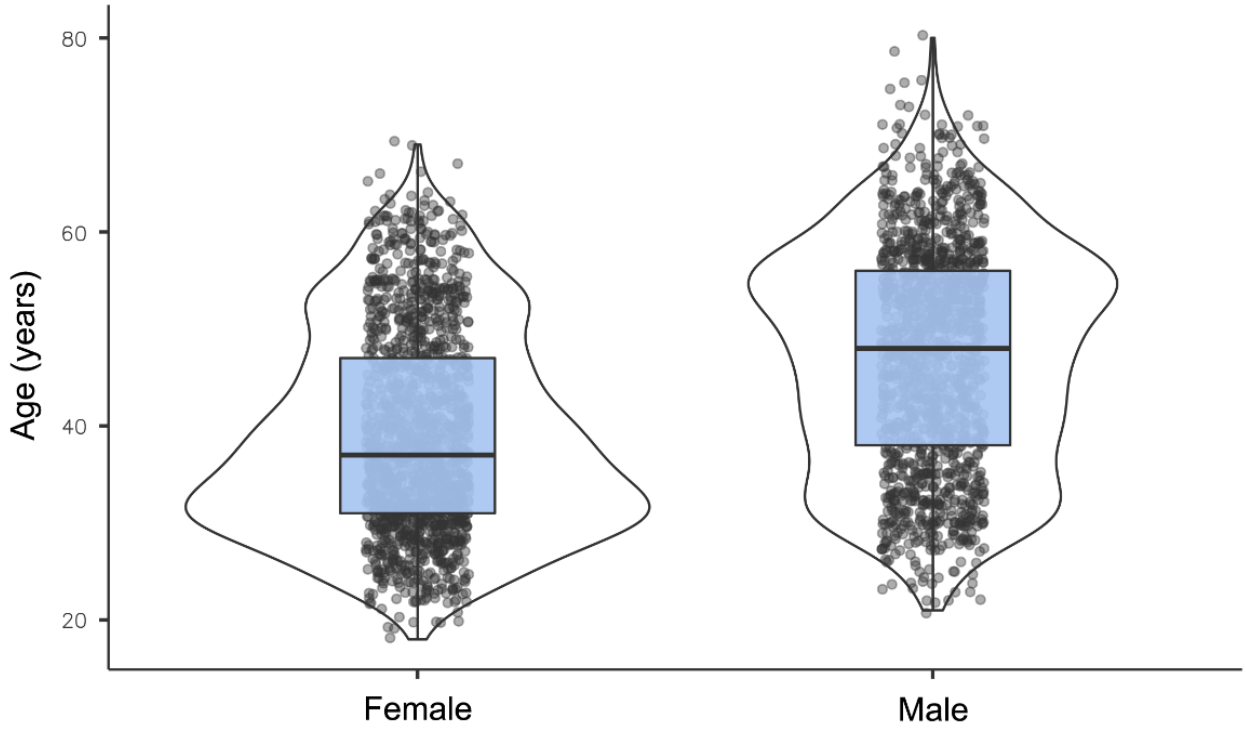
females, 47.6% (n=1345) identified as males, and 0.3% (n=69) chose not to disclose gender. The majority of the survey participants worked within a university hospital setting (n=1238, 43.8%), while 26.9% (n=749) worked within smaller regional hospitals. The ambulatory sector was represented by 21.6% (n=611) participants within private practices, 2.1% (n=60) within private clinics, and 1.6% (n=45) in a rehabilitation clinic setting. A total of 111 (3.9%) participants disclosed their work environment as "other" (eg, ambulatory nursing services). When asked about their profession, 65.6% (n=1855) identified as doctors, 29.5% (n=833) as nursing staff, and 4.9% (n=139) as other medical professionals such as therapeutic staff; Table 1).

Visualization of the gender distribution already showed indications of a non-Gaussian distribution. We further assessed the pattern by performing a Shapiro-Wilk test, resulting in a statistic value of 0.98 (P<.001), thus confirming the hypothesis of a nonnormal distribution (Figure 2).

Table 1. Demographic data of the study population.

Variable	Value
Participant total, N	2827
Age (years), mean (SD)	43.0 (11.8)
Gender, n (%)	
Female	1446 (51.1)
Male	1345 (47.6)
Missing/nondisclosed	69 (0.3)
Work setting, n (%)	
University hospital, high level care	1238 (43.8)
Regional hospital	749 (26.5)
Ambulatory care/medical practice	611 (21.6)
Private clinic	60 (2.1)
Rehabilitation clinic	45 (1.6)
Others (eg, ambulatory nursing service)	111 (3.9)
Missing/nondisclosed	13 (0.4)
Hospital environment, if working within hospital, n (%)	
Outpatient clinic	283 (10.0)
Standard care ward	750 (26.5)
Intensive care ward	486 (17.2)
Operating theater	401 (14.2)
Diagnostics	148 (5.2)
Missing/nondisclosed	759 (27.8)
Professional group, n (%)	
Doctors	1855 (65.6)
Nursing staff	833 (29.5)
Others (eg, therapeutic staff)	139 (4.9)

Figure 2. Age distribution split by gender within the study participants.



Perception on the Use of Telemedicine and Telehealth Infrastructure

Participants were asked to estimate the significance of telemedicine during the coronavirus crisis ("How high do you estimate the significance of telemedicine and teleconsultations in the current crisis?"). Of the 2730 participants that responded to this question, the importance of telemedicine and teleconsultation was rated as high (n=1065, 39%) to neutral (n=720, 26.4%). The majority of the doctors (1036/1806, 57.4%), nurses (508/797, 63.8%), and other medical professionals (90/127, 70.9%) estimated the significance of telemedicine and teleconsultation within the COVID-19 crisis as either high (4) or very high (5).

A subcohort analysis of the different groups, however, showed significant (χ^2_5=15.8; *P*<.001) differences in the perception of telemedicine. In a subcohort analysis, differences between doctors and nurses (*W*=4.05, *P*=.01) as well as between doctors and others (*W*=4.36, *P*<.001) were shown to be significant, while differences between nurses and others did not reach statistical significance (*W*=2.62, *P*=.15).

Closer analysis of the perception of telemedicine between health care professionals (Figure 3a) working in a diverse work environment (Figure 3b) revealed different response patterns within the observed groups. One difference was the significantly higher perception of telemedicine amongst health care workers within the stationary and hospital sector (χ^2_5=190; *P*<.001) compared to the ambulatory and practice sector, and a significantly higher estimation in university hospital workers (*W*=–18.39, *P*<.001) and regional hospital workers (*W*=–9.99, *P*<.001) compared to private practices (Table 2).

Further, correlation analysis revealed a significant negative correlation of the participant's age with the perception of telemedicine's significance (p=–0.23; *P*<.001).

Figure 3. Subgroup analysis: relative significance of telemedicine among professional groups (A) and different work settings (B).

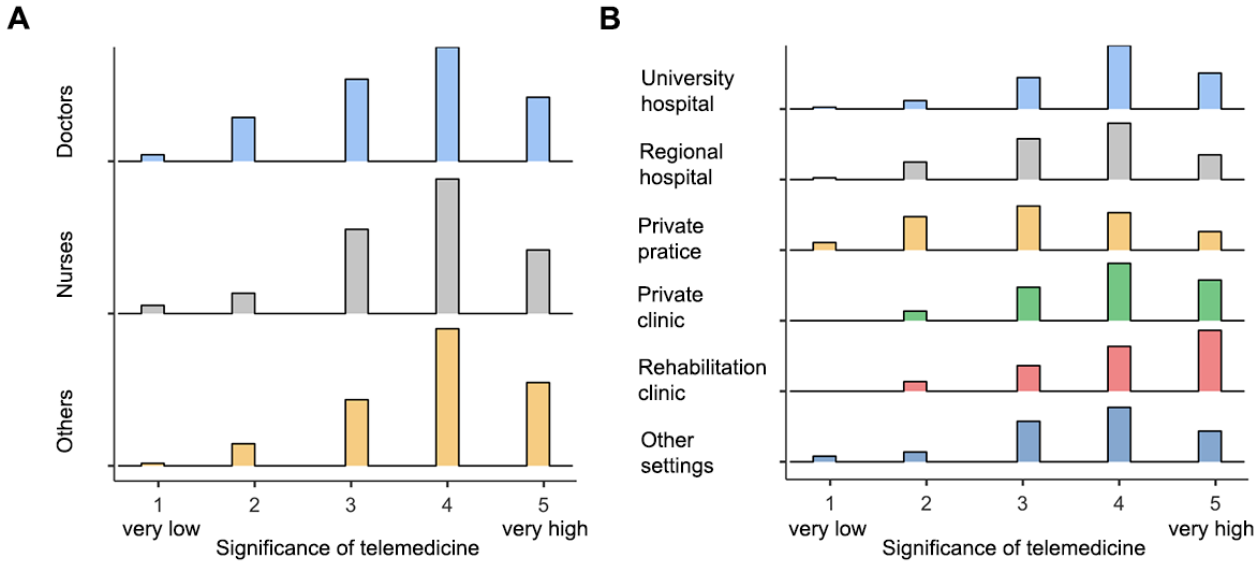


Table 2. Significance of telemedicine in the coronavirus crisis.

Profession	How high do you estimate the significance of telemedicine in the current crisis?					
	1 (very low), n (%)	2, n (%)	3, n (%)	4, n (%)	5 (very high), n (%)	Total, n (%)
Doctors	38 (2.1)	255 (14.1)	477 (26.4)	664 (36.8)	372 (20.6)	1806 (100.0)
Nurses	21 (2.6)	52 (6.5)	216 (27.1)	345 (43.3)	163 (20.5)	797 (100.0)
Others	1 (0.8)	9 (7.1)	27 (21.3)	56 (44.1)	34 (26.8)	127 (100.0)
Total	60 (2.2)	316 (11.6)	720 (26.4)	1065 (39.0)	569 (20.8)	2730 (100.0)

Availability of Telemedical Infrastructure and Establishment in Daily Routine

We further assessed the current availability of telemedical services during the pandemic. Despite being acknowledged as a significant tool during the COVID-19 crisis by the majority of participants, telemedicine was not yet part of the daily routine in most of the work environments. In only 20.2% (240/1189) of the university hospitals, 20.3% (12/59) of the private clinics, and 5.6% (41/726) of the regional hospitals telemedical consulting was in routine use. In contrast, 36% (214/595) of the participants working in a private practice setting already used telemedical tools routinely. Further analysis revealed a large proportion of hospitals that did not have access to telemedicine at all (405/1189, 34.1% for university hospitals and 393/726, 54.1% for regional hospitals; Table 3).

Furthermore, significance analysis and post hoc analysis revealed significant differences (χ^2_4=295, P<.001) between university hospitals and regional hospitals (W=–15.26, P<.001), between private practices and private clinics (W=–5.04, P=.01), and between private clinics and rehabilitation clinics (W=–5.88, P<.001).

To further investigate the reasons for the low use of telemedical tools, we addressed potential regulatory or technical obstacles within the participants' work environment. Although participants working in private practices (239/594, 40.2%) or private clinics (23/59, 39.0%) experienced no regulatory or technical obstacles, most of the medical professionals working in university hospitals (586/1190, 49.2%) experienced at least partial obstacles. In total, only 22.7% (616/2711) answered the question "Do you experience regulatory or technical obstacles for telemedicine within you work environment?" with "yes" (Table 4).

Table 3. Availability of telemedicine in different medical settings.

Answer	Do you have the possibility to reduce direct patient contact by using telemedicine?						
	University hospital, n (%)	Regional hospital, n (%)	Private practice, n (%)	Private clinic, n (%)	Rehab clinic, n (%)	Others, n (%)	Total, n (%)
None at all	405 (34.1)	393 (54.1)	88 (14.8)	21 (35.6)	34 (77.3)	47 (48.0)	988 (36.4)
In rare cases	471 (39.6)	279 (38.4)	284 (47.7)	25 (42.4)	8 (18.2)	23 (23.5)	1090 (40.2)
Daily routine	240 (20.2)	41 (5.6)	214 (36.0)	12 (20.3)	1 (2.3)	24 (24.5)	532 (19.6)
Others	73 (6.1)	13 (1.8)	9 (1.5)	1 (1.7)	1 (2.3)	4 (4.1)	101 (3.7)
Total	1189 (100.0)	726 (100.0)	595 (100.0)	59 (100.0)	44 (100.0)	98 (100.0)	2711 (100.0)

Table 4. Subcohort analysis: regulatory or technical obstacles for telemedicine within different work environments.

Answer	Do you experience regulatory or technical obstacles for telemedicine within you work environment?						
	University hospital, n (%)	Regional hospital, n (%)	Private practice, n (%)	Private clinic, n (%)	Rehab clinic, n (%)	Others, n (%)	Total, n (%)
None	357 (30.0)	262 (36.1)	239 (40.2)	23 (39.0)	19 (43.2)	34 (34.7)	934 (34.5)
Partially	586 (49.2)	281 (38.7)	214 (36.0)	23 (39.0)	13 (29.5)	44 (44.9)	1161 (42.8)
Yes	247 (20.8)	183 (25.2)	141 (23.7)	13 (22.0)	12 (27.3)	20 (20.4)	616 (22.7)
Total	1190 (100.0)	726 (100.0)	594 (100.0)	59 (100.0)	44 (100.0)	98 (100.0)	2711 (100.0)

Perception of Information Quality and Quantity

We assessed the perception of information quality and quantity within the different professional groups and the preferred communication channels to identify the most appropriate information strategy in a pandemic situation. Information was rated on a 5-point Likert scale, ranging from very poor (1) to very good (5). A majority of the participants rated the information quality either neutral (606/2827, 21.4%) or good (n=1341, 47.4%). Information quantity of public information concerning the coronavirus crisis was rated similar with neutral (981/2827, 34.7%) or good (n=1101, 38.9%). The detailed subgroup analysis, however, revealed that, although the majority of doctors rated the information quality as good (942/1855, 50.8%) or very good (n=213, 11.5%), a majority of the participating nurses rated the quality of public information lower on a Likert scale with 3 (neutral; 233/833, 28.0%) or 4 (good; n=330, 39.6%). When addressing the satisfaction on information quantity, the differences were smaller; 42.0% (779/1855) of the doctors and 33.4% (278/833) of nursing staff rated the information quantity with a 4 (“good”; Tables 5 and 6).

We performed a subcohort analysis to address the significance of differences between the professional groups. Differences between groups were significant for both quality (χ^2_2=69.3; *P*<.001) and quantity (χ^2_2=47.9; *P*<.001). Post hoc analysis with Dwass-Steel-Chritchlow-Figner pairwise comparisons showed significant differences in perception of information quality between doctors and nursing staff (*P*<.001), and nursing staff and other groups (*P*<.001), while differences between nurses and others were not shown to be significant. Addressing the information quantity, significant differences between doctors and nursing staff (*P*<.001), and doctors and others (*P*=.003) was shown, while the differences between nursing staff and other medical professionals were not significant (*P*=.96; Table 7).

We further assessed the main information sources for information within the COVID-19 pandemic. Out of the 2733 participants that responded to this question, a total of 83.6% (n=2284) of medical professionals used websites, 61.1% (n=1671) used television, and 30.7% (n=840) used newspapers within their top 3 sources of information on the COVID-19 pandemic. Social media was relatively low in use (n=388, 14.2%), and 34.6% (n=945) informed themselves through colleagues (Table 8).

Table 5. Perception of the information quality on the coronavirus among professional groups.

Professional group	How satisfied are you with the public information quality on coronavirus?					
	1 (very poor), n (%)	2, n (%)	3, n (%)	4, n (%)	5 (very good), n (%)	Total, n (%)
Doctors	119 (6.4)	235 (12.7)	346 (18.7)	942 (50.8)	213 (11.5)	1855 (100.0)
Nurses	61 (7.3)	165 (19.8)	233 (28.0)	330 (39.6)	44 (5.3)	833 (100.0)
Others	6 (4.3)	21 (15.1)	27 (19.4)	69 (49.6)	16 (11.5)	139 (100.0)
Total	186 (6.6)	421 (14.9)	606 (21.4)	1341 (47.4)	273 (9.7)	2827 (100.0)

Table 6. Perception of the information quantity on the coronavirus among professional groups.

Professional group	How satisfied are you with the public information quantity on coronavirus?					
	1 (very poor)	2	3	4	5 (very good)	Total
Doctors	19 (1.0)	195 (10.5)	697 (37.6)	779 (42.0)	165 (8.9)	1855 (100.0)
Nurses	30 (3.6)	219 (26.3)	225 (27.0)	278 (33.4)	81 (9.7)	833 (100.0)
Others	4 (2.9)	23 (16.5)	59 (42.4)	44 (31.7)	9 (6.5)	139 (100.0)
Total	53 (1.9)	437 (15.5)	981 (34.7)	1101 (38.9)	255 (9.0)	2827 (100.0)

Table 7. Post hoc Dwass-Steel-Critchlow-Fligner pairwise comparison.

Pairwise comparisons	Quality of coronavirus information		Quantity of coronavirus information	
	W	*P* value	W	*P* value
Doctors x nurses	–11.653	<.001	–9.252	<.001
Doctors x others	–0.139	>.99	–4.587	.003
Nurses x others	5.137	<.001	0.388	0.96

Table 8. Information sources about the pandemic used by participants; responses to the question "What are the top 3 sources you mainly use to inform yourself about the current coronavirus disease pandemic?" (N=2827).

Answer	n (%)
Websites	2284 (83.6)
Television	1671 (61.1)
Newspapers	840 (30.7)
Social media	388 (14.2)
Podcasts	563 (20.6)
Colleagues	945 (34.6)
Email newsletters	892 (32.6)
Mobile phone apps	191 (7.0)
Friends and relatives	265 (9.7)
Other sources	241 (8.8)

Discussion

In reaction to the COVID-19 pandemic, telemedicine has emerged as a key technology to bring high-level medical care to patients while reducing the transmission of COVID-19 among patients, families, and clinicians. We conducted this study to assess the extent of the current implementation within different health care settings, user acceptance and perception, public information politics, and regulatory burdens that are potentially impeding implementation and consequently withholding lifesaving telemedical treatment from patients who are infected that require medical attention.

Within this study, we asked a total of 2927 German medical professionals about their perceptions of telemedicine and telehealth during the current coronavirus crisis. With Germany maintaining one of the largest health care sectors worldwide, this survey can also be seen as a blueprint for other industrialized nations with similar infrastructural resources.

The study received broad interests and was supported by numerous medical societies, quickly surpassing the required 1849 participants needed to reach the calculated study sample size. The perception of telemedicine during the current COVID-19 crisis was generally high throughout all professional groups. This is in line with several studies addressing acceptance of telemedical solutions, generally supporting the high significance in clinical routine for both nurses [26] and primary health care providers [27].

Almost all medical specialties and professional societies have developed COVID-19 specific telemedical approaches catered to their specific medical needs (eg, allergologist [28], neurologists [29], or urologists [30].

Another striking finding of our study, however, was the significantly higher perception of telemedicine among health care workers within the stationary or hospital sector compared to the ambulatory or practice sector during the current crisis. This might be the result of many telemedical apps designed specifically for the hospital setting (eg, tele-intensive care) and less apps for the home care or ambulatory setting. As a major result of our study, this should lead to further development of telemedical solutions for the ambulatory sector to contain infections, reduce unnecessary practice visits, and consequently decrease infection risk.

Interestingly, over one-third of private practices already had the possibility to reduce patient contact through telemedicine. In many cases, practitioners used makeshift and highly accessible tools, like messenger services to communicate with patients in a home setting [31].

However, in particular to the hospital setting, regulatory and technical burdens seem to be hindering the progress. This is particularly true for the public sectors, where a majority of the study participants reported at least occasional obstacles for telemedicine within their work environments. Reducing these obstacles as well as investing in technical infrastructure to provide optimal care for the patients harmed by COVID-19 within the crisis should be a priority for regulatory bodies and governments. It is, however, important to underline that only approximately one-fifth of the survey participants described a severe regulatory or technical hindering to telemedicine (answering “yes”).

Public information in particular plays a crucial role in the implementation of adherence to public health guidelines and fact-based communication toward peers and patients. Although a majority of health care providers rated the communication quality and quantity positively, a subcohort analysis highlighted a significantly lower perception for nurses and other medical personnel. As medical nurses and other medical professionals play an essential role in patient care during the crisis, specially targeted information material is urgently needed to address this deficiency. For example, special COVID-19–related information platforms are already in place (eg, within the American Nurses Association [32]). To reach the recipient, it is also important to note that most medical professionals used websites and television to keep informed, while social media only served a minor role for medical professionals. Interestingly, the general acceptance of telemedicine was negatively correlated with age. This could clearly be an indication of a targeted information gap concerning older medical professionals. It is important to underline that regulatory and technical aspects are not the only burdens hindering the establishment of telemedicine in routine care. In particular financial aspects, such as a lack of additional investment in infrastructure, expansion of the clinical staff, and additional training for clinical and administrative staff, can potentially impact implementation. Furthermore, social aspects, such as language barriers in direct audio-video conversation and lack of technical skills, can impede the progress in this regard. Regulatory aspects, such as the need for supplementary documentation, should be addressed early on to avoid the creation of an additional workload. The aforementioned aspects have to be taken into account when creating new telemedical infrastructures.

Clearly, this study has some limitations. First, the study population was limited to German-speaking participants. Larger studies with international participants have to be conducted to confirm the results. Second, due to the emergent nature of this survey, the professional groups were not evenly distributed within the participants, leading to potential implications for certain groups. This can potentially be overcome by a larger study, in which other medical professional groups are specifically targeted. Third, arising from the nature of online surveys, there was potentially more attraction for technophile participants, resulting in a distorted picture concerning things such as technical hurdles.

The COVID-19 pandemic is creating a historic global challenge for health care providers, patients, and societies alike. When technological, regulatory, and infrastructural burdens can quickly be overcome, telemedicine has the chance to transform from model implementations to a global supply structure, potentially saving thousands of patients' lives.

Conflicts of Interest

GM is the president of the German Society of Telemedicine and an Honoria for 2 lectures from Philips. AP, GM, and LM are co-founders of Clinomic GmbH. AP is chief executive officer of Clinomic GmbH. All other authors declare that they have no conflict of interests.

Multimedia Appendix 1
Calculation of ideal sample size.
[DOCX File , 20 KB - jmir_v22i8e19745_app1.docx]

References

1. Kluge S, Janssens U, Welte T, Weber-Carstens S, Marx G, Karagiannidis C. [Recommendations for critically ill patients with COVID-19]. Med Klin Intensivmed Notfmed 2020 Apr;115(3):175-177 [FREE Full text] [doi: 10.1007/s00063-020-00674-3] [Medline: 32166350]
2. Breslow MJ, Rosenfeld BA, Doerfler M, Burke G, Yates G, Stone DJ, et al. Effect of a multiple-site intensive care unit telemedicine program on clinical and economic outcomes: An alternative paradigm for intensivist staffing*. Crit Care Med 2004;32(1):31-38. [doi: 10.1097/01.ccm.0000104204.61296.41]
3. Pronovost PJ, Waters H, Dorman T. Impact of critical care physician workforce for intensive care unit physician staffing. Curr Opin Crit Care 2001 Dec;7(6):456-459. [doi: 10.1097/00075198-200112000-00015] [Medline: 11805551]
4. Murphy DJ, Fan E, Needham DM. ICU staffing and patient outcomes: more work remains. Crit Care 2009;13(1):101 [FREE Full text] [doi: 10.1186/cc7113] [Medline: 19183427]
5. Treggiari MM, Martin DP, Yanez ND, Caldwell E, Hudson LD, Rubenfeld GD. Effect of intensive care unit organizational model and structure on outcomes in patients with acute lung injury. Am J Respir Crit Care Med 2007 Oct 01;176(7):685-690 [FREE Full text] [doi: 10.1164/rccm.200701-165OC] [Medline: 17556721]
6. Rosenfeld BA, Dorman T, Breslow MJ, Pronovost P, Jenckes M, Zhang N, et al. Intensive care unit telemedicine: alternate paradigm for providing continuous intensivist care. Crit Care Med 2000 Dec;28(12):3925-3931. [doi: 10.1097/00003246-200012000-00034] [Medline: 11153637]
7. Armaignac DL, Saxena A, Rubens M, Valle CA, Williams LS, Veledar E, et al. Impact of telemedicine on mortality, length of stay, and cost among patients in progressive care units: experience from a large healthcare system. Crit Care Med 2018 May;46(5):728-735 [FREE Full text] [doi: 10.1097/CCM.0000000000002994] [Medline: 29384782]
8. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]

9. Rademacher NJ, Cole G, Psoter KJ, Kelen G, Fan JWZ, Gordon D, et al. Use of telemedicine to screen patients in the emergency department: matched cohort study evaluating efficiency and patient safety of telemedicine. JMIR Med Inform 2019 May 08;7(2):e11233. [doi: 10.2196/11233] [Medline: 31066698]
10. Young LB, Chan PS, Lu X, Nallamothu BK, Sasson C, Cram PM. Impact of telemedicine intensive care unit coverage on patient outcomes: a systematic review and meta-analysis. Arch Intern Med 2011 Mar 28;171(6):498-506. [doi: 10.1001/archinternmed.2011.61] [Medline: 21444842]
11. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine during the COVID-19 pandemic: experiences from Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
12. Startschuss für das Virtuelle Krankenhaus. Das Landesportal Wir in NRW. 2020. URL: https://www.land.nrw/de/pressemitteilung/unikliniken-aachen-und-muenster-stellen-nordrhein-westfaelischen-krankenhaeusern [accessed 2020-04-23]
13. Deisz R, Rademacher S, Gilger K, Jegen R, Sauerzapfe B, Fitzner C, et al. Additional telemedicine rounds as a successful performance-improvement strategy for sepsis management: observational multicenter study. J Med Internet Res 2019 Jan 15;21(1):e11161 [FREE Full text] [doi: 10.2196/11161] [Medline: 30664476]
14. Marx G, Beckers R. [Telemedicine in Germany]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2015 Oct;58(10):1053-1055. [doi: 10.1007/s00103-015-2232-4] [Medline: 26293221]
15. Zhai Y, Wang Y, Zhang M, et al. From isolation to coordination: how can telemedicine help combat the COVID-19 outbreak? [preprint]. medRxiv 2020 Feb 23:1. [doi: 10.1101/2020.02.20.20025957]
16. Chen C, Jyan H, Chien S, Jen H, Hsu C, Lee P, et al. Containing COVID-19 among 627,386 persons in contact with the Diamond Princess cruise ship passengers who disembarked in Taiwan: big data analytics. J Med Internet Res 2020 May 05;22(5):e19540 [FREE Full text] [doi: 10.2196/19540] [Medline: 32353827]
17. Sterling R, LeRouge C. On-demand telemedicine as a disruptive health technology: qualitative study exploring emerging business models and strategies among early adopter organizations in the United States. J Med Internet Res 2019 Nov 15;21(11):e14304 [FREE Full text] [doi: 10.2196/14304] [Medline: 31730038]
18. Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA. Cost-effectiveness of telemedicine in remote orthopedic consultations: randomized controlled trial. J Med Internet Res 2019 Feb 19;21(2):e11330 [FREE Full text] [doi: 10.2196/11330] [Medline: 30777845]
19. Mileski M, Kruse CS, Catalani J, Haderer T. Adopting telemedicine for the self-management of hypertension: systematic review. JMIR Med Inform 2017 Oct 24;5(4):e41 [FREE Full text] [doi: 10.2196/medinform.6603] [Medline: 29066424]
20. COVID-GUIDE. URL: https://covidguide.health/de/ [accessed 2020-04-23]
21. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
22. The Jamovi Project. URL: https://www.jamovi.org/about.html [accessed 2020-04-20]
23. The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2019-10-25]
24. Anderson DR. Statistics for Business and Economics. England: Cengage Learning EMEA; 2010.
25. Gesundheitspersonal. Stat Bundesamt. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Gesundheitspersonal/_inhalt.html [accessed 2020-04-23]
26. Mullen-Fortino M, DiMartino J, Entrikin L, Mulliner S, Hanson CW, Kahn JM. Bedside nurses' perceptions of intensive care unit telemedicine. Am J Crit Care 2012 Jan;21(1):24-31. [doi: 10.4037/ajcc2012801] [Medline: 22210696]
27. Tudiver F, Wolff LT, Morin PC, Teresi J, Palmas W, Starren J, et al. Primary care providers' perceptions of home diabetes telemedicine care in the IDEATel Project. J Rural Health 2007 Jan;23(1):55-61. [doi: 10.1111/j.1748-0361.2006.00068.x]
28. Portnoy J, Waller M, Elliott T. Telemedicine in the era of COVID-19. J Allergy Clin Immunol Pract 2020 May;8(5):1489-1491 [FREE Full text] [doi: 10.1016/j.jaip.2020.03.008] [Medline: 32220575]
29. Klein BC, Busis NA. COVID-19 is catalyzing the adoption of teleneurology. Neurology 2020 May 26;94(21):903-904. [doi: 10.1212/WNL.0000000000009494] [Medline: 32238505]
30. Gadzinski AJ, Gore JL, Ellimoottil C, Odisho AY, Watts KL. Implementing telemedicine in response to the COVID-19 pandemic. J Urol 2020 Jul;204(1):14-16. [doi: 10.1097/ju.0000000000001033]
31. Maeder A, Ho K, Marcelo A. he Promise of New Technologies in an Age of New Health Challenges: Selected Papers from Global Telehealth. Amsterdam: IOS Press; 2016.
32. COVID-19 resource center. ANA Enterprise. URL: https://www.nursingworld.org/practice-policy/work-environment/health-safety/disaster-preparedness/coronavirus/ [accessed 2020-04-24]

Abbreviations

COVID-19: coronavirus disease
ICU: intensive care unit
MySQL: My Structured Query Language

Edited by G Eysenbach; submitted 11.05.20; peer-reviewed by N Yamada, T Aslanidis, JL Ditchburn; comments to author 02.06.20; revised version received 15.06.20; accepted 21.06.20; published 05.08.20.

Please cite as:
Peine A, Paffenholz P, Martin L, Dohmen S, Marx G, Loosen SH
Telemedicine in Germany During the COVID-19 Pandemic: Multi-Professional National Survey
J Med Internet Res 2020;22(8):e19745
URL: https://www.jmir.org/2020/8/e19745
doi:10.2196/19745
PMID:32568724

Original Paper

Feasibility of Real Time Internet-Based Teleconsultation in Patients With Multiple Sclerosis: Interventional Pilot Study

Miguel D'Haeseleer[1,2,3], MD, PhD; Piet Eelen[2], RNCS; Nima Sadeghi[3], BMBS; Marie B D'Hooghe[1,2,3], MD, PhD; Jeroen Van Schependom[1,2,3], MEng, PhD; Guy Nagels[1,2,3,4], MD, MEng, PhD

[1]Universitair Ziekenhuis Brussel, Brussels, Belgium

[2]Nationaal Multiple Sclerose Centrum, Melsbroek, Belgium

[3]Center for Neurosciences, Vrije Universiteit Brussel, Brussels, Belgium

[4]Zebra Academy, Brussels, Belgium

Corresponding Author:
Miguel D'Haeseleer, MD, PhD
Universitair Ziekenhuis Brussel
Laarbeeklaan 101
Brussels
Belgium
Phone: 32 2 477 60 12
Email: miguel.dhaeseleer@uzbrussel.be

Abstract

Background: Telemedicine (TM) is currently flourishing in rural and emergency settings, but its implementation in the routine management of chronic neurological disorders has developed with more hesitation. Limited access to specialized care facilities and expanding patient populations, combined with unprecedented mobility restrictions imposed by the coronavirus disease pandemic, are currently stressing the need for remote solutions in this field. Studies in patients with multiple sclerosis (MS) have been heterogeneous in objectives and methodology but generally support the concept that TM interventions produce clinical benefits, cost-effectiveness, and user satisfaction. Nonetheless, data on live interaction between patients and health care providers for MS teleconsultation purposes remain scarce.

Objective: The aim of this study is to demonstrate the feasibility of planned real time audiovisual teleconsultation over the internet for patients with MS.

Methods: A total of 20 patients with MS presenting at a specialized MS center in Belgium were recruited for this study. One teleconsultation was scheduled for each participant. Patients were provided a unique hyperlink by mail in advance, leading them automatically and directly to the virtual waiting room, where they could accept or decline our incoming call. All teleconsultations were performed by a trained medical student with the intention to keep the conversation similar to what is usually discussed during a classic face-to-face MS consultation; no remote physical exams were performed. The approach was considered feasible if at least 80% of the planned TM visits could be successfully completed at the foreseen moment. Patient satisfaction (technical quality, convenience, and overall quality of care) was evaluated at the end of each teleconsultation by means of 5-point Likert scales containing the categories very unsatisfied, unsatisfied, neutral, satisfied, and highly satisfied.

Results: Out of 20 consultations, 17 were successfully completed (85%). Failures were due to patients not responding (n=2) and technical issues (n=1). Out of the 17 consultations, 17 patients declared themselves satisfied or highly satisfied for technical quality, 15 patients for convenience, and 16 patients for overall quality of care.

Conclusions: Planned real time audiovisual teleconsultation over the internet is feasible and highly appreciated in patients with MS. Incorporation of such services in routine clinical MS practice is expected to improve access to specialized care facilities for affected patients.

(J Med Internet Res 2020;22(8):e18178) doi:10.2196/18178

KEYWORDS

multiple sclerosis; teleconsultation; internet; feasibility; eHealth

Introduction

Telemedicine (TM) is defined as the exchange of medical information between patients and health care providers in distinct locations using electronic communication technology [1]. Its primary aim is to improve public health by allowing medical services that cannot be easily established in a face-to-face manner. TM can, from a practical perspective, essentially be broken down into three major categories: real time interactive sessions (eg, telephone conversations, videoconferences), store-and-forward technology (eg, email), and remote monitoring (eg, self-assessment devices) [2]. Most efforts so far have been devoted to the implementation of TM in rural and emergency medicine, as illustrated by the successful example of telestroke [3], while common chronic neurological disorders seem to have adopted digital health technology with more hesitation.

Nervous system diseases are the leading cause of disability around the world, and their burden is expected to double over the next 25 years, an evolution mainly driven by expansion of the aging population [4]. Current access to neurological facilities is limited already, and classic care models will likely not be able to match a future demand of that magnitude [5]. Several high-impact papers have recently highlighted the potential of TM to help close this gap and to prove the benefits for millions of people with chronic conditions such as epilepsy, headache syndromes, dementia, movement disorders, and multiple sclerosis (MS) [5-7]. Accelerated and wide-scale incorporation of digital health services is generally, and more specifically in the field of neurology, expected and even recommended, viewpoints that have recently been sharpened by the coronavirus disease [8] outbreak, where social isolation has been deployed as the primary measure of constraint [9-12].

Individuals with MS typically receive their diagnosis during young adulthood [13], when busy social and professional schedules risk to limit the time available for medical attention. Strikingly, a high percentage of them appear to be interested in online interaction with health care providers and peers [14], consistent with the infodemiological finding (infodemiology refers to a newly described area of epidemiological research that deals with the question of how health information is accessed on the internet [15]) that an ever-increasing part of the general population exploits the internet to obtain health information [16]. Studies in patients with MS have applied different types of TM and were conducted for various purposes (eg, management, disability assessment, treatment, rehabilitation). A comprehensive and structured review of the literature has recently been published by Yeroushalmi and colleagues [17], in which the authors conclude that most randomized controlled trials have demonstrated either no difference between a TM versus an in-person intervention or an association of the former with a more beneficial outcome. Furthermore, the majority of these studies were considered low to medium cost. Another recent overview has primarily focused on the potential of digital communication tools to integrate real-world patient-centered data into clinical MS registries, which could then be easily shared between caregivers and lead, in conjunction with machine learning and artificial intelligence, to improved disease understanding, decision support, and self-empowerment [18]. Nonetheless, data on live interaction between patients and health care providers in a teleconsultation setting are relatively scarce. The main objective of our study is to explore whether planned real time audiovisual teleconsultation over the internet is feasible in patients with MS.

Methods

Ethics

This study was approved by the ethics committee of the Nationaal MS Center in Melsbroek (local; internal reference: AvN/AVDZ) and the Universitair Ziekenhuis Brussel (leading; internal reference: 2018/269 - Belgian Unique Number: 143201836797). Written informed consent was obtained from all participants prior to inclusion.

Patient Cohort

A total of 20 French- or Dutch-speaking patients with MS, according to the 2017 revised McDonald criteria [19], were recruited at the Nationaal MS Center in Melsbroek. Home access to the internet with a webcam-equipped device was required for inclusion. Recruitment of individuals with a high suspicion of moderate to severe cognitive impairment (based on common sense judgement of the medical record or initial patient contact) was actively avoided. A known history of cognitive dysfunction (defined as scoring less than 21 on the Mini Mental State Evaluation, if present in the medical record; no cognitive evaluations were performed within this study) accounted as a formal exclusion criterion. Age, sex, clinical subtype (ie, relapsing-remitting versus secondary or primary progressive MS), and disability (the latter represented by Expanded Disability Status Scale [EDSS] scores, with higher scores representing a more pronounced degree of functional impairment [20]) were extracted as demographic variables from the local existing patient records. EDSS scores are routinely assessed at nearly every visit during regular neurological follow-ups in the Melsbroek center, and the most recent available result was consistently selected in each participant. The same accounted for the determination of the clinical MS subtype.

Teleconsultations

One teleconsultation was scheduled for every participant, in addition to standard neurological follow-ups, and performed by a trained medical student (NS) under the supervision of the principle investigator of the study (MD), using a novel internet-based communication platform obtained from Zebra Academy. This technology was originally developed at the Vrije Universiteit Brussel with the purpose of supporting prehospital management of acute stroke [21]. A checklist (with questions about items such as general and neurological health, medication, and lifestyle factors; see Textbox 1) similar to the usual content of a classic face-to-face MS consultation was used as a backbone to guide the conversation, but deviations at the initiative of the patient were allowed. Patients were provided a unique hyperlink by mail in advance, leading them directly to the virtual waiting room where they could notice and accept our incoming call at the time of the scheduled appointment (Figure 1). Access was possible from any device with a webcam (ie, laptop, desktop,

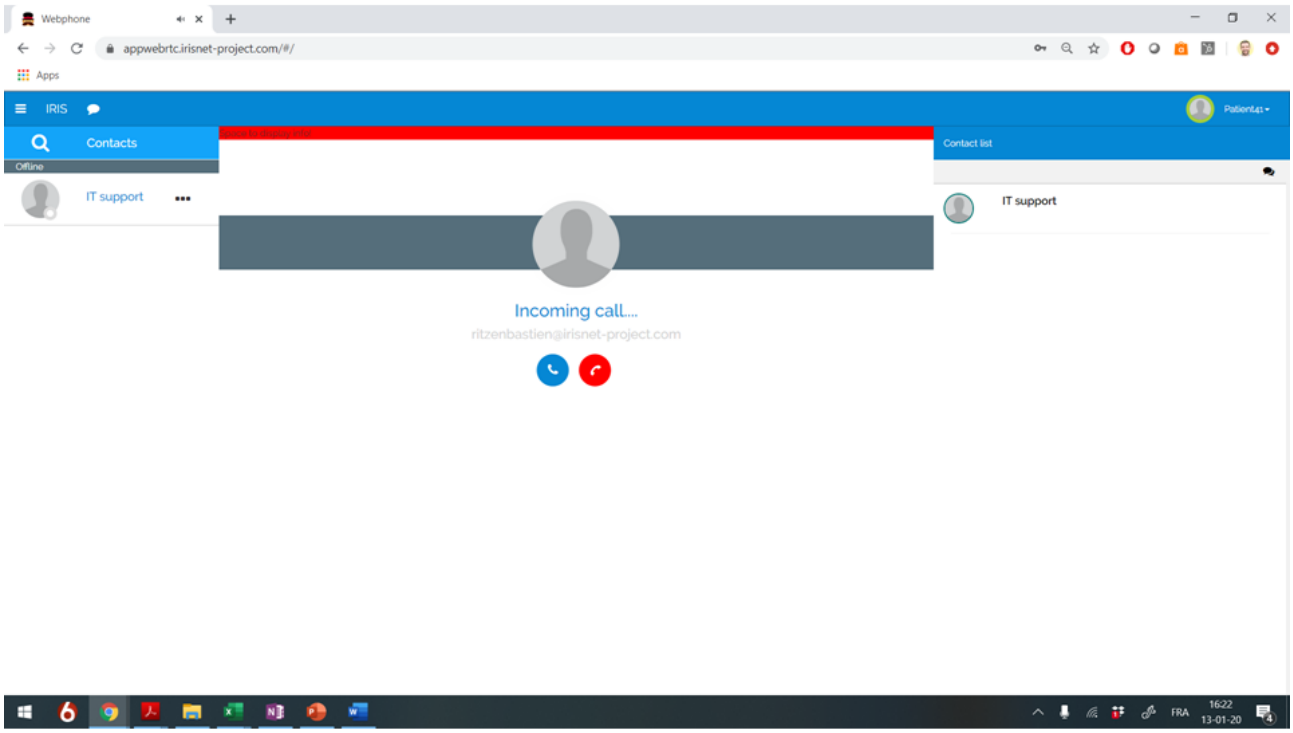
tablet, smartphone). Google Chrome was used as a web browser on both sides of the connection, as advised by Zebra Academy. All patients had to respond to the call within the next 30 minutes following the scheduled time, with a maximum of three attempts during this window. A report was written and forwarded to the treating neurologist after each visit to allow additional medical interventions outside of the scope and protocol of this study, if deemed necessary based on the content of the teleconsultation. The act of teleconsultation was considered feasible if at least 80% of the planned visits could be successfully completed.

Textbox 1. Checklist used to guide the teleconsultation.

Checklist

- How would you describe your current general health status?
- Did you experience any relapses or neurological deterioration over the past 3 months?
- Did you experience any other medical problems (not related to multiple sclerosis [MS]) over this time period?
- Do you have any of the following MS-related symptoms: muscle weakness, sensory loss, visual disturbances, swallowing or speech difficulties, bladder or bowl dysfunction, pain or spasticity (muscle spasms), fatigue, mood swings, mental slowness?
- Which of these symptoms have recently progressed or currently have a significant impact on your quality of life?
- What medication do you take?
- Do you experience any side effects or other inconvenience related to your medical treatment?
- Do you follow a regular rehabilitation scheme? If so, please estimate the average time of weekly physical activity.
- Would you describe yourself as compliant to your medical and rehabilitation treatment? If not, please explain why.
- Do you smoke or use recreational drugs?
- How would you evaluate your professional performance? How many days of work did you miss over the past 3 months?
- How would you evaluate your social and family member interactions?

Figure 1. Zebra Academy teleconsultation platform. The screenshot is from the perspective of a fictional patient receiving our incoming internet call. Clicking on the blue icon will establish the connection. Bastien Ritzen is a member of the Zebra Academy crew (for illustrative purposes). IT: information technology.



Satisfaction

Patient satisfaction was evaluated at the end of each teleconsultation by means of 5-point Likert scales containing the categories very unsatisfied, unsatisfied, neutral, satisfied, and highly satisfied. Assessments were independently carried out for technical quality, convenience, and overall quality of care (QoC).

Data Availability

Anonymized data will be shared by request from any qualified investigator.

Results

Patient Cohort

The median age and EDSS scores of the participants (as extracted from the Melsbroek patient database) at the time of inclusion were 41 (range 27-62) years of age and 4.0 (range 0.0-6.5), respectively. The female to male ratio was 11:9. A total of 14 participants had relapsing-remitting MS, while the other 6 had a progressive disease course (4/20 secondary and 2/20 primary progressive MS).

Teleconsultations

A total of 17 (85%) out of 20 planned teleconsultations were successfully completed. Failures were due to participants not responding (n=2) and technical issues (n=1). The nonresponders were contacted at a later time by telephone and both let us know that they had forgotten the appointment. The technical issue was a blank video screen from the patient's perspective, appearing after responding to each of the three allowed attempts to connect.

Satisfaction

Out of the total 17 consultations, 17 patients declared themselves to be satisfied or highly satisfied with the teleconsultation for technical quality, 15 patients for convenience, and 16 patients for overall QoC.

Discussion

We present the first study demonstrating feasibility of planned real time audiovisual teleconsultation over the internet in patients with MS, in which feasibility was a priori defined as the ability to complete a fixed number of scheduled TM visits at the foreseen moment. In addition, patient appreciation regarding the technical aspects and general approach was excellent. Our results are in line with a recent comparative crossover trial from Robb and colleagues [22], in which there was less than 15% difference between successfully performed internet-based video house calls and personal in-hospital visits (prespecified study target) after each participant agreed to receive both consultation modalities consecutively. Eventually 25 (67.6%) out of the 37 scheduled TM visits could be completed according to that protocol. Reasons for failure were not mentioned. The vast majority of participants reported that they would recommend teleconsultation to others (97.1%) and stated that establishing the virtual connection was easy (94.3%) [22]. To the best of our knowledge, there are no other studies exploiting this particular type of TM intervention. Previous efforts at connecting patients with health care providers, in the context of general MS management, were based on online texting, telephone hotlines, or noninternet videoconferencing outside the clinic hours [23-29]. Respective methodologies and objectives were heterogeneous but most of these studies had in common that telehealth services provided patient well-being and user satisfaction [17].

Our work was, to a certain extent, inspired by a previous randomized controlled trial demonstrating the feasibility of clinical monitoring over a 12-month period via video house calls in patients with Parkinson disease living throughout the United States. QoC perception was not significantly influenced, but the intervention did lead to a high degree of patient satisfaction and time gain [30]. Several reasons can be given to assume that patients with MS are at least as equally good candidates to benefit from such an approach. As already mentioned in the introduction, these individuals seem to have a strong interest in online communication, and the disease typically affects young adults who might have other priorities than seeking pathology-specific medical attention. In addition, MS commonly leads to cumulative and substantial physical disability [13], which can create additional logistic boundaries even in areas highly saturated with neurologists. Periodic assessment is crucial to monitor disease activity, treatment response, and health-related quality of life, and inevitably requires frequent visits to the neurology office. One study reported that nearly 30% of patients with MS do not receive neurological care at all, which undeniably decreases the likelihood of access to appropriate disease-modifying treatment and specialized facilities [31]. Over the past decades, it has become increasingly clear that absence or delayed start of an immunological maintenance therapy is a risk factor for a worse prognosis [32]. Interestingly, digital health services seem to have been positively welcomed by MS neurologists as well [33], for whom they can serve as a medium for education and case discussion [34,35].

Demonstrating feasibility is an early but essential hurdle toward the implementation of technological development in medicine. Various TM interventions have already been deployed to support medical management, disability assessment, treatment, and physical or cognitive rehabilitation in individuals with MS [17]. In contrast, real time audiovisual communication over the internet between health care providers and patients, aimed at routine clinical follow-up, is novel in this field. One of the most attractive features of the system applied in our study is its user-friendliness by simplicity, as it takes only two clicks from patients to participate in the call. On the other hand, we have to acknowledge that drawing conclusions from this pilot is not free from potential pitfalls. First, the total number of participants and scheduled TM visits was small. One should be particularly careful extrapolating our findings to the full community of individuals with MS, as results might be less positive when considering populations with more severe clinical disability. Previous work has revealed that patients with cognitive or visual impairment experienced more difficulties while using home-based TM systems [23,25], whereas we have actively avoided recruitment of patients with apparent cognitive dysfunction. Second, teleconsultations were performed by a medical student and did not include a physical exam. Uncertainty about the ability to conduct an accurate and detailed remote clinical neurological evaluation remains a general weakness of TM. However, it is worth mentioning that Bove and colleagues [36] recently reported agreement within 1 point between in-person and televideo-enabled EDSS scores for 88% of the cases, which is on par with the in-person interrater variability described by others. Third, feasibility was considered to be a single binary outcome measure ("either the teleconsultation could be successfully completed or not"), while other potentially interesting variables such as time- and

cost-effectiveness have not been taken into account. Satisfaction was assessed for exploratory reasons with Likert scales, which are easy to use but lack validation for these particular trials. There are recent data suggesting that such scales may be more vulnerable to bias from confounding factors and that a ceiling effect may be more difficult to avoid when compared with a visual analogue scale [37]. Fourth, our study had a cross-sectional design and was not able to inform us on patient compliance or technical reliability over time. Longitudinal data regarding real time interactive patient monitoring are scarce, but an Israeli study suggested that regular videoconferencing over the telephone line over 6 months improves clinical outcomes while reducing medical expenses [29]. In a future research project, we will investigate the feasibility and potential of patient follow-ups with the Zebra communication platform over 1 year. Positive results may open new ways to explore TM technology for multiple purposes in MS (eg, validation of clinical rating scales, noninferiority trial designs in relapse evaluation and routine follow-up, remote virtual participation of neurologist in multidisciplinary consultation, and therapeutic group sessions such as mindfulness). It would also be of interest to explore the utility of TM for diagnostic counselling and to do comparative studies across distinct MS populations, as there might be variability in results due to geographical, cultural, or social differences. However, even with today's knowledge, and if legal and reimbursement policies allow, incorporating TM in routine MS practice is likely to reduce barriers between affected patients and specialized care facilities, potentially leading to improved clinical prognosis and health-related quality of life in a cost- and time-efficient manner.

Acknowledgments

The authors thank Zebra Academy for offering their technology free of charge for this study. We would also like to express our gratitude toward Thomas Van Oosthuysen (Information and Communications Technology Department of the Nationaal MS Center in Melsbroek) and the Zebra Academy crew for providing technical support.

Authors' Contributions

MD and GN conceptualized the study. MD wrote the first draft of the paper, and all authors were involved in the critical reading and revision process.

Conflicts of Interest

This study was supported by a noncompetitive research grant from Roche (Basel, Switzerland). GN is a shareholder of Zebra Academy. The remaining authors declare no conflicts of interest.

References

1. Sola-Valls N, Blanco Y, Sepúlveda M, Martinez-Hernandez E, Saiz A. Telemedicine for monitoring MS activity and progression. Curr Treat Options Neurol 2015 Nov;17(11):47. [doi: 10.1007/s11940-015-0377-y] [Medline: 26423914]
2. Rubin MN, Wellik KE, Channer DD, Demaerschalk BM. Systematic review of teleneurology: neurohospitalist neurology. Neurohospitalist 2013 Jul;3(3):120-124 [FREE Full text] [doi: 10.1177/1941874413483754] [Medline: 24167644]
3. Hess DC, Audebert HJ. The history and future of telestroke. Nat Rev Neurol 2013 Jun;9(6):340-350. [doi: 10.1038/nrneurol.2013.86] [Medline: 23649102]
4. Dorsey ER, Constantinescu R, Thompson JP, Biglan KM, Holloway RG, Kieburtz K, et al. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. Neurology 2007 Jan 30;68(5):384-386. [doi: 10.1212/01.wnl.0000247740.47667.03] [Medline: 17082464]
5. Dorsey ER, Glidden AM, Holloway MR, Birbeck GL, Schwamm LH. Teleneurology and mobile technologies: the future of neurological care. Nat Rev Neurol 2018 May;14(5):285-297. [doi: 10.1038/nrneurol.2018.31] [Medline: 29623949]
6. Hatcher-Martin JM, Adams JL, Anderson ER, Bove R, Burrus TM, Chehrenama M, et al. Telemedicine in neurology. Neurology 2019 Dec 04;94(1):30-38. [doi: 10.1212/wnl.0000000000008708]
7. Guzik AK, Switzer JA. Teleneurology is neurology. Neurology 2019 Dec 04;94(1):16-17. [doi: 10.1212/wnl.0000000000008693]
8. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020 May;109:102433 [FREE Full text] [doi: 10.1016/j.jaut.2020.102433] [Medline: 32113704]
9. Klein BC, Busis NA. COVID-19 is catalyzing the adoption of teleneurology. Neurology 2020 Apr 01;94(21):903-904. [doi: 10.1212/wnl.0000000000009494]
10. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
11. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/nejmp2005835]
12. Mahmood S, Hasan K, Colder Carras M, Labrique A. Global preparedness against COVID-19: we must leverage the power of digital health. JMIR Public Health Surveill 2020 Apr 16;6(2):e18980 [FREE Full text] [doi: 10.2196/18980] [Medline: 32297868]

13. Reich DS, Lucchinetti CF, Calabresi PA. Multiple sclerosis. N Engl J Med 2018 Jan 11;378(2):169-180. [doi: 10.1056/nejmra1401483]
14. Lejbkowicz I, Paperna T, Stein N, Dishon S, Miller A. Internet usage by patients with multiple sclerosis: implications to participatory medicine and personalized healthcare. Mult Scler Int 2010;2010:640749. [doi: 10.1155/2010/640749] [Medline: 22096625]
15. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
16. Lavorgna L, Brigo F, Moccia M, Leocani L, Lanzillo R, Clerico M, et al. e-Health and multiple sclerosis: an update. Mult Scler 2018 Nov;24(13):1657-1664. [doi: 10.1177/1352458518799629] [Medline: 30231004]
17. Yeroushalmi S, Maloni H, Costello K, Wallin MT. Telemedicine and multiple sclerosis: a comprehensive literature review. J Telemed Telecare 2019 May 01:1357633X1984009. [doi: 10.1177/1357633x19840097]
18. Matthews PM, Block VJ, Leocani L. E-health and multiple sclerosis. Curr Opin Neurol 2020;33(3):271-276. [doi: 10.1097/wco.0000000000000823]
19. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 2018 Feb;17(2):162-173. [doi: 10.1016/s1474-4422(17)30470-2]
20. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983 Nov;33(11):1444-1452. [doi: 10.1212/wnl.33.11.1444] [Medline: 6685237]
21. Valenzuela Espinoza A, Van Hooff R, De Smedt A, Moens M, Yperzeele L, Nieboer K, PreSSUB-Consortium. Development and pilot testing of 24/7 in-ambulance telemedicine for acute stroke: prehospital stroke study at the Universitair Ziekenhuis Brussel-Project. Cerebrovasc Dis 2016;42(1-2):15-22 [FREE Full text] [doi: 10.1159/000444175] [Medline: 26950076]
22. Robb JF, Hyland MH, Goodman AD. Comparison of telemedicine versus in-person visits for persons with multiple sclerosis: a randomized crossover study of feasibility, cost, and satisfaction. Mult Scler Relat Disord 2019 Nov;36:101258. [doi: 10.1016/j.msard.2019.05.001]
23. Atreja A, Mehta N, Miller D, Moore S, Nichols K, Miller H, et al. One size does not fit all: using qualitative methods to inform the development of an Internet portal for multiple sclerosis patients. AMIA Annu Symp Proc 2005:16-20 [FREE Full text] [Medline: 16778993]
24. Miller DM, Moore SM, Fox RJ, Atreja A, Fu AZ, Lee J, et al. Web-based self-management for patients with multiple sclerosis: a practical, randomized trial. Telemed J E Health 2011;17(1):5-13 [FREE Full text] [doi: 10.1089/tmj.2010.0133] [Medline: 21214498]
25. Settle JR, Maloni HW, Bedra M, Finkelstein J, Zhan M, Wallin MT. Monitoring medication adherence in multiple sclerosis using a novel web-based tool: a pilot study. J Telemed Telecare 2015 Aug 06;22(4):225-233. [doi: 10.1177/1357633x15597115]
26. Mercier HW, Ni P, Houlihan BV, Jette AM. Differential impact and use of a telehealth intervention by persons with MS or SCI. Am J Phys Med Rehabil 2015;94(11):987-999. [doi: 10.1097/phm.0000000000000291]
27. Strupp J, Groebe B, Knies A, Mai M, Voltz R, Golla H. Evaluation of a palliative and hospice care telephone hotline for patients severely affected by multiple sclerosis and their caregivers. Eur J Neurol 2017 Dec;24(12):1518-1524. [doi: 10.1111/ene.13462] [Medline: 28926157]
28. Ehde DM, Arewasikporn A, Alschuler KN, Hughes AJ, Turner AP. Moderators of treatment outcomes after telehealth self-management and education in adults with multiple sclerosis: a secondary analysis of a randomized controlled trial. Arch Phys Med Rehabil 2018 Jul;99(7):1265-1272. [doi: 10.1016/j.apmr.2017.12.012] [Medline: 29337024]
29. Zissman K, Lejbkowicz I, Miller A. Telemedicine for multiple sclerosis patients: assessment using Health Value Compass. Mult Scler 2012 Apr;18(4):472-480. [doi: 10.1177/1352458511421918] [Medline: 21965420]
30. Beck CA, Beran DB, Biglan KM, Boyd CM, Dorsey ER, Schmidt PN, et al. National randomized controlled trial of virtual house calls for Parkinson disease. Neurology 2017 Aug 16;89(11):1152-1161. [doi: 10.1212/wnl.0000000000004357]
31. Minden SL, Hoaglin DC, Hadden L, Frankel D, Robbins T, Perloff J. Access to and utilization of neurologists by people with multiple sclerosis. Neurology 2008 Mar 25;70(13 Pt 2):1141-1149. [doi: 10.1212/01.wnl.0000306411.46934.ef] [Medline: 18362274]
32. Gasperini C, Prosperini L, Tintoré M, Sormani MP, Filippi M, Rio J, et al. Unraveling treatment response in multiple sclerosis. Neurology 2018 Dec 26;92(4):180-192. [doi: 10.1212/wnl.0000000000006810]
33. Kern R, Haase R, Eisele JC, Thomas K, Ziemssen T. Designing an electronic patient management system for multiple sclerosis: building a next generation multiple sclerosis documentation system. Interact J Med Res 2016 Jan 08;5(1):e2 [FREE Full text] [doi: 10.2196/ijmr.4549] [Medline: 26746977]
34. Johnson K, Hertz D, Stobbe G, Alschuler K, Kalb R, Alexander K, et al. Project extension for community healthcare outcomes (ECHO) in multiple sclerosis: increasing clinician capacity. Int J MS Care 2017;19(6):283-289 [FREE Full text] [doi: 10.7224/1537-2073.2016-099] [Medline: 29270085]
35. Turner AP, Wallin MT, Sloan A, Maloni H, Kane R, Martz L, et al. Clinical management of multiple sclerosis through home telehealth monitoring: results of a pilot project. Int J MS Care 2013;15(1):8-14 [FREE Full text] [doi: 10.7224/1537-2073.2012-012] [Medline: 24453757]

36. Bove R, Bevan C, Crabtree E, Zhao C, Gomez R, Garcha P, et al. Toward a low-cost, in-home, telemedicine-enabled assessment of disability in multiple sclerosis. Mult Scler 2019 Oct;25(11):1526-1534. [doi: 10.1177/1352458518793527] [Medline: 30141729]
37. Voutilainen A, Pitkäaho T, Kvist T, Vehviläinen-Julkunen K. How to ask about patient satisfaction? The visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. J Adv Nurs 2016 Apr;72(4):946-957. [doi: 10.1111/jan.12875] [Medline: 26689434]

Abbreviations

EDSS: Expanded Disability Status Scale
MS: multiple sclerosis
QoC: quality of care
TM: telemedicine

Edited by G Eysenbach; submitted 09.02.20; peer-reviewed by A Hidki, E Neter, M Lozano-Lozano; comments to author 06.04.20; revised version received 13.05.20; accepted 20.05.20; published 13.08.20.

Please cite as:
D'Haeseleer M, Eelen P, Sadeghi N, D'Hooghe MB, Van Schependom J, Nagels G
Feasibility of Real Time Internet-Based Teleconsultation in Patients With Multiple Sclerosis: Interventional Pilot Study
J Med Internet Res 2020;22(8):e18178
URL: https://www.jmir.org/2020/8/e18178
doi:10.2196/18178
PMID:32447274

Original Paper

Glycemic Outcomes in Adults With Type 2 Diabetes Participating in a Continuous Glucose Monitor–Driven Virtual Diabetes Clinic: Prospective Trial

Amit R Majithia[1], MD; Coco M Kusiak[2], MS; Amy Armento Lee[2], MS; Francis R Colangelo[3], MD; Robert J Romanelli[4], PhD; Scott Robertson[2], RN; David P Miller[2], MS; David M Erani[5], MD; Jennifer E Layne[6], PhD; Ronald F Dixon[6], MD; Howard Zisser[2], MD

[1]Department of Medicine and Department of Pediatrics, University of California San Diego School of Medicine, La Jolla, CA, United States
[2]Verily Life Sciences, South San Francisco, CA, United States
[3]Allegheny Health Network, Monroeville, PA, United States
[4]Sutter Health, Palo Alto, CA, United States
[5]Onduo Professionals, PC, Newton, MA, United States
[6]Onduo, LLC, Newton, MA, United States

Corresponding Author:
Amit R Majithia, MD
Department of Medicine and Department of Pediatrics
University of California San Diego School of Medicine
Biomedical Research Facility II (3A19)
9500 Gilman Drive
La Jolla, CA, 92093
United States
Phone: 1 858 822 0727
Email: amajithia@ucsd.edu

Abstract

Background: The Onduo virtual diabetes clinic (VDC) for people with type 2 diabetes (T2D) combines a mobile app, remote personalized lifestyle coaching, connected devices, and live video consultations with board-certified endocrinologists for medication management and prescription of real-time continuous glucose monitoring (RT-CGM) devices for intermittent use.

Objective: This prospective single-arm study evaluated glycemic outcomes associated with participation in the Onduo VDC for 4 months.

Methods: Adults aged ≥18 years with T2D and a baseline glycated hemoglobin (HbA1c) of ≥8% to ≤12% were enrolled from 2 primary care centers from February 2019 to October 2019. Participants were asked to engage at ≥1 time per week with their care team and to participate in a telemedicine consultation with a clinic endocrinologist for diabetes medication review. Participants were asked to use a RT-CGM device and wear six 10-day sensors (total 60 days of sensor wear) intermittently over the course of 4 months. The primary outcome was change in HbA1c at 4 months from baseline. Other endpoints included change in weight and in RT-CGM glycemic metrics, including percent time <70, 70-180, 181-250, and >250 mg/dL. Changes in blood pressure and serum lipids at 4 months were also evaluated.

Results: Participants (n=55) were 57.3 (SD 11.6) years of age, body mass index 33.7 (SD 7.2), and 40% (22/55) female. HbA1c decreased significantly by 1.6% (SD 1%; *P*<.001). When stratified by baseline HbA1c of 8.0% to 9.0% (n=36) and >9.0% (n=19), HbA1c decreased by 1.2% (SD 0.6%; *P*<.001) and 2.4% (SD 1.3%; *P*<.001), respectively. Continuous glucose monitoring–measured (n=43) percent time in range (TIR) 70-180 mg/dL increased by 10.2% (SD 20.5%; *P*=.002), from 65.4% (SD 23.2%) to 75.5% (SD 22.7%), which was equivalent to a mean increase of 2.4 hours TIR per day. Percent time 181-250 mg/dL and >250 mg/dL decreased by 7.2% (SD 15.4; *P*=.005) and 3.0% (SD 9.4; *P*=.01), respectively. There was no change in percent time <70 mg/dL. Mean weight decreased by 9.0 lb (SD 10.4; *P*<.001). Significant improvements were also observed in systolic blood pressure, total cholesterol, low-density lipoprotein cholesterol, and triglycerides (*P*=.04 to P=<.001).

Conclusions: Participants in the Onduo VDC experienced significant improvement in HbA1c, increased TIR, decreased time in hyperglycemia, and no increase in hypoglycemia at 4 months. Improvements in other metabolic health parameters including

weight and blood pressure were also observed. In conclusion, the Onduo VDC has potential to support people with T2D and their clinicians between office visits by increasing access to specialty care and advanced diabetes technology including RT-CGM.

Trial Registration: ClinicalTrials.gov NCT03865381; https://clinicaltrials.gov/ct2/show/NCT03865381

(J Med Internet Res 2020;22(8):e21778) doi:10.2196/21778

KEYWORDS

continuous glucose monitoring; telemedicine; telehealth; digital health; type 2 diabetes; HbA1c

Introduction

In recent years, there has been a dramatic increase in telehealth programs for the management of diabetes [1]. These programs have the potential to support about 34 million individuals with diabetes in the United States and may play a larger role in future diabetes care, given the expected increase in the incidence and prevalence of diabetes in the next decade and beyond [2]. Telehealth programs for diabetes typically include a smartphone app, connected devices such as blood glucose meters, and remote coaching that may be automated or provided by a live health coach. Current telehealth programs, however, do not address the limitations of the traditional health care model of diabetes management: limited access to endocrinologists, specialist education, and advanced diabetes management technology.

There is growing recognition that advanced technology, including continuous glucose monitoring (CGM) devices, can play an essential role in diabetes care for people with type 2 diabetes (T2D), regardless of their treatment regimen [3-7]. The Onduo Virtual Diabetes Clinic (VDC), a telehealth model for people with T2D, is unique in incorporating CGM in its care model. Availability of live video consultations with board-certified endocrinologists for medication management and the ability to remotely prescribe CGM devices are also unique components of the VDC. Real-world evidence suggests that participation in the VDC is associated with significant improvement in HbA_{1c} [8] and a significant reduction in diabetes distress [9]. Here we report outcomes of a prospective single-group assignment trial, examining changes in HbA_{1c} in adults with T2D after 4 months of participation in the Onduo VDC.

Methods

Study Objective

The primary objective of this prospective single-arm study was to evaluate the change in HbA_{1c} in adults with T2D after 4 months of participation in the Onduo VDC. Additional outcomes included change in glycemic metrics from CGM (mean glucose, coefficient of variation, and percent time <70 mg/dL, 70 to 180 mg/dL [time in range (TIR)], 181 to 250 mg/dL, and >250 mg/dL at 4 months from baseline.Changes in weight, blood pressure (BP), and serum lipids—namely, total cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, low-density lipoprotein (LDL) cholesterol, cholesterol/HDL ratio, and non-HDL cholesterol—were also evaluated at 4 months from baseline.

Participants

Participants were enrolled from 2 primary care networks—Allegheny Health Network, Pittsburgh, PA and Sutter Health Palo Alto Medical Foundation, Palo Alto, CA. Inclusion criteria were as follows: ≥18 years of age or older, confirmed diagnosis of T2D, HbA_{1c} level 8.0% and 12.0%, willingness to use a blood glucose meter and CGM device, and own a smartphone. Exclusion criteria were as follows: use of an insulin pump; pregnant or breastfeeding; malignant cancer in the previous 12 months; any solid organ transplant; end-stage (stage 4 or 5) renal disease or dialysis; liver failure; cystic fibrosis; chronic heart failure (Class C, D); diabetes-related pancreatic failure; current use of a blood thinner; and self-reported adhesive allergy. All participants provided written informed consent. The study protocol and consent forms were approved by the Western Institutional Review Board and registered with ClinicalTrials.gov NCT03865381.

Protocol

Baseline and final assessments were conducted in-person at the designated study sites including, physical measures, blood draws, and questionnaires. The intervention was conducted remotely through the Onduo VDC.

Virtual Diabetes Clinic Participation

The Onduo VDC telehealth program for people with T2D has been previously described [8,9]. In brief, the program combines mobile app technology, remote personalized lifestyle coaching from Certified Diabetes Care and Education Specialists (CDCES) and health coaches, and connected blood glucose meters and real-time continuous glucose monitoring (RT-CGM) devices. Live video consultations with board-certified endocrinologists are available as needed for medication management in addition to prescribing CGM. Participants interact with their care team and are sent educational materials by messaging through the app. Participants use the app to track data relevant to diabetes care (such as, blood glucose readings and CGM data) and to log medication use, physical activity, meal photos, and other information.

In this study, participants were asked to engage ≥1 time per week with their health coach or care team and to participate in a telemedicine consultation with VDC endocrinologists. All participants were mailed a RT-CGM device—Dexcom G6 (Dexcom)—for intermittent use. Participants were asked to wear six 10-day sensors (total 60 days of sensor wear) intermittently over the course of 4 months. Initial period of CGM device–wearing lasted for 20 days (consisting of 2 sensors worn back to back). Subsequently, the remaining 4 sensors were deployed in a 10-days “on” and 11-days “off” cycle. Sensor

glucose data were used by the care team for coaching and monitoring and as an educational feedback loop to assist participants in associating their glucose levels with their diet, lifestyle, and other factors to optimize diabetes self-management. Glucose data were also used by the VDC endocrinologists for medication management. During the 4-month period, additional CGM sensor wear may have been requested by VDC endocrinologists to evaluate the efficacy of medication changes and monitor impact on blood glucose.

Statistical Analysis

A sample size of 60 participants was selected to provide 90% power to detect a 0.5% decrease in HbA_{1c} at 4 months from baseline (primary outcome) using a 2-sided test with α=.05, after assuming 20% loss to follow-up. RT-CGM glycemic metrics, including mean glucose, coefficient of variation, and percent time <70 mg/dL, 70 to 180 mg/dL, 181 to 250 mg/dL, and >250 mg/dL) were calculated from an initial 10-day period within 30 days of enrollment to a 10-day follow-up period >90 days from enrollment, with a data sufficiency requirement of >70% of possible readings. All outcomes were evaluated by paired *t* test except for ranges <70 mg/dL and >250 mg/dL, which were evaluated by Wilcoxon signed rank test. Nominal significance levels (*P* values) are presented with statistical significance defined as $P<.05$. All statistical analyses were performed using Python 3.6.7.

Results

A total of 60 participants enrolled in the study, and 92% (55/60) completed the 4-month intervention. Reasons for withdrawal included protocol violations/non-compliance (n=3) and withdrawal of consent (n=2). Out of 55 participants who completed the study, 89% (49/55) had a medication change. Baseline demographic and clinical characteristics of the participants are presented in Table 1.

Table 1. Participant demographics at baseline.

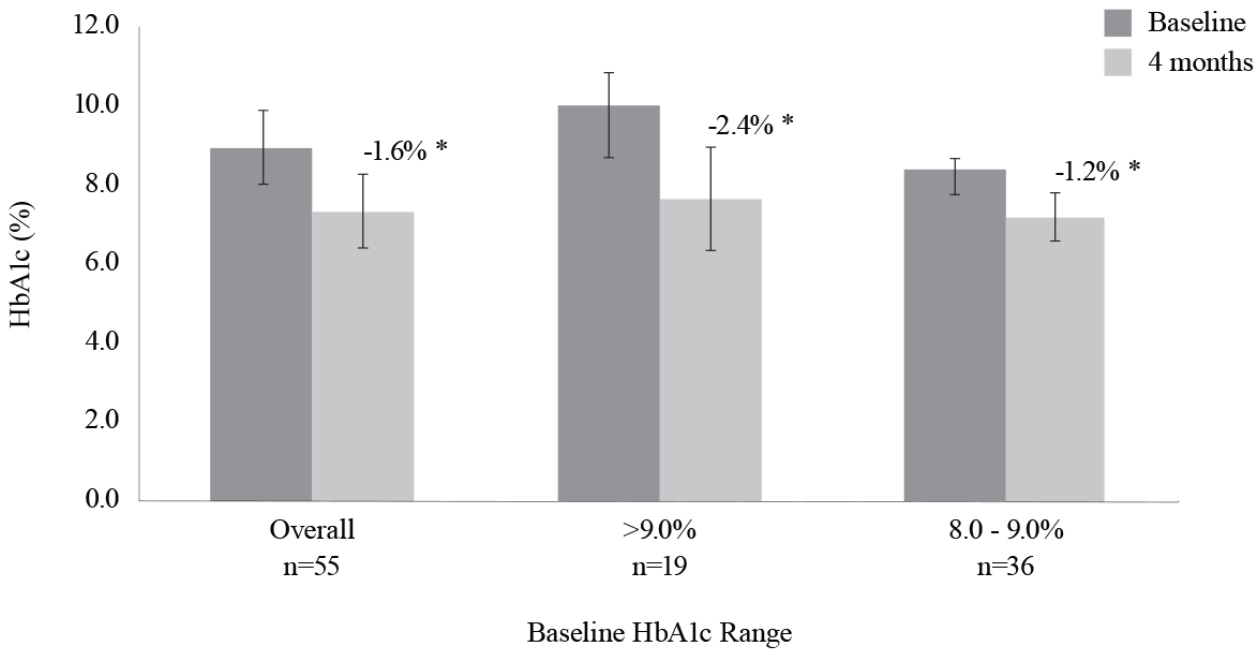
Variables	Values
Female, n (%)	22 (40)
Age (years) mean (SD)	57.3 (11.6)
Weight (lb) mean (SD)	218.7 (59.7)
BMI,[a] mean (SD)	33.7 (7.2)
Baseline HbA_{1c}, (%) mean (SD)	8.9 (1.0)
Systolic blood pressure (mm Hg) mean (SD)	132.1 (15.8)
Diastolic blood pressure (mm Hg) mean (SD)	80.7 (10.4)
Total cholesterol (mg/dL) mean (SD)	168.3 (42.8)
HDL[b] cholesterol (mg/dL) mean (SD)	40.4 (9.1)
LDL[c] cholesterol (mg/dL) mean (SD)	100.1 (36.5)
Non-HDL cholesterol (mg/dL) mean (SD)	128.0 (42.70)
Total cholesterol/HDL ratio, mean (SD)	4.4 (1.4)
Triglycerides (mg/dL) mean (SD)	236.7 (194.3)
Diabetes medications, n (%)	
0	0 (0)
1	6 (11)
2	21 (38)
≥3	28 (51)
Type of diabetes medications, n (%)	
Alpha glucosidase inhibitor	1 (2)
Biguanide	46 (84)
DPP-4[d] inhibitor	10 (18)
GLP-1[e] analogue	14 (25)
Insulin	20 (36)
SGLT2[f] inhibitor	20 (36)
Sulfonylurea	30 (55)
Thiazolidinedione	2 (4)
Lipid-lowering medications	44 (80)

[a]BMI: body mass index.
[b]HDL: high-density lipoprotein.
[c]LDL: low-density lipoprotein.
[d]DPP-4: dipeptidyl peptidase-4.
[e]GLP-1: glucagon-like peptide-1.
[f]SGLT2: sodium/glucose cotransporter 2.

Change in HbA1c

HbA_{1c} decreased significantly by 1.6% (SD 1%; *P*<.001), from 8.9% (SD 1%) at baseline to 7.3% (SD 0.9%) at 4 months (Figure 1). When stratified by baseline HbA_{1c} of 8.0% to 9.0% (n=36) and >9.0% (n=19), HbA_{1c} decreased by 1.2% (SD 0.6%) and 2.4% (SD 1.3%), respectively (both *P*<.001).

Figure 1. Change in HbA1c at 4 months from baseline. *P<.001.



CGM Device Use

Out of 55 participants in the study, 78.2% (43/55) met the criteria of follow-up CGM sensor–wear periods >90 days from baseline. Sensor-wear period was 94.8% (SD 8.2%) of the time specified per protocol (60 days) in this cohort. CGM metrics (n=43) are presented in Table 2. The increase in percent time in range 70 to 180 mg/dL was equivalent to a mean increase of 2.4 hours TIR per day.

Table 2. Changes in CGM metrics at 4 months from baseline.

Parameter	Baseline Mean (SD)	Follow-up Mean (SD)	Change Mean (SD)	*P* value
Mean glucose (mg/dL)	169.2 (29.3)	154.6 (33.0)	–14.6 (27.5)	.001
Coefficient of variation (%)	24.5 (4.9)	22.9 (5.5)	–1.6 (4.5)	.02
Percent time (%)				
<70 mg/dL	0.2 (0.4)	0.3 (0.6)	0.1 (0.7)	.49
70 to 180 mg/dL	65.4 (23.2)	75.5 (22.7)	10.2 (20.5)	.002
181 to 250 mg/dL	26.5 (14.9)	19.2 (15.5)	–7.2 (15.4)	.005
>250 mg/dL	8.0 (10.9)	5.0 (11.0)	–3.0 (9.4)	0.01

Weight, BP, and Serum Lipids

Change in weight, BP, and serum lipids at 4 months from baseline are presented in Table 3. Significant decreases were observed in weight, body mass index (BMI), systolic BP, total cholesterol, LDL cholesterol, total cholesterol/HDL ratio and triglycerides (P=.04 to P<.001).

Table 3. Change in BP and serum lipids at 4 months from baseline.

Parameter	Baseline Mean (SD)	Follow-up Mean (SD)	Change Mean (SD)	*P* value
Weight (lb)[a]	217.5 (59.5)	208.5 (53.7)	–9.0 (10.4)	<.001
BMI[ab]	33.6 (7.2)	32.2 (6.5)	–1.34 (1.5)	<.001
Systolic BP[ac] (mm Hg)	132.4 (15.8)	128.0 (16.6)	–4.4 (13.1)	.04
Diastolic BP (mm Hg)[a]	80.5 (10.5)	79.8 (10.5)	–0.8 (7.5)	0.48
Total cholesterol (mg/dL)	168.3 (42.8)	151.7 (41.1)	–16.6 (46.0)	<.001
HDL[ad] cholesterol, (mg/dL)	40.4 (9.1)	40.0 (10.9)	–0.4 (7.1)	.90
LDL[ae] cholesterol (mg/dL)	100.1 (36.5)	93.6 (31.9)	–6.5 (27.5)	.04
Total cholesterol/HDL Ratio	4.4 (1.4)	3.9 (1.3)	–0.5 (1.4)	.003
Triglycerides (mg/dL)	236.7 (194.30)	193.0 (163.2)	–43.7 (115.4)	.008

[a]For n=54 (1 subject did not complete the 4-month assessment at the study site, but submitted results from an external laboratory).

[b]BMI: body mass index.

[c]BP: blood pressure.

[d]HDL: high-density lipoprotein.

[e]LDL: low-density lipoprotein.

Adverse Events

There were no serious adverse events.

Discussion

In this prospective single-arm trial of the Onduo VDC, adults with T2D and suboptimal glycemic control experienced a statistically significant and clinically meaningful reduction in HbA_{1c} at 4 months. Analysis of RT-CGM metrics demonstrated a significant increase in TIR, decreased time in hyperglycemia, and found no increase in hypoglycemia. Participants also experienced significant decreases in weight, systolic BP, and serum lipids. Delivering RT-CGM devices, incorporating insights from intermittent RT-CGM use, and evaluating glycemic outcomes using RT-CGM data are unique aspects of the overall Onduo VDC care model for people with T2D.

Erhardt et al [4] and Vigersky et al [5] have reported on the use of intermittent RT-CGM in a population with T2D in a 52-week, two-arm, randomized controlled trial (RCT) that compared a 12-week active intervention of intermittent RT-CGM use to self-monitoring blood glucose. Similar to the present study, this RCT [4,5] also utilized Dexcom RT-CGM devices on an intermittent basis (2 weeks use, 1 week off) and measured short-term change in HbA_{1c}. Participants in the RCT (RT-CGM group n=50) and the present study were predominately male, similar for age, baseline HbA_{1c} and insulin use. Health insurance coverage differed: RCT participants were military health care beneficiaries compared to a mainly commercially insured population in the present study. In addition, the present study integrated intermittent RT-CGM data in Onduo VDC's coaching and telemedicine care model, while limited information was reported regarding the use of RT-CGM data in the RCT. A mean decrease in HbA_{1c} of 1% was observed in the RCT RT-CGM group vs 1.6% in the present study, suggesting a potential additive or synergistic benefit of combining intermittent RT-CGM and telehealth in people with T2D. Interestingly, TIR at the final assessment was similar in the RCT intervention group and the present study, 75.3% and 75.5%, respectively. During the 40-week follow-up period, RCT participants did not use RT-CGM, yet durable improvement in HbA_{1c} was observed at the 52-week follow-up assessment in the RT-CGM group compared with the self-monitoring blood glucose group. Similarly, the duration of our intervention was 4 months, and an evaluation of outcomes at 12-month outcomes is planned.

Previous studies of telehealth interventions in individuals with T2D have reported reductions in HbA_{1c} ranging from 0.7% to 2.1% [10-14]. Some of these studies have also reported improvements in secondary outcomes such as weight, lipids, and BP, although results are variable [11,13-16]. Direct comparisons to prior telehealth studies are complicated by differences in participant demographics, length of intervention, program features, and method of reporting HbA_{1c}, for example, directly measured vs estimated from fingerstick blood glucose readings. The magnitude of HbA_{1c} reduction in this prospective trial is consistent with that observed at an average of 4 months in a recently published retrospective analysis of 740 VDC participants [8]. Specifically, in participants with a baseline HbA_{1c} ≥8%, HbA_{1c} declined by 1.5% in the retrospective study and by 1.6% in the present study. In participants with baseline HbA_{1c} >9.0%, HbA_{1c} declined by 2.3% in the retrospective study and 2.4%, in this study.

Importantly, this is the first telehealth study in which participants with T2D were provided RT-CGM devices and changes in RT-CGM metrics beyond HbA_{1c} were quantified. Providing RT-CGM to participants in the VDC program supports individualized diabetes management in 2 ways. First, it improves self-management by showing participants how specific diet and lifestyle choices impact their blood glucose fluctuations. Second,

rich information on glycemic variability that is best derived from CGM data provides information to VDC endocrinologists, pharmacists, nutritionists, CDCES, and health coaches to guide lifestyle and therapeutic interventions.

At a population level, assessing changes in RT-CGM–derived glycemic outcomes supplements HbA_{1c}, offering additional insight into the impact of an intervention on diabetes management. Notably, the mean 10% increase in TIR observed in this study—equivalent to an increase of 2.4 hours per day—is considered a clinically meaningful increase [17]. Increased TIR is associated with lower vascular complication burden [5]. People with T2D report that TIR has at least as significant impact as HbA_{1c} has on their daily life [18]. The dual application of RT-CGM as a therapeutic and diagnostic tool is a unique strength of the Onduo VDC program and this study.

Limitations of our study include the sample size, duration of the intervention, and lack of a randomized control arm. It is important to note that some of the observed decrease in HbA_{1c} may be attributed to regression to the mean. Further studies are planned.

This prospective clinical trial of the Onduo VDC demonstrated improvements in glycemic outcomes in adults with suboptimally controlled T2D. Improvements in risk factors for diabetes complications, including weight, BP, and serum lipids were also observed.

Acknowledgments

The authors thank the study participants and Onduo clinical team, including the care leads and certified diabetes educators. We would like to thank John Schmucker and Anne Vermeulen for their leadership. We thank the staff at the participating research centers, particularly, Ashley Wilferd and the staff of the Allegheny Singer Research Institute of the Allegheny Health Network, Pittsburgh, PA; Elizabeth Mesghina and the staff of Sutter Health; and Palo Alto Medical Foundation Research Institute, Palo Alto, CA. We appreciate the efforts of the research and clinical teams at Verily Life Sciences, including Riddhi Kapoor, Maryam Yaghini, and Karina Brennecke. Funding for the study was provided by Onduo LLC, Newton, MA. Some of these data were presented at the 80th Scientific Sessions of the American Diabetes Association, A Virtual Experience, June 12-16, 2020.

Authors' Contributions

All authors contributed to the review of the report and approved the final version for submission. FRC, DME, and RJR contributed to the acquisition of data and all authors contributed to the interpretation of data. JEL developed the first draft of the manuscript. All authors contributed with a critical revision of the first and subsequent manuscript versions. RFD, CMK, AAL, ARM, DPM, HZ, and SR contributed to the study design. ARM is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

ARM has received consulting fees from Onduo LLC. RFD and JEL are employees of Onduo LLC, a joint venture of Verily Life Sciences and Sanofi. CMK, AAL, HZ, and DPM are employees and shareholders of Verily Life Sciences. SR was an employee of Verily Life Sciences at the time the work was completed. DME is an employee of Onduo Professionals, PC, which provides clinical services to Onduo LLC. RJR and FRC report no conflicts of interest.

References

1. Levine BJ, Close KL, Gabbay RA. Reviewing U.S. connected diabetes care: the newest member of the team. Diabetes Technol Ther 2020;22(1):1-9. [doi: 10.1089/dia.2019.0273] [Medline: 31483160]
2. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2020. Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services; 2020. URL: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [accessed 2020-08-11]
3. American Diabetes Association. 7. Diabetes Technology: Standards of Medical Care in Diabetes — 2020.. Diabetes Care 2020 Jan;43(Suppl 1):S77-S88. [doi: 10.2337/dc20-S007] [Medline: 31862750]
4. Ehrhardt NM, Chellappa M, Walker MS, Fonda SJ, Vigersky RA. The effect of real-time continuous glucose monitoring on glycemic control in patients with type 2 diabetes mellitus. J Diabetes Sci Technol 2011 May 01;5(3):668-675 [FREE Full text] [doi: 10.1177/193229681100500320] [Medline: 21722581]
5. Vigersky RA, Fonda SJ, Chellappa M, Walker MS, Ehrhardt NM. Short- and long-term effects of real-time continuous glucose monitoring in patients with type 2 diabetes. Diabetes Care 2012;35(1):32-38 [FREE Full text] [doi: 10.2337/dc11-1438] [Medline: 22100963]
6. Beck RW, Riddlesworth TD, Ruedy K, Ahmann A, Haller S, Kruger D, DIAMOND Study Group. Continuous glucose monitoring versus usual care in patients with type 2 diabetes receiving multiple daily insulin injections: a randomized trial. Ann Intern Med 2017;167(6):365-374. [doi: 10.7326/M16-2855] [Medline: 28828487]
7. Park C, Le QA. The effectiveness of continuous glucose monitoring in patients with type 2 diabetes: a systematic review of literature and meta-analysis. Diabetes Technol Ther 2018 Sep;20(9):613-621. [doi: 10.1089/dia.2018.0177] [Medline: 30095980]

8. Dixon RF, Zisser H, Layne JE, Barleen NA, Miller DP, Moloney DP, et al. A virtual type 2 diabetes clinic using continuous glucose monitoring and endocrinology visits. J Diabetes Sci Technol 2019 Nov 25:1932296819888662. [doi: 10.1177/1932296819888662] [Medline: 31762302]
9. Polonsky W, Layne J, Parkin C, Kusiak C, Barleen N, Miller D, et al. Impact of participation in a virtual diabetes clinic on diabetes-related distress in individuals with type 2 diabetes. Clin Diabetes 2020 Jun 12:cd190105-cd190106 [FREE Full text] [doi: 10.2337/cd19-0105]
10. Bollyky JB, Bravata D, Yang J, Williamson M, Schneider J. Remote lifestyle coaching plus a connected glucose meter with certified diabetes educator support improves glucose and weight loss for people with type 2 diabetes. J Diabetes Res 2018;2018:3961730 [FREE Full text] [doi: 10.1155/2018/3961730] [Medline: 29888288]
11. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
12. Bandi A, Larson M, Beattie J. MON-LB002 Diabetes Care Network: A telehealth based collaborative approach to scale endocrine expertise. Journal of the Endocrine Society 2019;3(Supplement_1):1-4. [doi: 10.1210/js.2019-MON-LB002]
13. Wang G, Zhang Z, Feng Y, Sun L, Xiao X, Wang G, et al. Telemedicine in the management of type 2 diabetes mellitus. Am J Med Sci 2017 Jan;353(1):1-5. [doi: 10.1016/j.amjms.2016.10.008] [Medline: 28104096]
14. Sun C, Sun L, Xi S, Zhang H, Wang H, Feng Y, et al. Mobile phone-based telemedicine practice in older chinese patients with type 2 diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e10664 [FREE Full text] [doi: 10.2196/10664] [Medline: 30609983]
15. Hallberg SJ, McKenzie AL, Williams PT, Bhanpuri NH, Peters AL, Campbell WW, et al. Effectiveness and safety of a novel care model for the management of type 2 diabetes at 1 year: an open-label, non-randomized, controlled study. Diabetes Ther 2018;9(2):583-612 [FREE Full text] [doi: 10.1007/s13300-018-0373-9] [Medline: 29417495]
16. Jeong JY, Jeon J, Bae K, Choi Y, Park K, Kim J, et al. Smart care based on telemonitoring and telemedicine for type 2 diabetes care: multi-center randomized controlled trial. Telemed J E Health 2018;24(8):604-613. [doi: 10.1089/tmj.2017.0203] [Medline: 29341843]
17. Battelino T, Danne T, Bergenstal RM, Amiel SA, Beck R, Biester T, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. Diabetes Care 2019;42(8):1593-1603. [doi: 10.2337/dci19-0028] [Medline: 31177185]
18. Runge AS, Kennedy L, Brown AS, Dove AE, Levine BJ, Koontz SP, et al. Does time-in-range matter? Perspectives from people with diabetes on the success of current therapies and the drivers of improved outcomes. Clin Diabetes 2018;36(2):112-119 [FREE Full text] [doi: 10.2337/cd17-0094] [Medline: 29686449]

Abbreviations

BMI: body mass index
BP: blood pressure
CDCES: Certified Diabetes Care and Education Specialist
CGM: continuous glucose monitoring
HbA1c: glycated hemoglobin
HDL: high-density lipoprotein
LDL: low-density lipoprotein
RCT: randomized controlled trial
RT-CGM: real-time continuous glucose monitoring
T2D: type 2 diabetes
TIR: time in range
VDC: virtual diabetes clinic

Edited by G Eysenbach; submitted 07.07.20; peer-reviewed by D Rodbard, L Heinemann; comments to author 11.07.20; revised version received 20.07.20; accepted 23.07.20; published 28.08.20.

Please cite as:
Majithia AR, Kusiak CM, Armento Lee A, Colangelo FR, Romanelli RJ, Robertson S, Miller DP, Erani DM, Layne JE, Dixon RF, Zisser H
Glycemic Outcomes in Adults With Type 2 Diabetes Participating in a Continuous Glucose Monitor–Driven Virtual Diabetes Clinic: Prospective Trial
J Med Internet Res 2020;22(8):e21778
URL: https://www.jmir.org/2020/8/e21778
doi:10.2196/21778
PMID:32856597

Original Paper

An Image-Based Mobile Health App for Postdrainage Monitoring: Usability Study

Chien-Hung Liao[1*], MD, FACS, FICS; Yu-Tung Wu[1*], MD; Chi-Tung Cheng[1], MD; Chun-Hsiang Ooyang[1], MD; Shih-Ching Kang[1], MD; Chih-Yuan Fu[1], MD; Yu-Pao Hsu[1], MD; Chi-Hsun Hsieh[1], MD, PhD; Chih-Chi Chen[2], MD

[1]Department of Trauma and Emergency Surgery, Linkou Chang Gung Memorial Hospital, Chang Gang University, Taoyaun, Taiwan

[2]Department of Physical Medicine and Rehabilitation, Linkou Chang Gung Memorial Hospital, Chang Gang University, Taoyaun, Taiwan

[*]these authors contributed equally

Corresponding Author:
Chih-Chi Chen, MD
Department of Physical Medicine and Rehabilitation
Linkou Chang Gung Memorial Hospital
Chang Gang University
5, Fuxin street, Kweishiang district
Taoyaun, 333
Taiwan
Phone: 886 33281200 ext 3651
Email: Claudia5477@gmail.com

Abstract

Background: The application of mobile health (mHealth) platforms to monitor recovery in the postdischarge period has increased in recent years. Despite widespread enthusiasm for mHealth, few studies have evaluated the usability and user experience of mHealth in patients with surgical drainage.

Objective: Our objectives were to (1) develop an image-based smartphone app, SurgCare, for postdrainage monitoring and (2) determine the feasibility and clinical value of the use of SurgCare by patients with drainage.

Methods: We enrolled 80 patients with biliary or peritoneal drainage in this study. A total of 50 patients were assigned to the SurgCare group, who recorded drainage monitoring data with the smartphone app; and 30 patients who manually recorded the data were assigned to the conventional group. The patients continued to record data until drain removal. The primary aim was to validate feasibility for the user, which was defined as the proportion of patients using each element of the system. Moreover, the secondary aim was to evaluate the association of compliance with SurgCare and the occurrence of unexpected events.

Results: The average submission duration was 14.98 days, and the overall daily submission rate was 84.2%. The average system usability scale was 83.7 (SD 3.5). This system met the definition of “definitely feasible” in 34 patients, “possibly feasible” in 10 patients, and “not feasible” in 3 patients. We found that the occurrence rates of complications in the SurgCare group and the conventional group were 6% and 26%, respectively, with statistically significant differences P=.03. The rate of unexpected hospital return was lower in the SurgCare group (6%) than in the conventional groups (26%) (P=.03).

Conclusions: Patients can learn to use a smartphone app for postdischarge drainage monitoring with high levels of user satisfaction. We also identified a high degree of compliance with app-based drainage-recording design features, which is an aspect of mHealth that can improve surgical care.

(J Med Internet Res 2020;22(8):e17686) doi:10.2196/17686

KEYWORDS

telemedicine; smartphone; surgical drainage; postdrainage care, mHealth

Introduction

Surgical drainage is a therapeutic procedure with multiple purposes, including relieving symptoms, bypassing occlusions, and monitoring postoperative conditions [1-4]. In patients with acute biliary diseases, such as acute cholecystitis or cholangitis, drainage can relieve symptoms and stabilize the patient’s condition [5-7]. Drainage can also postpone emergency surgery, undergoing interval surgery instead [8,9]. Sometimes, patients need to monitor their own drains because of a prolonged

therapeutic course. Traditional postdischarge drainage care depends on medical professionals asking patients to record the amount and characteristics of the fluid drained. However, these self-report measures are not only unreliable in elderly adults and those with impaired cognition [10,11], but also time-consuming with regard to processing the data [12]. Inadequate monitoring and care might prolong drainage insertion, delay recovery, reduce quality of life, and induce sequential complications such as an electrolyte imbalance, dehydration, sepsis, or physical injury related to disruption of the drain placement [13,14]. Therefore, proper monitoring of the drainage status is a critical issue.

In the last decade, apps for mobile devices have radically changed modern lifestyles. Additionally, the healthcare sector has been enriched by numerous apps [15,16]. Because of the increase in popularity of this new technology, the World Health Organization (WHO) has defined these tools as electronic health (eHealth) and mobile health (mHealth) applications. As the ownership of mobile devices has become more common [17], patients and their caregivers are increasingly willing to use technology to access health care [18,19]. Several studies have demonstrated that mHealth technology improves the control of cardiac function and glycemic hemostasis, enhances medication compliance, and shortens hospital stays [20-24]. Additionally, prior research on app protocols for surgical patients has focused on routine procedures that already have a low baseline rate of postoperative and discharge complications [25-27]. Although the experience with using mHealth apps in surgical care is limited, it has been suggested that surgical patients benefit from this new technological mode of support [28,29].

We developed an internet-based remote app to monitor drainage and conducted a study to investigate the adequacy of the remote app with regard to helping patients and caregivers properly manage drains at home. This study focused on the feasibility and clinical value of the remote app for patients with percutaneous or surgical drainage.

Methods

Study Population

Patients who were eligible to participate in this study were adult inpatients (aged ≥20 years) in the acute care surgery department of a medical center. We enrolled patients who were undergoing percutaneous or surgical drainage of the biliary tract or peritoneal cavity at our department from May 2019 to October 2019. All patients who fulfilled the inclusion criteria were approached to participate in this study. Notably, patients were excluded if they had neurologic or cognitive disorders prohibiting their usage of the app or ability to give informed consent. To calculate the sample size, we used the following parameters: α=.05, a power of 80%, an enrollment ratio of 1.6, and a complication decreasing rate of 18%. We recruited 50 patients to participate in the app group and 30 patients to participate in the conventional group.

The institutional review board of Chang Gung Memorial Hospital approved this study protocol (201900495B0). A research assistant helped the enrolled patients and caregivers install the app on their smartphones and instructed them how to use the app before the operation. Before patients were enrolled in this study, the research assistant evaluated their familiarity with wearable devices and smartphones. If the patients were not confident about using these devices, we provided further instructions to their caregivers.

SurgCare App

SurgCare (Figure 1) is an iOS/Android app that facilitates the recording of postprocedural clinical variables (drainage amount, discharge color, associated discomfort, body weight, and analgesic ingestion) by patients who have had drains placed.

The app transmits digital images of the surgical wound and drainage to the medical staff as shown in Figure 2.

SurgCare was developed by surgical professionals and software programmers to fulfill the needs of patients caring for drains at home and was only used by patients who were followed in our institute. Figure 3 presents the system architecture. Once the user inputs the data into the app, the information is synchronized with the server when internet access is available on the smartphone. This interdevice data transmission worked well, with no abnormal events reported by the patients.

Furthermore, the research assistant monitored the synchronizing of the data on weekdays and called the patients if their information was missing. If there was a change in fluid color, persistent changes in vital signs or abnormal drainage, this information was provided by smart devices to the medical doctors who then arranged further management. Patients continued to transmit data until their drains were removed. The research assistant did not have contact with the conventional group when patients were at home. Another medical staff member who did not participate in designing this system independently reviewed and analyzed the data.

Figure 1. Screenshots of the SurgCare app showing the records of date, drain method, drainage amount, color, and image of drained material and wound status.

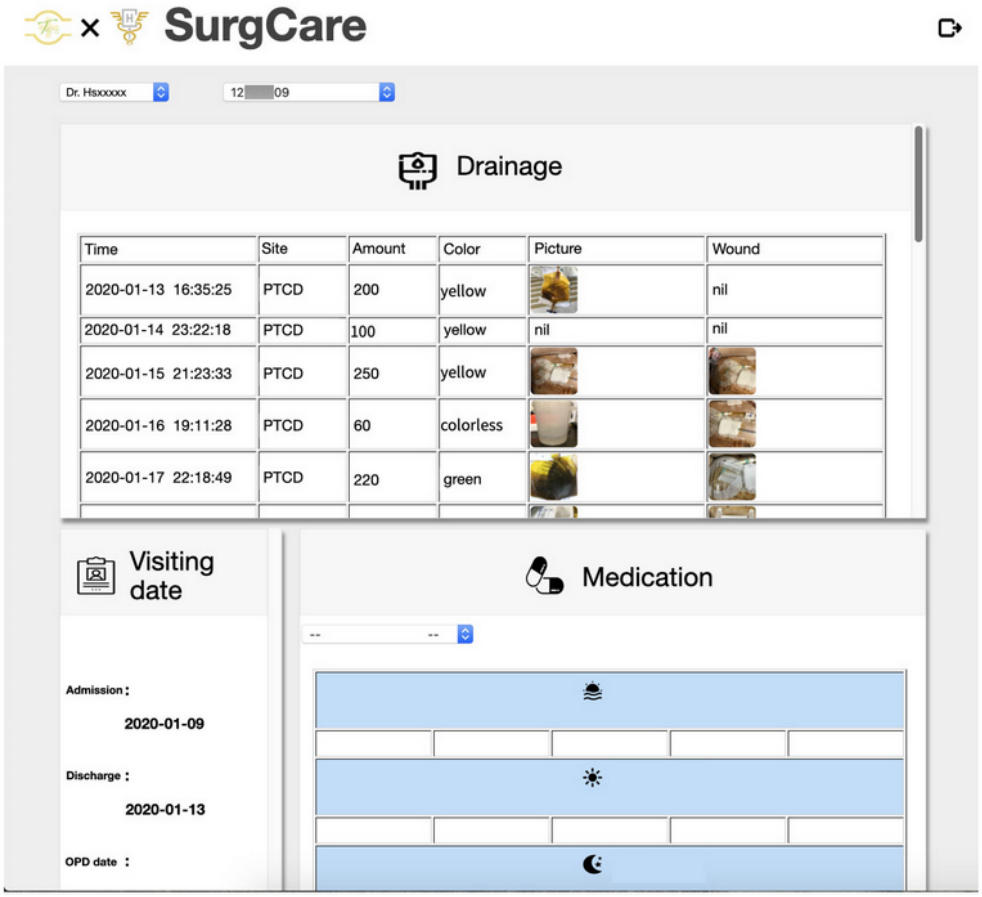


Figure 2. Image of drainage-inserted wound and drainage content. A: wound with percutaneous gallbladder drainage; B: wound with percutaneous biliary drainage; C: yellowish bile drainage; D: mucus-like white bile drainage.

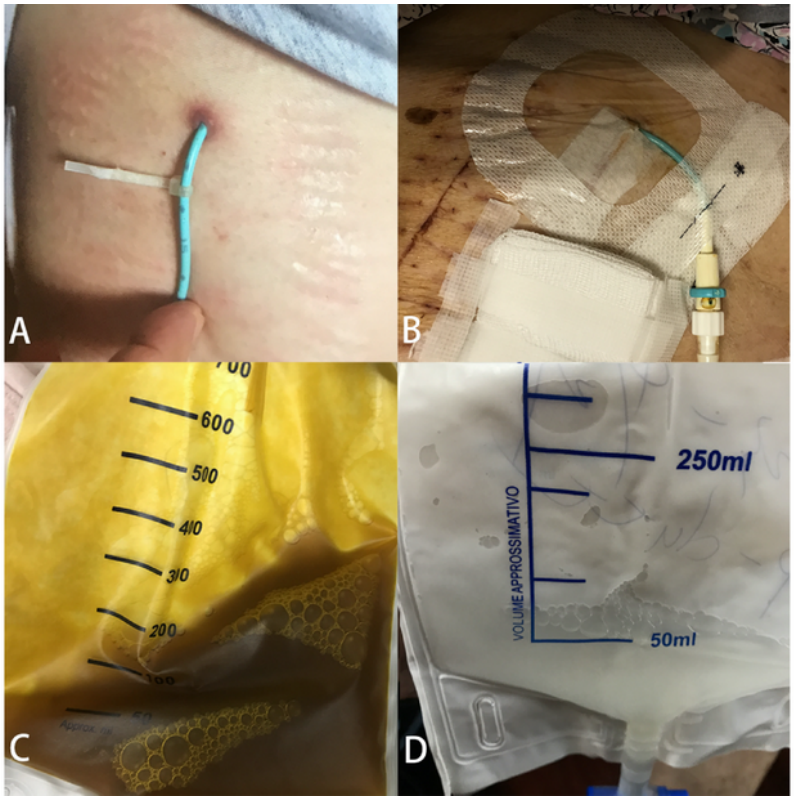
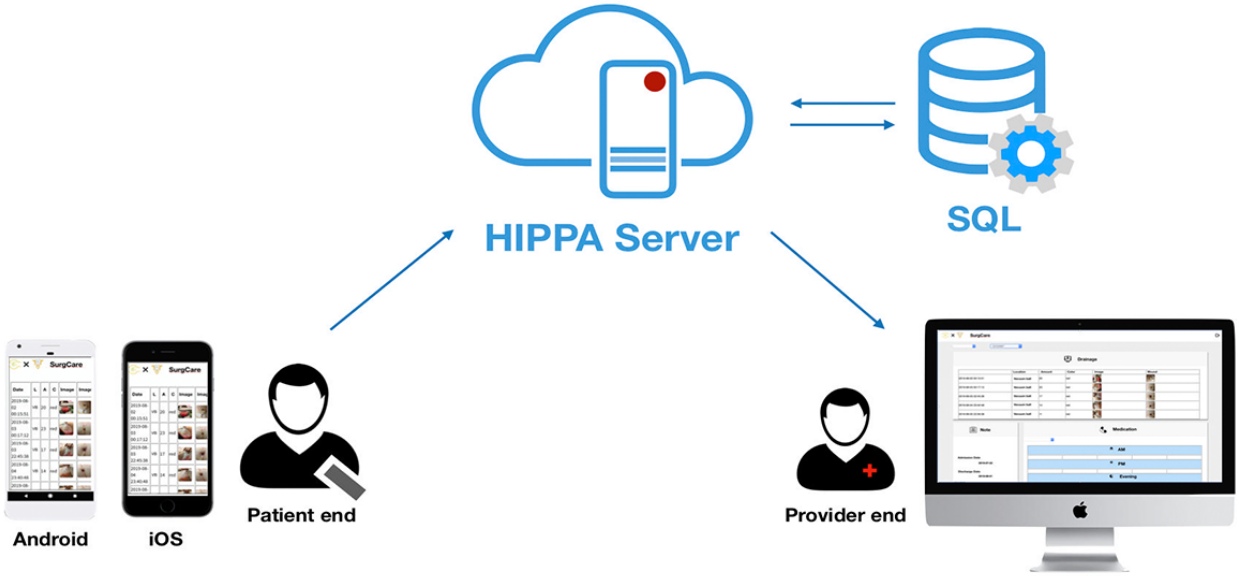


Figure 3. System architecture of the mobile device for recording postdrainage care.



User Tasks

We formally tested the usability and feasibility of the app with postdischarge drainage patients at a major academic medical center. The app was loaded onto an iOS or Android smartphone or tablet. We assessed patients' baseline familiarity with smartphones prior to testing. User tasks included waking up the device, launching the app, inputting information (including drainage amount, color, and the presence of discomfort), capturing an image, reviewing and retaking or accepting captured images, responding to questions, and submitting the data.

Measures and Analysis

Feasibility: Protocol Completion

Following usability testing of the app, participants were asked to rate their performance and to provide feedback on the app. Participants also used a system usability scale to evaluate their satisfaction with the app [30].

We evaluated the compliance of the patients with the use of SurgCare. If patients submitted data on more than 80% of days on which they had a drain, they were classified in the good compliance group. If the number of days on which data were submitted was less than 80% of the total number of days for which they had the drains, we classified them in the poor compliance group.

In this study, the feasibility was assessed as in past studies [31]. The feasibility was defined as the proportion of participants using each element of the system for at least 70% of the period. "Definitely feasible," "possibly feasible," and "not feasible" were defined as ≥70%, 50%-69%, and <50% of the participants meeting that criterion, respectively.

Clinical Value: Association of App Usage With Early Complication Rate

The clinical outcome of interest was validation that the use of the app leads to fewer complications and a lower incidence of unexpected hospital return. We compared the high compliance group with the poor compliance group and conventional group with regard to the total compliance rate, incidence of drain dysfunction, incidence of drain dislodgement, and rate of infection. We also analyzed the rates of unexpected hospital return and readmission in these groups.

Statistical Analysis

Pearson's chi-square test and Fisher exact test were used to compare categorical variables. Quantitative variables were compared with Student's *t* test. Levene's test was used to correct for intergroup variations before the application of Student's *t* test. Statistical analysis was performed with SPSS v 20.0 for Macintosh (SPSS Inc). A value of $P<.05$ was considered statistically significant.

Results

Participant Characteristics

In total, 105 patients underwent 108 procedures during the study period. Out of these 105 patients, 3 patients with repeated procedures were excluded. After being approached, 8 patients refused to participate in this project, and another 14 patients had cognitive problems and were excluded from this study. Of the 80 patients who were approached, were eligible, and agreed to participate, 6 patients were lost to follow-up. A total of 47 participants completed the usability testing, 26 of whom had caregiver assistance or proxy participation. Another 27 patients were included in the conventional group; these patients used the traditional hand recording method to track the drainage amount and color. The study flow diagram is shown in Figure 4.

Demographics and basic clinical information of SurgCare group are presented in Table 1. The average duration of app use was 14.5 (SD 3.2) days. The common termination of follow-up was drainage removal and wound closure, followed by a lack of satisfaction and complications.

Figure 4. Flow diagram of the study design.

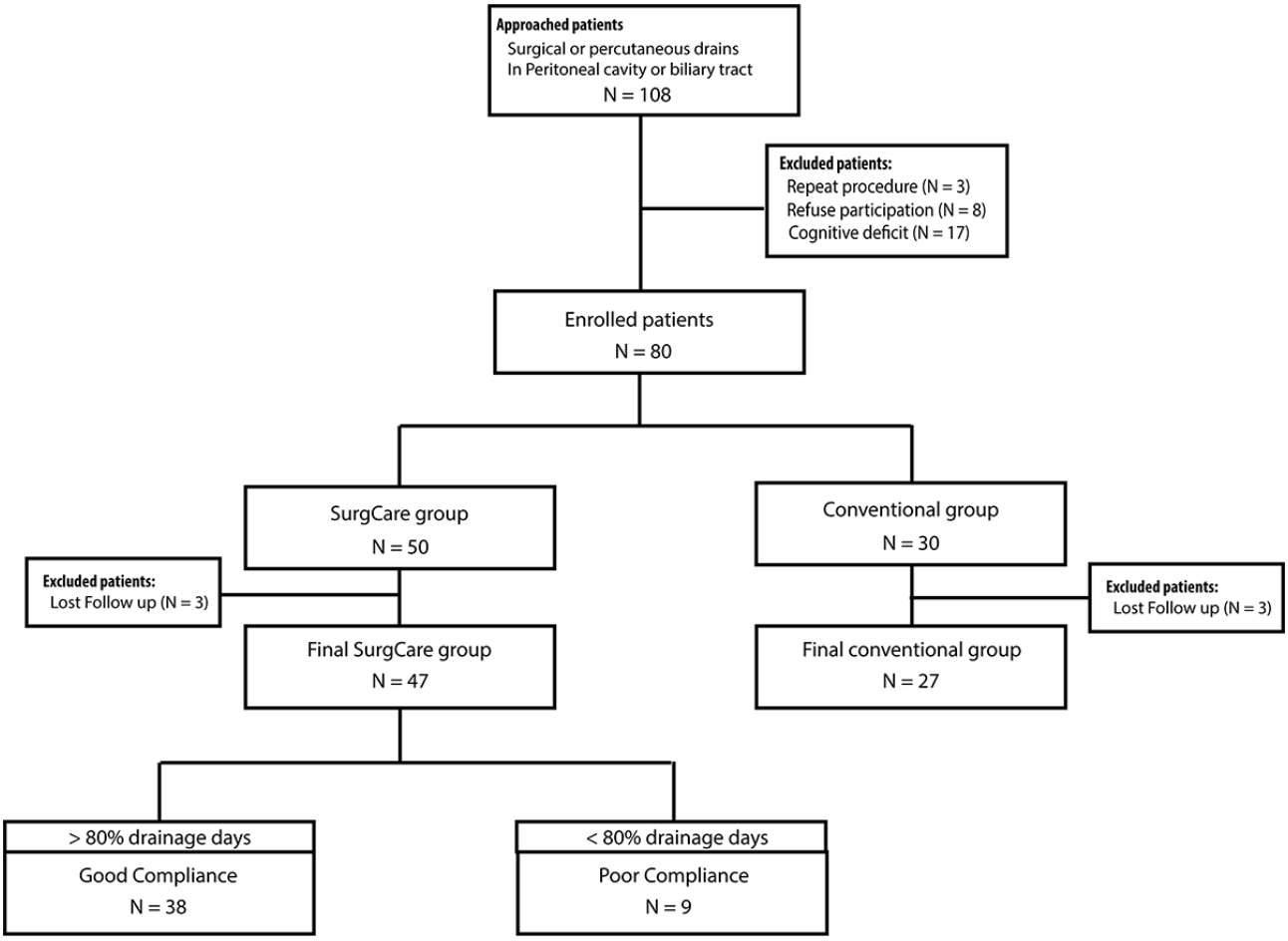


Table 1. Demographic characteristics of SurgCare participants with surgical and percutaneous drainage (N=47).

Characteristics	Value
Age, mean (SD)	61.4 (15.9)
Gender, n (%)	
Male	29 (62)
Female	18 (38)
Underlying disease, n (%)	
Hypertension	21 (45)
Diabetes mellitus	16 (34)
Renal insufficiency	7 (15)
Chronic obstructive pulmonary disease	3 (6)
Malignancy	3 (6)
Drainage site, n (%)	
Biliary drainage	25 (53)
Peritoneal drainage	22 (47)
Method of participation, n (%)	
Independent	21 (45)
Caregivers	26 (55)
System usability scale, mean (SD)	83.7 (3.5)
Feasibility test, n (%)	
Definite feasible	34 (72)
Possibly feasible	10 (21)
Not feasible	3 (6)
Overall daily compliance	
Total drainage days, N	836
Day submitted, n (%)	704 (84)
Day missed, n (%)	132 (16)
Complication, n (%)	
Dysfunction	1 (2)
Dislodge	0 (0)
Infection	2 (4)
Unexpected return	3 (6)
Unexpected readmission	2 (4)

To compare SurgCare group with the conventional group, we identified gender, age, underlying chronic disorders, drainage site, and method of participation for both the groups. Total occurrence of complication was lower in SurgCare group (6%) than in the conventional group (26%) with statistical significance of *P*=.03. The incidence of drainage dislodge was lower in SurgCare group (2%) than in the conventional group (11%) with statistical significance of *P*=.045. Moreover, the unexpected hospital return was lower in SurgCare group (6%) than in the conventional group (26%) with significant difference of *P*=.03 (Table 2).

Table 2. Comparison of the characteristics and prognosis of patients within the SurgCare and conventional groups.

Characteristics	SurgCare	Conventional	*P* value
Patients, n	47	27	—[a]
Age, mean (SD)	60.2 (17.1)	63.6 (13.4)	.36
Gender, n (%)			.60
Male	29 (62)	15 (56)	
Female	16 (38)	12 (44)	
Underlying disease, n (%)			
Hypertension	21 (45)	8 (30)	.23
Diabetes mellitus	16 (34)	6 (22)	.43
Renal insufficiency	7 (15)	5 (19)	.75
Chronic obstructive pulmonary disease	3 (6)	2 (7)	>.99
Malignancy	3 (6)	3 (11)	.66
Drainage site, n (%)			.08
Biliary drainage	25 (53)	20 (74)	
Peritoneal drainage	22 (47)	7 (26)	
Method of participation, n (%)			.37
Independent	21 (45)	15 (56)	
Caregivers	26 (55)	12 (44)	
Complication, n (%)	3 (6)	7 (26)	.03[b]
Dysfunction	1 (2)	3 (11)	.14
Dislodge	0 (0)	3 (11)	.045[b]
Infection	2 (4)	1 (4)	>.99
Unexpected hospital return	3 (6)	7 (26)	.03[b]
Unexpected readmission	2 (4)	4 (15)	.11

[a]Not applicable.

[b]Fisher Exact test.

Feasibility and Usability Evaluation of SurgCare

After evaluation, there were 34 patients for whom it was definitely feasible to use this app. For 10 patients, use of the app might be feasible, although they needed more support from the research assistant to help them operate the system. Another 3 patients were in the infeasible group because they completed less than 30% of the elements. The overall system usability score for the app was 83.3, which is considered good in usability testing.

Association of SurgCare App Usage Compliance and Early Complication Rate

In the good compliance group, we found that the rate of complications related to drainage was 3%, which was much lower than that in the poor compliance group (11%). With regard to unexpected hospital return (3% vs 11%) and readmission (3% vs 11%), the good compliance group had better results than the poor compliance group and the conventional group as shown in Table 3.

Table 3. Comparison of the prognosis of patients with SurgCare with good and poor compliance.

Characteristics	Good compliance	Poor compliance	*P* value
Patients, n	38	9	—[a]
Age, mean (SD)	60.3 (18.1)	60.0 (13.1)	.96
Gender, n (%)			.27
Male	22 (58)	7 (78)	
Female	16 (42)	2 (22)	
Underlying disease, n (%)			
Hypertension	16 (42)	5 (56)	.47
Diabetes mellitus	15 (40)	1 (11)	.11
Renal insufficiency	7 (18)	0 (0)	.32
Chronic obstructive pulmonary disease	1 (3)	2 (22)	.09
Malignancy	2 (5)	1 (11)	.48
Drainage site, n (%)			.87
Biliary drainage	20 (53)	5 (56)	
Peritoneal drainage	18 (47)	4 (44)	
Method of participation, n (%)			.14
Independent	15 (40)	6 (67)	
Caregivers	23 (61)	3 (33)	
Complication, n (%)			
Dysfunction	1 (3)	0 (0)	>.99
Dislodge	0 (0)	0 (0)	>.99
Infection	1 (3)	1 (11)	.32
Unexpected hospital return	2 (5)	1 (11)	.52
Unexpected readmission	1 (3)	1 (11)	.26

[a]Not applicable.

Discussion

This study demonstrated the feasibility of using a mobile app to monitor the recovery status of patients with drains and to assist patients and caregivers in detecting abnormalities in a timely manner. Remote apps could support self-care and allow close follow-up [32-34]. We also identified an unexpected reduction in the rates of hospital return and readmission in the SurgCare group (6.4%) compared with the conventional group (25.9%) (*P*=.02). We found that the occurrences of complications such as dislodgement (0%), infection (4%), and dysfunction (2%) were relatively fewer in the SurgCare group. To the best of our knowledge, this is one of the first innovative studies focusing on the development of comprehensive app functions to assist surgical patients with drain care and monitoring. With the evaluation of uploaded image of wound and drainage content, the health care team can identify the abnormalities earlier to prevent the sequential complications, which is one of the causes to reduce unexpected hospital return. The current standard of care for the majority of surgical patients following hospital discharge involves little formal communication between patients and their care team until their routine clinic follow-up 2-3 weeks after discharge [35,36]. Some mHealth protocols have been developed to improve patient monitoring or replace routine postoperative clinic visits [37-39]. In addition to this trend, the national policy priority mandates improving transitions of care following hospital discharge and reducing hospital readmissions [40,41]. We created an image-based mobile app aimed at increasing communication between patients and health care personnel after discharge from the hospital as part of an effort to detect drainage complications in an early stage and reduce hospital readmissions.

The compliance with using apps is generally high. Apps for surgical patients must be developed carefully, keeping in mind that the users are very vulnerable [42,43]. In this study, we find that compliance with using the SurgCare was acceptable. Most participants found the app easy to use (%), though the questions that did not elicit an unanimously positive response indicate that there is a degree of tentativeness regarding participants' ability to independently perform the functions in the SurgCare. The SurgCare can provide assistance not only by facilitating monitoring but also by providing psychological support [44,45]. For patients and caregivers who are not familiar with postoperative wound care, remote support helps them detect the early dislodgement or dysfunction of the drain, and it can reduce the subsequent occurrence of infections and other

complications. Self-report questionnaires are the most common method of monitoring drainage because of their cost-effectiveness and ease of administration [46]. However, the disadvantages of self-report questionnaires include a lack of reliability and the influences of social desirability, age, questionnaire complexity, and recall ability [47,48]. Therefore, an easily used mobile app can solve these problems. With immediate recording and image capture, the app is an excellent tool for close monitoring. Moreover, SurgCare can offer a rapid response to help patients psychologically, as patients and caregivers can receive responses from healthcare personnel before returning to the clinic. The daily monitoring messages are delivered via mobile messaging (eg, short message service) instead of email to facilitate more immediate communication. These adjustments are expected to improve user satisfaction and compliance and may ultimately further enhance the efficacy of the intervention.

Furthermore, 4 key categories of age-related barriers are associated with the use of mHealth by older adults, namely, barriers related to cognition, motivation, physical ability, and perception [26]. As surgical drainage patients are usually older adults (median age >60 years) who have the potential to develop cognitive or memory impairments, it is crucial to select an easy-to-follow app to use in clinical research with this population. We noticed that if the patients were well trained, they were able to input their health data by themselves without any dependence on a physician's assistant, study nurse, or other caregivers [49,50]. After the interview, we also noticed that the surgeons were interested in the electronic assessment of patient-reported outcomes rather than the conventional manual reporting of the drainage amount and status. Previous studies have suggested that the electronic assessment of patient-reported outcomes was as accurate as the conventional method of manual recording [51,52]. Another survey of 108 health care personnel showed a high level of acceptance (84.3%) of app-assisted recording [53]. Digital medicine is unstoppable, and patient empowerment plays a new and growing role in disease management. With support from both patients and health providers, we can determine the impact of mHealth on the postprocedural care of surgical drains.

In contrast to the generally rapid growth of mHealth in medical fields, research on surgical topics has been limited. Among studies focusing on the application of mHealth to surgical issues, several have investigated wound care and pain scaling to validate the clinical usefulness of mHealth [54-57]. A recent study used another wearable device to track the step counts of patients who had undergone various abdominal surgeries for 1 month after discharge and showed that the mHealth app could effectively track recovery [24,58]. Because drain care is an issue that is unique to surgical patients, and telephone conversations and questionnaires cannot be used to access the visual component. In this study, we developed an app that can improve compliance with postdischarge drainage care and monitoring and reduce the risk of drainage complications.

Limitations

This study had several limitations that should be considered when interpreting the results. First, all assessments were conducted online, and inclusion relied exclusively on self-reported data. Therefore, the internal validity and generalizability to a larger clinical population might have been compromised. Second, there was a rather low postassessment response rate. The lack of data from approximately one-fifth of the sample limits the validity of our findings since it remains unclear how satisfied the nonresponders were and how they differed in terms of symptomatology. Third, the sample size was limited, and the sample did not represent all surgical patients. Subsequent research should, therefore, investigate the efficacy and cost-effectiveness of SurgCare in a fully powered randomized control trial. The study provided valuable information about the feasibility and adequacy of an internet-based intervention for the management of drains, which can be used to guide subsequent research.

Conclusion

In this study, we present a remote app that can improve patient compliance with postdischarge drainage care and monitoring and reduce the rate of major complications. The patients were enthusiastic about partnering with their health providers in novel ways to optimize their healthcare. Although mHealth will certainly not replace physician contact, it will serve as a digital assistant for diagnostic, therapeutic, and follow-up purposes, supporting patient recovery.

Conflicts of Interest

None declared.

References

1. Kortram K, de Vries Reilingh TS, Wiezer MJ, van Ramshorst B, Boerma D. Percutaneous drainage for acute calculous cholecystitis. Surg Endosc 2011 Nov;25(11):3642-3646 [FREE Full text] [doi: 10.1007/s00464-011-1771-5] [Medline: 21638173]
2. Chousleb E, Szomstein S, Podkameni D, Soto F, Lomenzo E, Higa G, et al. Routine abdominal drains after laparoscopic Roux-en-Y gastric bypass: a retrospective review of 593 patients. Obes Surg 2004 Oct;14(9):1203-1207. [doi: 10.1381/0960892042387101] [Medline: 15527635]
3. Wallace MJ, Chin KW, Fletcher TB, Bakal CW, Cardella JF, Grassi CJ, Society of Interventional Radiology (SIR). Quality improvement guidelines for percutaneous drainage/aspiration of abscess and fluid collections. J Vasc Interv Radiol 2010 Apr;21(4):431-435. [doi: 10.1016/j.jvir.2009.12.398] [Medline: 20346880]

4. Swan MC, Oliver DW, Cassell OCS, Coleman DJ, Williams N, Morritt DG, et al. Randomized controlled trial of fibrin sealant to reduce postoperative drainage following elective lymph node dissection. Br J Surg 2011 Jul;98(7):918-924. [doi: 10.1002/bjs.7462] [Medline: 21456091]
5. Pessa ME, Hawkins IF, Vogel SB. The treatment of acute cholangitis. Percutaneous transhepatic biliary drainage before definitive therapy. Ann Surg 1987 Apr;205(4):389-392 [FREE Full text] [doi: 10.1097/00000658-198704000-00008] [Medline: 3566375]
6. Sartelli M, Viale P, Catena F, Ansaloni L, Moore E, Malangoni M, et al. 2013 WSES guidelines for management of intra-abdominal infections. World J Emerg Surg 2013 Jan 08;8(1):3 [FREE Full text] [doi: 10.1186/1749-7922-8-3] [Medline: 23294512]
7. Miura F, Okamoto K, Takada T, Strasberg SM, Asbun HJ, Pitt HA, et al. Tokyo Guidelines 2018: initial management of acute biliary infection and flowchart for acute cholangitis. J Hepatobiliary Pancreat Sci 2018 Jan;25(1):31-40 [FREE Full text] [doi: 10.1002/jhbp.509] [Medline: 28941329]
8. Kim KH, Sung CK, Park BK, Kim WK, Oh CW, Kim KS. Percutaneous gallbladder drainage for delayed laparoscopic cholecystectomy in patients with acute cholecystitis. Am J Surg 2000 Feb;179(2):111-113 [FREE Full text] [doi: 10.1016/s0002-9610(00)00247-6] [Medline: 10773145]
9. Bickel A, Hoffman RS, Loberant N, Weiss M, Eitan A. Timing of percutaneous cholecystostomy affects conversion rate of delayed laparoscopic cholecystectomy for severe acute cholecystitis. Surg Endosc 2016 Mar;30(3):1028-1033 [FREE Full text] [doi: 10.1007/s00464-015-4290-y] [Medline: 26139479]
10. Campbell NL, Zhan J, Tu W, Weber Z, Ambeuhl R, McKay C, et al. Self-Reported Medication Adherence Barriers Among Ambulatory Older Adults with Mild Cognitive Impairment. Pharmacotherapy 2016 Feb;36(2):196-202 [FREE Full text] [doi: 10.1002/phar.1702] [Medline: 26890914]
11. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
12. Pettee GK, McClain JJ, Schmid KK, Storti KL, Ainsworth BE. Reliability and convergent validity of the past-week Modifiable Activity Questionnaire. Public Health Nutr 2011 Mar;14(3):435-442. [doi: 10.1017/S1368980010002612] [Medline: 20843404]
13. McMillan MT, Fisher WE, Van Buren G, McElhany A, Bloomston M, Hughes SJ, et al. The value of drains as a fistula mitigation strategy for pancreatoduodenectomy: something for everyone? Results of a randomized prospective multi-institutional study. J Gastrointest Surg 2015 Jan;19(1):21-30; discussion 30 [FREE Full text] [doi: 10.1007/s11605-014-2640-z] [Medline: 25183409]
14. Wiggers JK, Coelen RJS, Rauws EAJ, van Delden OM, van Eijck CHJ, de Jonge J, et al. Preoperative endoscopic versus percutaneous transhepatic biliary drainage in potentially resectable perihilar cholangiocarcinoma (DRAINAGE trial): design and rationale of a randomized controlled trial. BMC Gastroenterol 2015 Feb 14;15:20 [FREE Full text] [doi: 10.1186/s12876-015-0251-0] [Medline: 25887103]
15. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
16. Goyal S, Lewis G, Yu C, Rotondi M, Seto E, Cafazzo JA. Evaluation of a Behavioral Mobile Phone App Intervention for the Self-Management of Type 2 Diabetes: Randomized Controlled Trial Protocol. JMIR Res Protoc 2016 Aug 19;5(3):e174 [FREE Full text] [doi: 10.2196/resprot.5959] [Medline: 27542325]
17. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
18. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
19. Kessel KA, Vogel MM, Kessel C, Bier H, Biedermann T, Friess H, et al. Mobile Health in Oncology: A Patient Survey About App-Assisted Cancer Care. JMIR Mhealth Uhealth 2017 Jun 14;5(6):e81 [FREE Full text] [doi: 10.2196/mhealth.7689] [Medline: 28615159]
20. Hung C, Yu J, Lin Y, Chen Y, Huang C, Lee J, et al. Mortality Benefit of a Fourth-Generation Synchronous Telehealth Program for the Management of Chronic Cardiovascular Disease: A Longitudinal Study. J Med Internet Res 2016 May 13;18(5):e102 [FREE Full text] [doi: 10.2196/jmir.5718] [Medline: 27177497]
21. Doocy S, Paik KE, Lyles E, Hei TH, Fahed Z, Winkler E, et al. Guidelines and mHealth to Improve Quality of Hypertension and Type 2 Diabetes Care for Vulnerable Populations in Lebanon: Longitudinal Cohort Study. JMIR Mhealth Uhealth 2017 Oct 18;5(10):e158 [FREE Full text] [doi: 10.2196/mhealth.7745] [Medline: 29046266]
22. Ahmed I, Ahmad NS, Ali S, Ali S, George A, Saleem DH, et al. Medication Adherence Apps: Review and Content Analysis. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e62 [FREE Full text] [doi: 10.2196/mhealth.6432] [Medline: 29549075]
23. Machado SM, Wilson EH, Elliott JO, Jordan K. Impact of a telemedicine eICU cart on sepsis management in a community hospital emergency department. J Telemed Telecare 2018 Apr;24(3):202-208. [doi: 10.1177/1357633X17691862] [Medline: 29278979]

24. Wu J, Ho T, Chang Y, Hsu C, Tsai CJ, Lai F, et al. Wearable-Based Mobile Health App in Gastric Cancer Patients for Postoperative Physical Activity Monitoring: Focus Group Study. JMIR Mhealth Uhealth 2019 Apr 23;7(4):e11989 [FREE Full text] [doi: 10.2196/11989] [Medline: 31012858]
25. Kummerow BK, Oyefule OO, Phillips SE, Baucom RB, Holzman MD, Sharp KW, et al. Postoperative Care Using a Secure Online Patient Portal: Changing the (Inter)Face of General Surgery. J Am Coll Surg 2015 Dec;221(6):1057-1066. [doi: 10.1016/j.jamcollsurg.2015.08.429] [Medline: 26453260]
26. Eisenberg D, Hwa K, Wren SM. Telephone follow-up by a midlevel provider after laparoscopic inguinal hernia repair instead of face-to-face clinic visit. JSLS 2015;19(1):e2014.00205 [FREE Full text] [doi: 10.4293/JSLS.2014.00205] [Medline: 25848178]
27. Canon S, Shera A, Patel A, Zamilpa I, Paddack J, Fisher PL, et al. A pilot study of telemedicine for post-operative urological care in children. J Telemed Telecare 2014 Dec;20(8):427-430 [FREE Full text] [doi: 10.1177/1357633X14555610] [Medline: 25316038]
28. Sanger PC, Hartzler A, Han SM, Armstrong CAL, Stewart MR, Lordon RJ, et al. Patient perspectives on post-discharge surgical site infections: towards a patient-centered mobile health solution. PLoS One 2014;9(12):e114016 [FREE Full text] [doi: 10.1371/journal.pone.0114016] [Medline: 25436912]
29. Sanger PC, Hartzler A, Lordon RJ, Armstrong CA, Lober WB, Evans HL, et al. A patient-centered system in a provider-centered world: challenges of incorporating post-discharge wound data into practice. J Am Med Inform Assoc 2016 May;23(3):514-525. [doi: 10.1093/jamia/ocv183] [Medline: 26977103]
30. Bangor A, Kortum P, Miller J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies 2009;4(3):114-123 [FREE Full text]
31. Meyer N, Kerz M, Folarin A, Joyce DW, Jackson R, Karr C, et al. Capturing Rest-Activity Profiles in Schizophrenia Using Wearable and Mobile Technologies: Development, Implementation, Feasibility, and Acceptability of a Remote Monitoring Platform. JMIR Mhealth Uhealth 2018 Oct 30;6(10):e188 [FREE Full text] [doi: 10.2196/mhealth.8292] [Medline: 30377146]
32. Arostegui I, Legarreta MJ, Barrio I, Esteban C, Garcia-Gutierrez S, Aguirre U, IRYSS-COPD Group. A Computer Application to Predict Adverse Events in the Short-Term Evolution of Patients With Exacerbation of Chronic Obstructive Pulmonary Disease. JMIR Med Inform 2019 Apr 17;7(2):e10773 [FREE Full text] [doi: 10.2196/10773] [Medline: 30994471]
33. Yongping L, Juan Z, Zhou P, Yongfeng Z, Liu W, Shi Y. Evaluation of the Quadri-Planes Method in Computer-Aided Diagnosis of Breast Lesions by Ultrasonography: Prospective Single-Center Study. JMIR Med Inform 2020 May 05;8(5):e18251 [FREE Full text] [doi: 10.2196/18251] [Medline: 32369039]
34. Dittrich F, Beck S, Harren AK, Reinecke F, Serong S, Jung J, et al. Analysis of Secure Apps for Daily Clinical Use by German Orthopedic Surgeons: Searching for the "Needle in a Haystack". JMIR Mhealth Uhealth 2020 May 07;8(5):e17085 [FREE Full text] [doi: 10.2196/17085] [Medline: 32379054]
35. Hwa K, Wren SM. Telehealth follow-up in lieu of postoperative clinic visit for ambulatory surgery: results of a pilot program. JAMA Surg 2013 Sep;148(9):823-827 [FREE Full text] [doi: 10.1001/jamasurg.2013.2672] [Medline: 23842982]
36. Gurjar SV, Kulkarni D, Khawaja HT. Outpatient general surgical follow-up: are we using this resource effectively? Int J Surg 2009 Feb;7(1):62-65 [FREE Full text] [doi: 10.1016/j.ijsu.2008.10.014] [Medline: 19095510]
37. Rao R, Shukla BM, Saint-Cyr M, Rao M, Teotia SS. Take two and text me in the morning: optimizing clinical time with a short messaging system. Plast Reconstr Surg 2012 Jul;130(1):44-49. [doi: 10.1097/PRS.0b013e3182547d63] [Medline: 22743872]
38. Wiseman JT, Fernandes-Taylor S, Barnes ML, Tomsejova A, Saunders RS, Kent KC. Conceptualizing smartphone use in outpatient wound assessment: patients' and caregivers' willingness to use technology. J Surg Res 2015 Sep;198(1):245-251. [doi: 10.1016/j.jss.2015.05.011] [Medline: 26025626]
39. Chen DW, Davis RW, Balentine CJ, Scott AR, Gao Y, Tapia NM, et al. Utility of routine postoperative visit after appendectomy and cholecystectomy with evaluation of mobile technology access in an urban safety net population. J Surg Res 2014 Aug;190(2):478-483. [doi: 10.1016/j.jss.2014.04.028] [Medline: 24880202]
40. Wiseman JT, Guzman AM, Fernandes-Taylor S, Engelbert TL, Saunders RS, Kent KC. General and vascular surgery readmissions: a systematic review. J Am Coll Surg 2014 Sep;219(3):552-69.e2 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2014.05.007] [Medline: 25067801]
41. Goldstein MM, Thorpe Jane H. The First Anniversary of the Health Information Technology for Economic and Clinical Health (HITECH) Act: the regulatory outlook for implementation. Perspect Health Inf Manag 2010 Sep 01;7:1c [FREE Full text] [Medline: 20808607]
42. Scott AR, Alore EA, Naik AD, Berger DH, Suliburk JW. Mixed-Methods Analysis of Factors Impacting Use of a Postoperative mHealth App. JMIR Mhealth Uhealth 2017 Feb 08;5(2):e11 [FREE Full text] [doi: 10.2196/mhealth.6728] [Medline: 28179215]
43. Zhu Z, Zhang H, Li Y, Meng X, Jiao Y, Qin Y, et al. Establishment of a New Platform for the Management of Patients After Cardiac Surgery: Descriptive Study. JMIR Med Inform 2019 Apr 01;7(2):e13123 [FREE Full text] [doi: 10.2196/13123] [Medline: 30932865]

44. Dahlberg K, Jaensson M, Nilsson U, Eriksson M, Odencrants S. Holding It Together-Patients' Perspectives on Postoperative Recovery When Using an e-Assessed Follow-Up: Qualitative Study. JMIR Mhealth Uhealth 2018 May 25;6(5):e10387 [FREE Full text] [doi: 10.2196/10387] [Medline: 29802094]
45. Jaensson M, Dahlberg K, Eriksson M, Grönlund , Nilsson U. The Development of the Recovery Assessments by Phone Points (RAPP): A Mobile Phone App for Postoperative Recovery Monitoring and Assessment. JMIR Mhealth Uhealth 2015;3(3):e86 [FREE Full text] [doi: 10.2196/mhealth.4649] [Medline: 26362403]
46. Sylvia LG, Bernstein EE, Hubbard JL, Keating L, Anderson EJ. Practical guide to measuring physical activity. J Acad Nutr Diet 2014 Feb;114(2):199-208 [FREE Full text] [doi: 10.1016/j.jand.2013.09.018] [Medline: 24290836]
47. Jacobs DR, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. Med Sci Sports Exerc 1993 Jan;25(1):81-91 [FREE Full text] [doi: 10.1249/00005768-199301000-00012] [Medline: 8423759]
48. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med 2003 Jun;37(3):197-206; discussion 206 [FREE Full text] [doi: 10.1136/bjsm.37.3.197] [Medline: 12782543]
49. Quinn CC, Staub S, Barr E, Gruber-Baldini A. Mobile Support for Older Adults and Their Caregivers: Dyad Usability Study. JMIR Aging 2019 May 23;2(1):e12276 [FREE Full text] [doi: 10.2196/12276] [Medline: 31518271]
50. Or CK, Liu K, So MKP, Cheung B, Yam LYC, Tiwari A, et al. Improving Self-Care in Patients With Coexisting Type 2 Diabetes and Hypertension by Technological Surrogate Nursing: Randomized Controlled Trial. J Med Internet Res 2020 Mar 27;22(3):e16769 [FREE Full text] [doi: 10.2196/16769] [Medline: 32217498]
51. Rosner BI, Gottlieb M, Anderson WN. Accuracy of Internet-Based Patient Self-Report of Postdischarge Health Care Utilization and Complications Following Orthopedic Procedures: Observational Cohort Study. J Med Internet Res 2018 Jul 20;20(7):e10405 [FREE Full text] [doi: 10.2196/10405] [Medline: 30030212]
52. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009 Jun;12(4):419-429. [doi: 10.1111/j.1524-4733.2008.00470.x] [Medline: 19900250]
53. Kessel KA, Vogel MM, Schmidt-Graf F, Combs SE. Mobile Apps in Oncology: A Survey on Health Care Professionals' Attitude Toward Telemedicine, mHealth, and Oncological Apps. J Med Internet Res 2016 Nov 24;18(11):e312 [FREE Full text] [doi: 10.2196/jmir.6399] [Medline: 27884810]
54. Gunter R, Fernandes-Taylor S, Mahnke A, Awoyinka L, Schroeder C, Wiseman J, et al. Evaluating Patient Usability of an Image-Based Mobile Health Platform for Postoperative Wound Monitoring. JMIR Mhealth Uhealth 2016 Sep 28;4(3):e113 [FREE Full text] [doi: 10.2196/mhealth.6023] [Medline: 27683059]
55. Fernandes-Taylor S, Gunter RL, Bennett KM, Awoyinka L, Rahman S, Greenberg CC, et al. Feasibility of Implementing a Patient-Centered Postoperative Wound Monitoring Program Using Smartphone Images: A Pilot Protocol. JMIR Res Protoc 2017 Feb 22;6(2):e26 [FREE Full text] [doi: 10.2196/resprot.6819] [Medline: 28228369]
56. Gunter RL, Fernandes-Taylor S, Rahman S, Awoyinka L, Bennett KM, Weber SM, et al. Feasibility of an Image-Based Mobile Health Protocol for Postoperative Wound Monitoring. J Am Coll Surg 2018 Mar;226(3):277-286. [doi: 10.1016/j.jamcollsurg.2017.12.013] [Medline: 29366555]
57. Lalloo C, Shah U, Birnie KA, Davies-Chalmers C, Rivera J, Stinson J, et al. Commercially Available Smartphone Apps to Support Postoperative Pain Self-Management: Scoping Review. JMIR Mhealth Uhealth 2017 Oct 23;5(10):e162 [FREE Full text] [doi: 10.2196/mhealth.8230] [Medline: 29061558]
58. Symer MM, Abelson JS, Milsom J, McClure B, Yeo HL. A Mobile Health Application to Track Patients After Gastrointestinal Surgery: Results from a Pilot Study. J Gastrointest Surg 2017 Sep;21(9):1500-1505. [doi: 10.1007/s11605-017-3482-2] [Medline: 28685388]

Abbreviations

eHealth: electronic health
mHealth: mobile health
WHO: World Health Organization

Edited by G Eysenbach; submitted 04.01.20; peer-reviewed by R Ciptaningtyas, A Zink, CS Hung, R Kaczmarczyk; comments to author 12.06.20; revised version received 16.07.20; accepted 19.07.20; published 28.08.20.

Please cite as:
Liao CH, Wu YT, Cheng CT, Ooyang CH, Kang SC, Fu CY, Hsu YP, Hsieh CH, Chen CC
An Image-Based Mobile Health App for Postdrainage Monitoring: Usability Study
J Med Internet Res 2020;22(8):e17686
URL: http://www.jmir.org/2020/8/e17686/
doi:10.2196/17686
PMID:32857060

Original Paper

Patients' and Health Care Professionals' Perceptions of the Potential of Using the Digital Diabetes Questionnaire to Prepare for Diabetes Care Meetings: Qualitative Focus Group Interview Study

Katarina Eeg-Olofsson[1,2], MD, PhD; Unn-Britt Johansson[3,4], RN, PhD; Ebba Linder[5], RN; Janeth Leksell[6,7], RN, PhD

[1]Sahlgrenska University Hospital, Gothenburg, Sweden

[2]Institute of Medicine, Department of Molecular and Clinical Medicine, University of Gothenburg, Gothenburg, Sweden

[3]Sophiahemmet University, Stockholm, Sweden

[4]Department of Clinical Sciences and Education, Karolinska Institutet, Stockholm, Sweden

[5]Center of Registers Västra Götaland, Nationella Diabetesregistret, Gothenburg, Sweden

[6]School of Education, Health and Social Studies, Dalarna University, Falun, Sweden

[7]Department of Medical Sciences, Clinical Diabetology and Metabolism, Uppsala University, Uppsala, Sweden

Corresponding Author:
Katarina Eeg-Olofsson, MD, PhD
Sahlgrenska University Hospital
Blå Stråket 5
Gothenburg, 413 45
Sweden
Phone: 46 313421000
Email: katarina.eeg-olofsson@vgregion.se

Abstract

Background: In effective diabetes management, it is important that providers and health care systems prioritize the delivery of patient-centered care and that they are respectful of and responsive to individual patient preferences and barriers.

Objective: The objective of the study was to conduct focus group interviews to capture patients' and health care professionals' perceptions and attitudes regarding digital technology and to explore how the digital Diabetes Questionnaire can be used to support patient participation in diabetes care, as a basis for an implementation study.

Methods: A qualitative study was conducted with six focus group discussions with diabetes specialist nurses and medical doctors (n=29) and four focus group discussions with individuals with diabetes (n=23). A semistructured focus group interview guide was developed, including probing questions. The data were transcribed verbatim, and qualitative content analysis was performed using an inductive approach.

Results: Two main categories were revealed by the qualitative analysis: *perceptions of digital technology and the digital questionnaire in diabetes management and care* and *perceptions of participation in diabetes care*. An overarching theme that emerged from the focus group interviews was *patients' and professionals' involvement in diabetes care using digital tools.*

Conclusions: The analysis identified important factors to consider when introducing the digital Diabetes Questionnaire in clinical use. Both professionals and patients need support and training in the practical implementation of the digital questionnaire, as well as the opportunity to provide feedback on the questionnaire answers.

(J Med Internet Res 2020;22(8):e17504) doi:10.2196/17504

KEYWORDS

Digital questionnaire; health care professionals; diabetes care; focus group interview; qualitative research; eHealth

Introduction

Background

In effective diabetes management, it is important that providers and health care systems prioritize the delivery of patient-centered care, acknowledging multiple morbidity and being respectful of and responsive to individual patients' preferences and barriers, including the differential costs of therapies [1]. Practicing patient-centered psychosocial care requires that the context of the person with diabetes be considered in communications and interactions, problem identification, psychosocial screening, diagnostic evaluation, and intervention services [2].

Importance of Patient-Reported Outcomes

In line with these requirements for patient-centered care, our research group has developed and psychometrically tested a digital patient-reported outcome measure, ie, the Diabetes Questionnaire [3-7]. Patients respond digitally to the Diabetes Questionnaire, and these responses form the basis for the patient's and the health care staff's preparation to be well-informed to discuss the patient's care and make decisions based on the patient's wishes, needs, and barriers. The purpose of the Diabetes Questionnaire is to be user-friendly and immediately provide results to the patients and health care professionals as part of the consultation. The Diabetes Questionnaire is a tool for person-centered care that creates the necessary conditions for shared, well-founded decisions regarding patients living with diabetes. A recent review by Skovlund et al [8] concluded that patient-reported outcome measures have the potential to facilitate person-centered care and active participation. The long-term goal is for the Diabetes Questionnaire to be used to capture patient-reported outcome measures in integrated diabetes care and contribute to better health care meetings, facilitate thorough follow-up over time, and be considered together with the medical variables in the National Quality Registry for Diabetes.

Challenges and Possibilities for Digital Health Tools

Digital developments in health care have created a great need for research on how digital health care services affect, for example, health care quality, design, and accessibility. Despite great interest from users, decision makers, and academics, the research base on these issues is limited [9]. The research available in this area focuses mainly on specific implementations, such as the psychotherapeutic treatment of mental illnesses or the diagnosis of skin diseases using digital tools, in more traditional telemedicine or in related areas such as online journals [10]. However, more research is needed on digital health care services, both because many conditions that are currently handled in digital health care services have not yet been investigated and because new digital tools that can change digital and physical health care services and care processes are constantly being developed and need to be evaluated separately [11,12]. In this study, we defined digital health as "the cultural transformation of how disruptive technologies that provide digital and objective data accessible to both caregivers and patients leads to an equal level doctor-patient relationship with shared decision-making and the democratization of care [13]."

Rationale

The rationale of this study was based on earlier studies and reviews [14,15] that pointed out that the basis of knowledge for conducting an implementation study of patient-reported outcome measures should be tailored by identifying and addressing potential barriers and facilitators specific to the setting. In addition, the methodological quality of existing evidence with respect to digital health interventions for chronic diseases such as diabetes is low, and the results are unpredictable [16]. As a first step, we therefore decided to conduct focus group discussions with both patients and professionals involved in diabetes care, to inform a forthcoming implementation study of the Diabetes Questionnaire in a diabetes care setting.

Aim of the Study

The purpose of this study was to conduct focus group interviews to capture patients' and health care professionals' perceptions and attitudes regarding digital technology and to explore how the digital Diabetes Questionnaire can be used to support patient participation in diabetes care, as the basis for an implementation study.

Methods

Research Design

An exploratory and descriptive qualitative design was used in this study. The data were collected through focus group discussions conducted from June 2018 to November 2018 with diabetes specialist nurses and medical doctors and with adults with type 1 diabetes or type 2 diabetes.

Focus groups have been widely used to examine persons' experiences, and this method was chosen because the focus group environment is socially oriented and may increase the sense of belonging and cohesiveness among the participants, which can lead to increased openness [17].

Sample and Setting

The participants were recruited through purposive sampling, which involves a conscious selection of individuals with the appropriate experiences or characteristics [18]; 18 hospital-based outpatient clinics and 22 primary health care clinics were initially approached. Of these, 14 hospital-based outpatient clinics and 8 primary health care clinics agreed to participate in the study.

The study was approved by the Regional Ethical Review Board in Gothenburg, Sweden (No. 317-18). A letter to the participants informed them about the study's purpose, the voluntary nature of their participation, the confidentiality measures and methods of handling of their personal data, the National Diabetes Register, contact details, and the right to end participation. All participants gave written informed consent, and the research was performed in accordance with the Declaration of Helsinki [19].

Participants and Procedure

The focus group discussions were held at hospital outpatient centers, primary health care clinics, and the Center of Registers Västra Götaland. Each group consisted of 3 to 6 participants; 6 focus group discussions were held with diabetes specialist nurses and medical doctors (health care professionals), and 4 focus group discussions were conducted with persons with type 1 diabetes or type 2 diabetes (patients). Focus group participant characteristics are presented in Table 1.

At the beginning of each focus group discussion, the participants completed a brief questionnaire that asked about their demographic characteristics (age, gender, education, occupation, age at diagnosis, and type of diabetes). The discussion opened with a general introduction of the study, an overview of the purpose of the discussion, and its confidentiality. The duration of the focus group discussions ranged from 0.6 hours to 1.5 hours. KEO or JL moderated the focus groups, and EL facilitated the discussions. All participants were given sufficient opportunity to share their views. We conducted separate focus group discussions for patients and for health care professionals. A semistructured focus group interview guide was developed according to the study aims (Multimedia Appendix 1). Furthermore, probing questions were used, for example, "Could you please further describe the situation using a concrete example?" All focus group discussions were audio-recorded with a digital voice recorder and transcribed verbatim by a medical secretary.

Table 1. Focus group participant characteristics.

Participant characteristics	Value
Patients with diabetes (4 focus groups, n=23)	
Age (years), median (range[a])	60 (22-81)
Gender, n (%)	
Women	11 (48)
Men	12 (52)
Diabetes duration (years), median (range)	21 (3-64)
Diabetes type, n (%)	
Type 1	17 (74)
Type 2	6 (26)
Health-care professionals (6 focus groups, n=29)	
Age (years), median (range)	53 (36-70)
Gender, n (%)	
Women	24 (83)
Men	5 (17)
Role, n (%)	
Diabetes specialist nurses	23 (79)
Medical doctors	6 (21)

[a]Minimum to maximum.

Data Analysis

Qualitative content analysis inspired by Krippendorff [20] was used with an inductive approach. The focus group discussions and content analysis were performed in Swedish. The research group has deep knowledge and experience in this method. The quotations presented in this paper were translated into English by a professional translator. The data analysis was manually performed as follows:

Step 1: The transcribed focus groups discussions were read through several times to obtain an overall sense of the data (KEO, U-BJ, EL, and JL). After discussion, all authors agreed that saturation had been reached.

Step 2: The transcribed text was divided into units of meaning, which were condensed and labeled with codes and discussed (KEO, U-BJ, EL, and JL). The analysis was based on a manifest interpretation of the text.

Step 3: The various codes were compared, and similarities and differences were identified. The codes were then sorted into five categories. Thereafter, the two main categories were determined by consensus among all authors.

Step 4: The analysis was based on a manifest interpretation of the text. An overarching theme was identified when all authors performed a latent interpretation of the content.

Results

The qualitative analysis identified two main categories: *perceptions of digital technology and the digital questionnaire in diabetes management and care* and *perceptions of participation in diabetes care* (Table 2). The overarching theme

that emerged from the focus group discussions was *patients' and professionals' involvement in diabetes care using digital tools*.

Table 2. Theme, main categories, and categories.

Theme	Main categories	Categories
Patients' and professionals' involvement in diabetes care using digital tools	Perceptions of digital technology and the digital questionnaire in diabetes management and care	• Hope and concern • Opportunities and obstacles • Individual needs and supportive structure
	Perceptions of participation in diabetes care	• Give and take • Trust and communication: A cornerstone for relationships

Perceptions of Digital Technology and the Digital Questionnaire in Diabetes Management and Care

Hope and Concern

In terms of their perceptions of the use of digital technology in connection with the digital Diabetes Questionnaire and diabetes management and care, participants expressed both hope and concern. Patient focus groups expressed that using digital technology provides more information to the health care professionals. Regarding hope, the health care professional focus groups expressed that this form of care (digital technology) and the questionnaire could increase the availability of information required for retaining patients in the care unit. In addition, the health care professionals said that, with digital technology and the digital questionnaire, factors such as accessibility, individualization, and closer contact with the patient could lead to broader perspectives in individual encounters in diabetes clinics. In the patient focus groups, some argued that digital technology and the digital questionnaire would simplify life by providing opportunities to compile and analyze information on different aspects of living with the disease.

> *It's a lot of technology now. And so, I think we must do this, to keep up with the younger generation because they are there. That's where they communicate. You get so much more out of it. You get so much more information with the digital, you get many angles and lots to work with. Thus, you see so much more.* [Health care professional]

Participants in both the health care professional and the patient focus groups emphasized that this form of care (digital technology) can never take the place of physical encounters but that it should be an important complement to these encounters.

However, a concern was raised that health care professionals lack sufficient skills in digital technology, and that it was these professionals perceived to be responsible for stimulating the patients' interest in this form of care. The professionals noted that the patients they meet have varying attitudes and experiences regarding using digital technology. The patient focus groups confirmed that there is a great variety of perspectives and attitudes regarding using and seeing the benefits of digital technology. Participants in both the health care professional focus groups and the patient focus groups expressed concern that a digital questionnaire would be perceived as violating the patient's privacy, and some described a feeling of "being unprotected." In addition, concerns about data confidentiality were expressed.

> *I see the possibilities in the technology. I am a technology lover myself in many respects—not least in my job. But I am terrified that, if you face this, the politicians will soon see it, and "Here, we can take and reduce the staff; here, we invest in digital technology instead.*" [Patient]

Opportunities and Obstacles

Both health care professionals and patients discussed and reflected on what opportunities the questionnaire provides in terms of improving patient care. The participants expressed that the digital questionnaire would probably create opportunities for person-focused approaches, as well as supporting an in-depth dialogue. It was also seen as clarifying caregivers' interest in learning about what it is like to live with diabetes and facilitating discussion (such as allowing patients to talk about their moods). The participants argued that the digital questionnaire provides a more complete, overall picture that highlights individual needs and provides opportunities for preparation and feedback. They talked about a "feeling curve" and a "technical curve." The digital questionnaire was seen as able to inspire support for a learning climate, which may, in turn, result in a new way of working. However, the participants also described obstacles such as the possibility that not all patients would appreciate the questionnaire and that there might be a lack of competency to follow up on the questionnaire or a lack of technical support provided by the health care professionals.

The patients expressed a lack of competency in diabetes care as potential obstacles. The professionals also expressed a lack of time as a potential obstacle. Nevertheless, the questionnaire was simultaneously seen as presenting a positive challenge because it could lead to reflection for both professionals and patients. For the professionals, it could mean a new approach and working method (improvement work or quality development). For the patients, the questionnaire could be an opportunity to talk and reflect on "What is important for me in my life with diabetes?"

Both patients and health care professionals stated that the questionnaire provides a good basis for the opportunity for monitoring and evaluation at the group level, which can create

a foundation for positive changes in health care. In addition, the questionnaire provides a basis for comparing the results between health care units at the national level.

> *Nevertheless, I believe that it gives the patient the opportunity to reflect, when filling this in. … Therefore, that is a message we send with our questions actually…. And then I think it helps us to be able to sometimes, instead of talking blood sugar curves, get a little insight into what is behind this disease.* [Health care professional]

> *If we take this data in, we can't just drop it there. We must find some structure for it too. And how to get this information, which can be a little worrying.* [Health care professional]

Individual Needs and a Supportive Structure

Both patients and health care professionals' discussions identified two supportive structures as prerequisites for implementing the digital Diabetes Questionnaire. The first structure is the professionals' need for an introduction and training before the start of the implementation. The second supportive structure is that the organization (eg, hospital, primary care clinic) provide infrastructure and other complementary resources to support the work and development of the diabetes care teams.

Education and training were not only requested at the beginning of the implementation but also as a continuous intervention. The health care professionals called for a clearer structure in their work to create well-functioning routines. The importance of support from the head of the department was emphasized to ensure that time and resources would be made available.

> *Getting enough time for the invitation for the diabetes care meeting, interpreting the answers, preparing for the meeting and for the actual meeting itself. Time is needed to structure the work before and create routines. And of course, support from the management.* [Health care professional]

Perceptions of Participation in Diabetes Care

Give and Take

Primarily the health care professionals expressed the idea that patient participation in diabetes care is a requirement for change. They suggested that patient participation means that the health care professionals give support and make space for the patients to take responsibility for their own disease.

The patient focus groups also expressed that participation in their own diabetes care is a prerequisite for diabetes management and that the patient should be considered the principal actor in this process of providing the information about living with diabetes. Furthermore, the patients expressed the importance of having the right to decide for themselves and to be involved in the health care decisions based on their individual needs, which require both self-confidence and courage. They argued that there are different types of involvement, ie, participation in diabetes care, participation in diabetes care together with the health care professionals, and participation and involvement in the diabetes disease itself.

Furthermore, the professionals expressed the importance of the digital questionnaire in communicating the health care professionals interest in the patients' responses to the questionnaire. The questionnaire signals that the diabetes health care professionals are interested in the patients' life with diabetes. The patients expressed a similar idea in terms of participation being facilitated by the structure of the care meeting and that the questionnaire potentially increasing participation through structure and planning.

> *It must be those [patients] who control what actions they take. We are all consultants and informers, but participation is after all the Alpha and Omega.* [Health care professional]

> *It is a prerequisite for something to happen at all, that they are involved in it. Otherwise it doesn't happen that much.* [Health care professional]

> *Used correctly, the questionnaire can probably increase participation. If it is, then there are enough questions about what support I need, and I express it. And that it is taken care of when resources are given to it; then it will increase my… or the care of my illness. This can't be just a box-ticking affair, because that won't increase the participation in care. Without it being like, you have a plan for how to proceed in these questions. And that it will be a continuation.* [Patient]

Trust and Communication: A Cornerstone for Relationships

The health care professionals' focus groups expressed that trust and interaction between patients and health care professionals are the foundation for participation and a prerequisite for change. The patients expressed continuity as one cornerstone to build trust in their relationships with health care professionals. Another cornerstone for building these relationships is the interactive communication between patients and health care professionals. A prerequisite for interactive communication is speaking the same language and using the same vocabulary:

> *No, I think the patients may feel that they are being seen differently. That judgment ... than just coming to the visit, if they have now filled in for example this in at home, for example, before, and you bring it together, then maybe it becomes even clearer that you start from their needs, than when you just meet, I don't know.* [Health care professional]

> *After all, it is a difference if you are to participate in an operation that you do once in your life, or if you are to participate in an illness that you will have for the rest of your life. I see it in a different way really, it is that you are the main person responsible for yourself and you should seek participation from others as support.* [Patient]

Discussion

Principal Findings

The main finding from the focus group interviews was the theme of patients and health care professionals' involvement in diabetes care using digital tools. The main category of perceptions of digital technology and the digital questionnaire in diabetes management was built on the categories of hope and concern, opportunities and obstacles, and individual needs and supportive structure. The other main category, perceptions of participation in diabetes care, was built on categories of give and take, and trust and communication: a cornerstone for relationships. Although, the participants were not aware of the content of the Diabetes Questionnaire, they nevertheless expressed many views, thoughts, and feelings during the focus group interviews.

During the analysis of the focus group interviews, it became clear that the patients and health care professionals expressed similar expectations and reflections regarding both digital technology and the Diabetes Questionnaire. In particular, these similarities were seen to be prominent when the participants discussed digital technology in health care. Concerns such as a lack of knowledge about digital technology and a lack of privacy were evident. A previous study [21] has shown that the acceptance of digital technology relies on understanding of patients' fears and concerns about lack of security. Lupton [22] highlighted the importance of reducing fears or concerns regarding insecurity among both patients and health care professionals through implementing secure computer systems and protecting personal data. Consequently, patients and health care professionals must both be aware of the security systems that surround digital technology in health care. Additionally, if these security systems are presented in a transparent and pedagogical way, knowledge of digital technology will increase and these concerns will decrease. Optimistically, in our study, the participants expressed hope that using digital technology can contribute to facilitating and supporting the everyday lives of people with diabetes. This finding is in line with previous studies [23-25] indicating that digital technologies can enhance diabetes care in a real-world setting.

When we analyzed the participants' perceptions of the digital Diabetes Questionnaire, we found both differences and similarities between patients and health care professionals. The health care professionals saw an opportunity in that the Diabetes Questionnaire could facilitate person-centered care. They used phrases such as "an informed patient," the opportunity for "person-centered work," and "honest answers" from patients living with diabetes. The patients saw an opportunity for using the digital Diabetes Questionnaire to talk about and reflect on important factors in their everyday lives that influence their ability to take care of their diabetes. These results were in line with a newly published review showing that patient-reported outcome measures can facilitate a meaningful and focused conversation during the clinical encounter [8]. Conversely, patients in our study raised a concern about a lack of competency among health care professionals in terms of dealing with patients' responses to the questionnaire. Furthermore, the participants emphasized the importance of a supportive structure. This highlights the importance of health care professionals receiving training on how to use the digital Diabetes Questionnaire and on how to sufficiently address patients' responses to the questionnaire [26].

Furthermore, Greenhalgh et al [27] pointed out that the ways in which clinicians use patient-reported outcome measures are shaped by their relationships with patients. We found similar results under the category of trust and communication: a cornerstone for relationships. In addition, the participants reported that interactive communication between patients and professionals is important for building relationships. A prerequisite for interactive communication is speaking the same language and using the same vocabulary. There are recommendations for language used by health care professionals and others when discussing diabetes through spoken or written words [28]. An obstacle noted in our study, as well as in others, is that health care professionals lack time in everyday clinical care [29]. Both participant groups expressed that this lack of time could negatively affect the use of the digital Diabetes Questionnaire.

Perceptions of participation were discussed in the focus groups. The participants emphasized the importance of give and take in each care meeting for achieving participation. In addition, they reflected on whether participation meant that the patients become involved in their care or whether the health care professional must become involved in the patients' daily life. The participants talked about adding the "feeling curve" from the digital Diabetes Questionnaire to the already existing "technological curve" to present a more holistic picture of the patients' daily lives. The importance of participation and shared decision making in diabetes care and its association with optimal self-management has been shown in an earlier study [30]. Nevertheless, the participants in our study asserted that participation in diabetes care is suboptimal and perceived as an unclear concept.

Our focus group interview analysis showed that the participants expressed a positive attitude toward using the digital Diabetes Questionnaire. Discussing and reflecting on the questionnaire responses during physical encounters can improve diabetes care based on the person's experience of living with diabetes and facilitate adequate support and self-management in a structured way. In addition, it is important to confirm the patient in a positive way. The response of the questionnaire clarifies what resources the patients have that could enable optimal daily diabetes self-management.

The analysis identified important factors to consider when introducing the digital Diabetes Questionnaire in clinical use. Both professionals and patients need support and training in the practical implementation of this intervention, and they should have an opportunity to provide feedback on the questionnaire answers. Öberg et al [31] pointed out that targeted training could increase the digital skills used in diabetes care. These needs are also confirmed by the participants' discussion of the factor of insufficient time and the concern that patients' responses to the questionnaire should be treated with great integrity. A review [14] in palliative care settings pointed out that providing an

educational component prior to the implementation is crucial, in addition to interviewing health care professionals and patients.

Limitations

Our sample may consist of participants with an interest in digital technology and in the development of diabetes care, which may have introduced bias. Thus, a sample using patients with low health literacy may show different results. Poor health literacy can seriously impair people's interactions with health care professionals and their potential to benefit from digital health services. Qualitative studies are difficult to generalize because of small sample sizes. However, the results, especially on the perceptions of digital technology, are important for other groups of patients and for health care professionals in other disciplines.

Implications

This study identified important factors to consider when introducing the digital Diabetes Questionnaire in clinical use and will serve as a basis for continued work in a larger implementation study for the Diabetes Questionnaire. When introducing digital patient-reported outcome measures such as the digital Diabetes Questionnaire, it is important to consider what conditions exist. A prerequisite is ensuring that both patients and health care staff members can handle digital technology. A second prerequisite is the careful evaluation of the content of the intended patient-reported outcome measure questions. Finally, a trusting relationship between health care professionals and patients is required to make conversations based on the digital patient-reported outcome measure feel meaningful.

Acknowledgments

This work was supported by the Health and Welfare research profile, Dalarna University, Sophiahemmet Foundation, The Swedish Diabetes Foundation, and the Swedish state under the agreement between the Swedish government and the county councils, the ALF agreement (ALFGBG-698991). The authors thank all the participants, both patients and health care professionals, for sharing their experiences and opinions. We also thank Jennifer Barrett, PhD, from Edanz Group for editing a draft of this manuscript.

Authors' Contributions

All authors made substantial contributions to the design of the study, to the analysis and interpretation of the data, and in writing and revising the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group interview guide.
[DOCX File , 15 KB - jmir_v22i8e17504_app1.docx]

References

1. Davies MJ, D'Alessio DA, Fradkin J, Kernan WN, Mathieu C, Mingrone G, et al. Management of hyperglycaemia in type 2 diabetes, 2018. a consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia 2018 Dec;61(12):2461-2498. [doi: 10.1007/s00125-018-4729-5] [Medline: 30288571]
2. Young-Hyman D, de GM, Hill-Briggs F, Gonzalez JS, Hood K, Peyrot M. Psychosocial care for people with diabetes: a position statement of the American Diabetes Association. Diabetes Care 2016 Dec;39(12):2126-2140. [doi: 10.2337/dc16-2053] [Medline: 27879358]
3. Borg S, Eeg-Olofsson K, Palaszewski B, Svedbo Engström M, Gerdtham U, Gudbjörnsdottir S. Patient-reported outcome and experience measures for diabetes: development of scale models, differences between patient groups and relationships with cardiovascular and diabetes complication risk factors, in a combined registry and survey study in Sweden. BMJ Open 2019 Jan 04;9(1):e025033 [FREE Full text] [doi: 10.1136/bmjopen-2018-025033] [Medline: 30612113]
4. Borg S, Palaszewski B, Gerdtham U, Fredrik O, Roos P, Gudbjörnsdottir S. Patient-reported outcome measures and risk factors in a quality registry: a basis for more patient-centered diabetes care in Sweden. Int J Environ Res Public Health 2014 Nov 26;11(12):12223-12246 [FREE Full text] [doi: 10.3390/ijerph111212223] [Medline: 25431875]
5. Svedbo Engström M, Leksell J, Johansson U, Borg S, Palaszewski B, Franzén S, et al. Health-related quality of life and glycaemic control among adults with type 1 and type 2 diabetes - a nationwide cross-sectional study. Health Qual Life Outcomes 2019 Aug 14;17(1):141 [FREE Full text] [doi: 10.1186/s12955-019-1212-z] [Medline: 31412881]
6. Svedbo Engström M, Leksell J, Johansson U, Eeg-Olofsson K, Borg S, Palaszewski B, et al. A disease-specific questionnaire for measuring patient-reported outcomes and experiences in the Swedish National Diabetes Register: development and evaluation of content validity, face validity, and test-retest reliability. Patient Educ Couns 2018 Jan;101(1):139-146. [doi: 10.1016/j.pec.2017.07.016] [Medline: 28736071]

7. Svedbo Engström M, Leksell J, Johansson U, Gudbjörnsdottir S. What is important for you? a qualitative interview study of living with diabetes and experiences of diabetes care to establish a basis for a tailored Patient-Reported Outcome Measure for the Swedish National Diabetes Register. BMJ Open 2016 Mar 24;6(3):e010249 [FREE Full text] [doi: 10.1136/bmjopen-2015-010249] [Medline: 27013595]
8. Skovlund SE, Lichtenberg TH, Hessler D, Ejskjaer N. Can the routine use of patient-reported outcome measures improve the delivery of person-centered diabetes care? a review of recent developments and a case study. Curr Diab Rep 2019 Aug 16;19(9):84. [doi: 10.1007/s11892-019-1190-x] [Medline: 31420754]
9. Jandoo T. WHO guidance for digital health: what it means for researchers. Digit Health 2020;6:2055207619898984 [FREE Full text] [doi: 10.1177/2055207619898984] [Medline: 31949918]
10. Cahn A, Akirov A, Raz I. Digital health technology and diabetes management. J Diabetes 2018 Jan;10(1):10-17. [doi: 10.1111/1753-0407.12606] [Medline: 28872765]
11. Baltaxe E, Czypionka T, Kraus M, Reiss M, Askildsen JE, Grenkovic R, et al. Digital health transformation of integrated care in Europe: overarching analysis of 17 integrated care programs. J Med Internet Res 2019 Sep 26;21(9):e14956 [FREE Full text] [doi: 10.2196/14956] [Medline: 31573914]
12. Konttila J, Siira H, Kyngäs H, Lahtinen M, Elo S, Kääriäinen M, et al. Healthcare professionals' competence in digitalisation: a systematic review. J Clin Nurs 2019 Mar;28(5-6):745-761. [doi: 10.1111/jocn.14710] [Medline: 30376199]
13. Meskó B, Drobni Z, Bényei E, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017;3:38. [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
14. Antunes B, Harding R, Higginson IJ, EUROIMPACT. Implementing patient-reported outcome measures in palliative care clinical practice: a systematic review of facilitators and barriers. Palliat Med 2014 Feb;28(2):158-175. [doi: 10.1177/0269216313491619] [Medline: 23801463]
15. Stover AM, Tompkins Stricker C, Hammelef K, Henson S, Carr P, Jansen J, et al. Using stakeholder engagement to overcome barriers to implementing patient-reported outcomes (PROs) in cancer care delivery: approaches from 3 prospective studies. Med Care 2019 May;57 Suppl 5 Suppl 1:S92-S99. [doi: 10.1097/MLR.0000000000001103] [Medline: 30985602]
16. Bashi N, Fatehi F, Mosadeghi-Nik M, Askari MS, Karunanithi M. Digital health interventions for chronic diseases: a scoping review of evaluation frameworks. BMJ Health Care Inform 2020 Mar;27(1):e100066 [FREE Full text] [doi: 10.1136/bmjhci-2019-100066] [Medline: 32156751]
17. Krueger R, Casey M. Focus Groups: A practical guide for applied research. 5th Revised edition ed. Thousand Oaks, United States: SAGE Publications, Inc; 2014.
18. Patton M. Qualitative Research & Evaluation Methods - Integrating Theory and Practice. Thousand Oaks (CA), United states: SAGE Publications Ltd; 2015.
19. World MA. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
20. Krippendorff K. Content Analysis: An introduction to Its Methodology. Thousand Oaks (CA), United States: SAGE Publications, Inc; 2018.
21. Nymberg VM, Bolmsjö BB, Wolff M, Calling S, Gerward S, Sandberg M. 'Having to learn this so late in our lives…' Swedish elderly patients' beliefs, experiences, attitudes and expectations of e-health in primary health care. Scand J Prim Health Care 2019 Mar;37(1):41-52 [FREE Full text] [doi: 10.1080/02813432.2019.1570612] [Medline: 30732519]
22. Lupton D. Digital health now and in the future: findings from a participatory design stakeholder workshop. DIGITAL HEALTH 2017 Nov 09;3:1-17. [doi: 10.1177/2055207617740018]
23. Bakhach M, Reid MW, Pyatak EA, Berget C, Cain C, Thomas JF, et al. Home telemedicine (CoYoT1 Clinic): a novel approach to improve psychosocial outcomes in young adults with diabetes. Diabetes Educ 2019 Aug;45(4):420-430. [doi: 10.1177/0145721719858080] [Medline: 31244396]
24. Marcolino MS, Maia JX, Alkmim MBM, Boersma E, Ribeiro AL. Telemedicine application in the care of diabetes patients: systematic review and meta-analysis. PLoS One 2013;8(11):e79246 [FREE Full text] [doi: 10.1371/journal.pone.0079246] [Medline: 24250826]
25. Wang G, Zhang Z, Feng Y, Sun L, Xiao X, Wang G, et al. Telemedicine in the management of type 2 diabetes mellitus. Am J Med Sci 2017 Jan;353(1):1-5. [doi: 10.1016/j.amjms.2016.10.008] [Medline: 28104096]
26. Haverman L, van Oers H, Limperg P, Hijmans P, Schepers S, Sint NS. Implementation of electronic patient reported outcomes in pediatric daily clinical practice: the KLIK experience. Clin Prac in Pediatr Psychol 2014;2(1):50-67 [FREE Full text] [doi: 10.1037/cpp0000043]
27. Greenhalgh J, Gooding K, Gibbons E, Dalkin S, Wright J, Valderas J, et al. How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? a realist synthesis. J Patient Rep Outcomes 2018 Dec;2:42 [FREE Full text] [doi: 10.1186/s41687-018-0061-6] [Medline: 30294712]
28. Dickinson JK, Guzman SJ, Maryniuk MD, O'Brian CA, Kadohiro JK, Jackson RA, et al. The use of language in diabetes care and education. Diabetes Educ 2017 Dec;43(6):551-564. [doi: 10.1177/0145721717735535] [Medline: 29040034]
29. West E, Barron DN, Reeves R. Overcoming the barriers to patient-centred care: time, tools and training. J Clin Nurs 2005 Apr;14(4):435-443 [FREE Full text] [doi: 10.1111/j.1365-2702.2004.01091.x] [Medline: 15807750]

30. Rodriguez-Gutierrez R, Gionfriddo MR, Ospina NS, Maraka S, Tamhane S, Montori VM, et al. Shared decision making in endocrinology: present and future directions. Lancet Diabetes Endocrinol 2016 Aug;4(8):706-716. [doi: 10.1016/S2213-8587(15)00468-4] [Medline: 26915314]
31. Öberg U, Isaksson U, Jutterström L, Orre CJ, Hörnsten A. Perceptions of persons with type 2 diabetes treated in Swedish primary health care: qualitative study on using ehealth services for self-management support. JMIR Diabetes 2018 Mar 12;3(1):e7 [FREE Full text] [doi: 10.2196/diabetes.9059] [Medline: 30291075]

Edited by G Eysenbach; submitted 17.12.19; peer-reviewed by J Farzi, L Quintana; comments to author 10.03.20; revised version received 15.04.20; accepted 13.05.20; published 19.08.20.

Please cite as:
Eeg-Olofsson K, Johansson UB, Linder E, Leksell J
Patients' and Health Care Professionals' Perceptions of the Potential of Using the Digital Diabetes Questionnaire to Prepare for Diabetes Care Meetings: Qualitative Focus Group Interview Study
J Med Internet Res 2020;22(8):e17504
URL: https://www.jmir.org/2020/8/e17504
doi:10.2196/17504
PMID:32812884

Original Paper

Factors Influencing the Smartphone Usage Behavior of Pedestrians: Observational Study on “Spanish Smombies”

César Fernández[1], PhD; María Asunción Vicente[1], PhD; Irene Carrillo[2,3], PhD; Mercedes Guilabert[2], PhD; José Joaquín Mira[2,4], PhD

[1]Telematics Engineering Area, Miguel Hernández University, Elche, Spain

[2]Health Psychology Department, Miguel Hernández University, Elche, Spain

[3]Foundation for the Promotion of Health and Biomedical Research of Valencia Region, Alicante, Spain

[4]Alicante-Sant Joan Health District, Alicante, Spain

Corresponding Author:
María Asunción Vicente, PhD
Telematics Engineering Area
Miguel Hernández University
Avenida de la Universidad s/n
Edificio Innova
Elche, 03202
Spain
Phone: 34 966658385
Email: suni@umh.es

Abstract

Background: Smartphone addiction has become a reality accepted by all. Some previous studies have shown that the use of smartphones on public roads while walking is very common among the young population. The term “smombie” or smartphone zombie has been coined for this behavior. Such behavior causes a reduction in the attention given to other pedestrians and drivers and may result in accidents or collisions. However, there are no precise data about how many people use the phone while they are walking on the street. Smartphone usage habits are evolving rapidly, and more in-depth information is required, particularly about how users interact with their devices while walking: traditional phone conversations (phone close to the ear), voice chats (phone in front of the head), waiting for notifications (phone in hand), text chats (user touching the screen), etc. This in-depth information may be useful for carrying out specific preventive actions in both the education field (raising awareness about the risks) and in the infrastructure field (redesigning the cities to increase safety).

Objective: This study aimed to gather information about pedestrians’ smartphone usage and to identify population groups wherein interventions should be focused to prevent accidents. The main hypothesis was that gender, age, and city area can significantly influence the smartphone usage of the pedestrians while walking.

Methods: An observational study of pedestrians in the street was carried out in Elche, a medium-sized Spanish city of 230,000 inhabitants. The following data were gathered: gender, age group, location, and type of smartphone interaction. A specific smartphone app was developed to acquire data with high reliability. The statistical significance of each variable was evaluated using chi-squared tests, and Cramér’s V statistic was used to measure the effect sizes. Observer agreement was checked by the Cohen kappa analysis.

Results: The behavior of 3301 pedestrians was analyzed, of which 1770 (53.6%) were females. As expected, the effect of the main variables studied was statistically significant, although with a small effect size: gender (P<.001, V=0.12), age (P<.001, V=0.18), and city area (P<.001, V=0.16). The phone in hand or “holding” behavior was particularly dependent on gender for all age groups (P<.001, V=0.09) and to a greater extent in young people (P<.001, V=0.16). Approximately 39.7% (222/559) of the young women observed showed “holding” or “smombie” behavior, and they comprised the highest proportion among all age and gender groups.

Conclusions: An in-depth analysis of smartphone usage while walking revealed that certain population groups (especially young women) have a high risk of being involved in accidents due to smartphone usage. Interventions aimed at reducing the risk of falls and collisions should be focused in these groups.

(J Med Internet Res 2020;22(8):e19350) doi:10.2196/19350

KEYWORDS

smartphone addiction; smartphone overuse; smombies; pedestrian safety; mobile phone

Introduction

Use of Mobile Devices and Smartphones

Since the emergence of mobile technology, the use of mobile devices and services has continued to increase progressively and at different rates in both developed and developing countries [1]. By the end of 2018, more than 5 billion people around the world subscribed to mobile services, accounting for 67% of the global population, and this figure is expected to reach 71% by 2025 [2]. In Spain, the penetration rate of mobile devices is 98%, with most of them (80%) being smartphones [1]. In both cases, the penetration rate in Spain is above the European average (85% and 72%, respectively) [2]. There are differences in the penetration rate of smartphones by age; 95% of the Spaniards younger than 35 years own a smartphone, while only 60% of those older than 50 years own a smartphone—a trend that is widespread worldwide [1]. Further, 3.6 billion people are connected to the mobile internet, with 67% of the global connections occurring through smartphones [2,3].

These data show the globalized presence of mobile devices, which is inevitably linked to their increasing use in time and place. In 2018, users around the world spent an average of 800 hours per year on their smartphones. In Europe, the average time spent on smartphones is 3 hours daily and between 14 and 43 hours weekly. The most common activities among Europeans include (in this order) emailing, social networking, instant messaging, searching, reading, and gaming [4]. Given these high levels of dedication, it is not surprising that the use of smartphones overlaps with the execution of other activities (multitasking) with variable attention requirements such as watching television, eating, dressing, working, or walking, including frequent checks and alternating periods of activity to attend to possible notifications [5].

Problematic Smartphone Use

Despite the many benefits of smartphones, their unlimited use can lead to what is known as problematic smartphone use, which is related to the discomfort associated with "unsubstantiated or behavioral addictions" (eg, anxiety when the device is not accessible) [6]. According to a recent meta-analysis by Sohn et al [7] which included 41,871 children and young people, the median prevalence of problematic smartphone use was 23.3%. Age (17-19 years) and female gender were the risk factors for the development of problematic smartphone use, although in the case of the latter, the results are not conclusive. Problematic smartphone use is associated with higher odds of experiencing depression, anxiety, perceived stress, and a decrease in sleep quality. One of the problematic uses can arise, for example, when crossing a street; this is a complex exercise with a relatively high demand for perceptive and cognitive capacity. Even for those pedestrians who can successfully integrate the required information under normal circumstances, the distraction of holding a smartphone can interfere with the decision-making process at many points. Pedestrians may be unaware of important auditory or visual information, make incorrect judgments about speed (especially when multiple lanes or vehicles are involved), incorrectly attribute driver intent, or misjudge their ability to cross in a given gap. Distraction, therefore, has the potential to exacerbate the risk of a collision for pedestrians [8]. In a study [9] that aimed to explore the effect of gender on the use of smartphones while walking, a modest gender bias was observed, with walking behavior with the smartphone more frequent among women than men. Another effect observed was that when couples of the opposite sex walked together, the use of the smartphone was decreased. Some of the reasons for people to walk with smartphones in their hands could be the social pressure to be available, security concerns (reduced risk of theft), psychological dependence (anxiety over separation from the smartphone), or for display as a status symbol.

Smombies

The term "smombie" (smartphone-zombie) [10,11] or "phone walker" [9] has been coined as a result of increasingly frequent behavior involving the use of smartphones while walking on public roads [9]. This concept refers to the pedestrian who uses a smartphone while walking, with the physical or cognitive consequences that this type of behavior may have. The effects on physical health are the carrying out of behaviors that may endanger the pedestrian or other people who are circulating at that moment—mainly the lack of safety [12,13]. The other effects may be directly related to the way one walks and one's direct involvement in a traffic accident compared to people who do not use their smartphones while walking [13]. At the cognitive level, the lack of attention while walking and using the smartphone implies a lack of recognition of the other pedestrians, lower cognitive capacity, and greater attention deficit [14]. It is arguable whether smombie behavior represents a form of problematic smartphone use in itself or not. However, there is no doubt that this behavior represents a safety risk. It seems that this phenomenon is increasingly being studied in different countries and contexts [15,16] and evidences on the effects and consequences of this pattern of behavior are increasing [17,18], because of which this study was carried out. The goal of this study was to gather the information that helps us measure and understand the smombie behavior and to identify specific groups that may require special attention to reduce the risk of accidents.

Methods

Study Design

An observational study of the behavior of pedestrians with their smartphones was carried out in Elche, Spain, by a multidisciplinary research group composed mainly of behavioral scientists and smartphone engineers from Miguel Hernández University. It was executed from April 2019 until November

2019 as a project called “CountingSmombies.” This study was registered and validated by the ethics committee of the Miguel Hernández University with the research code COIR:AUT.DISP.CFP.01.19. Direct measurements by an observer was the method used to register the behavior of the pedestrians with their smartphones while they were walking on the street. The behavior of the pedestrians with their smartphones was categorized into the following 5 classes, which is ordered from lower to higher use of the smartphone.

1. NOT VISIBLE: The pedestrian does not visibly carry or use his/her smartphone.
2. TALKING: The pedestrian is talking on the phone in the traditional way, that is, the smartphone is held in the hand and close to the ear and mouth of the talking subject.
3. HEADPHONES: The pedestrian is wearing headphones visibly and these are supposed to be connected to a smartphone.
4. HOLDING: The pedestrian holds the smartphone in one of his/her hands while walking but is not looking directly at it.
5. SMOMBIE: The pedestrian holds the smartphone in his/her hand while walking and interacts with the screen by either staring at it or typing or talking toward the screen when the audio is sent or during videoconferencing.

These 5 categories were selected according to that reported in recent studies [19,20] and by considering 2 main factors: first, they were easily distinguishable by the observers and second, they represented different attitudes toward smartphone usage while walking. The ordering below reflects, what we considered, an increasing risk of behavioral addiction or attention loss:

1. Talking in a traditional way was considered as less invasive smartphone usage, as only one ear is involved.
2. Using headphones was associated with higher attention loss, with both ears involved.
3. Holding the smartphone in our hands, although apparently not causing attention loss, may reflect a psychological dependence, a need to be aware of incoming notifications, and can be part of an alternation of smombie-holding periods. That is why we considered such behavior almost on top of the problematic smartphone use list.

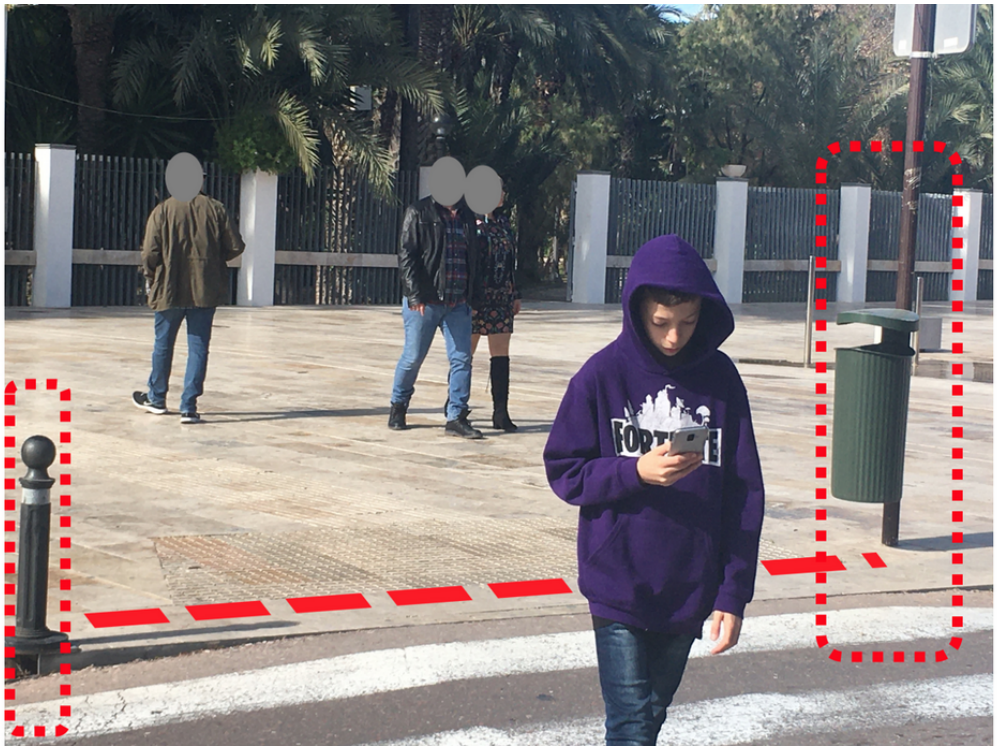
Figure 1 illustrates the 5 types of smartphone behaviors observed in street pedestrians. The observer also registered the gender and the approximate age of the pedestrian according to the perceived appearance (the interobserver agreement was validated through a Cohen kappa analysis in a simultaneous session with 2 observers and with N=100). Only 2 classes were used in the gender category: male and female. For the age category, 4 classes were used: (1) below 18 years (10-18 years), teenagers; (2) 18-35 years, young people; (3) 35-65 years, adults; (4) over 65 years, older people. Data were stored by using a quick annotation app that allowed saving the data of each of the performed experiments.

Figure 1. The 5 types of smartphone usage behaviors observed in street pedestrians. From left to right (lower to higher use of smartphone): smartphone "not visible," "talking" on a smartphone, using "headphones," "holding" a smartphone, and "smombie".

Pedestrian Observation Procedure

A pedestrian observation procedure was designed with 2 modalities: sitting observer and moving (walking) observer. Sitting observers stayed at a specific location and registered pedestrians in their field of vision. Moving observers walked through a predetermined path and registered pedestrians walking in the opposite direction. When the observer was walking, it was easy to avoid biases in the selection of the pedestrians; only those who randomly cross with the observer were recorded. When the observer was sitting, a stricter protocol was needed; there were many pedestrians around and only some of them may catch the observer's attention. According to our protocol, only those pedestrians who pass through a predefined crossing line were recorded. The crossing line was defined between 2 points on the street, for example, 2 trees, 2 bollards, or the sides of a shop window (Figure 2). Besides, the pedestrians walking through the crossing line could walk in either direction; therefore, a predefined direction was established and only pedestrians walking in such a direction were recorded.

Figure 2. Defining a visual crossing line between 2 points on the street, for example, the bollard and the street litter bin, to not introduce bias and to count only the pedestrians moving in one of the two directions between the crossing line.



A pedestrian eligibility list was given to the observers to homogenize their recordings (Textbox 1). Pedestrians were included if they were walking alone or in a group. However, groups of over 4 people were excluded to avoid counting errors. It was decided to include people who run but not those who use vehicles such as bicycles and wheel or electric scooters. People in wheelchairs (and their porters) and those who push a baby carriage were included. A common behavior that was observed among the pedestrians was that they took their smartphones from their pockets or purses and stopped or even sat or leaned on an urban structure to use the smartphone more slowly. When this occurred during an observational experiment, this subject was not counted, as this situation fell outside the established categories (ie, the subject was not walking).

Pedestrians who used a wristband or watch were also excluded because it was difficult to assess their category; the observer cannot distinguish whether they were wearing a traditional watch or a smartwatch. In the first case, the pedestrian may just be looking at the time; in the second case, the pedestrian may be accessing a secondary smartphone screen. Globally, the eligibility list introduced a bias in the measures: the actual proportions of the smombies may have been higher than that accounted for (eg, smartwatches are excluded). However, our goal was to establish very clear criteria so that different observers could obtain homogeneous measures.

The pedestrian could simultaneously present 2 established behaviors (or 2 mixed categories), for example, using headphones and looking at the smartphone screen. In these cases, the observer should select the behavior where the problematic use of smartphone is higher. Figure 3 shows 2 possible mixed categories: using headphones and holding the smartphone simultaneously was categorized as "holding" class and using headphones and interacting with the smartphone was categorized as "smombie" class.

Textbox 1. Eligibility criteria for the pedestrians.

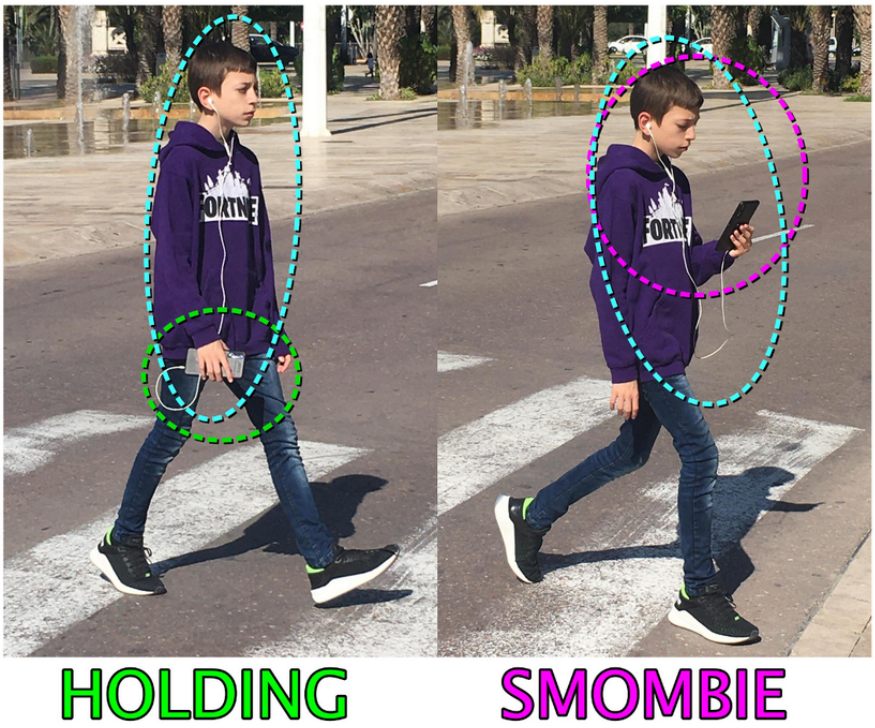
Pedestrians included in the observational data

- people walking alone or in a group of 4 people or less
- people running or jogging
- people in wheelchairs and their companions
- people walking with their dogs
- people carrying baby carriages

Pedestrians excluded from the observational data

- groups of more than 4 people
- bike or scooter travelers
- those who stopped walking and stood while using the smartphone
- people who were doing a job (eg, postmen, carriers, gardeners, waiters)
- those who checked a wristband or wristwatch
- any other situation not considered above

Figure 3. In pedestrians with mixed behavior, higher problematic smartphone usage was selected as the pedestrian behavior.



Observation Sessions

Data were registered during 43 observation sessions. An observation session requires the selection of a placement (sitting observer) or path (moving observer) as well as the selection of the date and time for the recording. The session duration was not predetermined, and the number of pedestrians registered per session were not predetermined. The quick annotation app allowed continuous recording without limits in time or number of registers.

Concerning date and time, observation sessions were performed from July 2019 to November 2019 during working days and during rush hours (except for the *Pokemon Go Community Day*, an extra observation session). Two periods of the day were considered: midday rush hour (noon to 2 PM) and evening rush hour (7 PM to 9:30 PM). According to the Spanish schedule, since most people leave work or school during these periods, there are plenty of people and activities on the street.

An extra observation session was performed on Saturday, October 12, 2019 on the *Pokemon Go Community Day* [21], a worldwide monthly event wherein *Pokemon Go* players get together to look for special game items (specific *Pokemons* that appear with high frequency at certain time lapses and city areas). This resulted in a notable increase in the number of *Pokemon Go* players on the street and consequently, in the number of

smombies observed. The goal was to analyze the differences with the remaining sessions.

Concerning the location, all the observation sessions were performed in Elche (medium-sized Spanish city, 230,000 inhabitants), where 3 different scenarios were selected: city center, residential areas (with large avenues frequented by runners and hikers), and different areas of the University campus. These scenarios covered a wide variety of city inhabitants and situations such as people walking to or from work in the city center, people in their leisure time in residential areas, and students in the University campus. The special observation session corresponding to the *Pokemon Go Community Day* was performed in the city center.

Quick Annotation App

The main screen of the quick annotation app is shown in Figure 4. The purpose of using the app was to be able to quickly and efficiently save the measurement made by the observer on a spreadsheet. Acquiring data was fast. It just required 3 taps per subject to register: gender button, age button, and behavior button. The selected options light up in green for a short time to provide visible feedback to the observer. When an observational session ends, simply by clicking on the "Finish Experiment" button, the data were saved on a spreadsheet and stored for further processing and analysis.

Figure 4. A screenshot of the quick annotation app. Data were stored in the csv format and processed in Matlab and R software.

Statistical Methods

Each observation session was stored in csv files, which were processed with Matlab (version R2019a). All statistical analyses were performed using R software (version 3.6.2). All variables were considered categorical (including age, which was grouped into 4 age ranges). According to this, the chi-square test was used to evaluate the statistical significance of each variable, while effect sizes were measured using Cramér's V statistic. Observer agreement was checked by the Cohen kappa analysis.

Results

Descriptive Statistics

A total of 3301 pedestrians were registered, of which 1770 (53.6%) pedestrians were women. According to the data from this observational study, 29.7% (982/3301) of the observed pedestrians were walking using a smartphone ("talking", "headphones", "holding", and "smombies" classes) during rush hours on the working days in this city. The descriptive statistics of the study are shown in Table 1. To validate the study tool (clearly exclusive categories and the quick annotation app performance), 2 judges (sitting observers) performed a Cohen kappa analysis in a simultaneous session with a sample population of 100. In all categories, the degree of acceptance of the judges was high: gender of the pedestrians (=1; P<.001), age group (=0.703; P<.001), and behavior with the smartphone (=0.953; P<.001). Multimedia Appendix 1 shows the confusion matrices obtained in the Cohen kappa analysis for each category of the study.

Table 1. Demographic characteristics of the observed pedestrians in the study (N=3301).

Characteristics	Value
Gender, n (%)	
Male	1531 (46.4)
Female	1770 (53.6)
Age (years), n (%)	
Teenagers (10-18 years)	260 (7.9)
Young people (18-35 years)	1179 (35.7)
Adults (35-65 years)	1338 (40.5)
Older people (>65 years)	524 (15.9)
Smartphone use, n (%)	
Not visible	2319 (70.3)
Talking	141 (4.3)
Headphones	99 (3.0)
Holding	381 (11.5)
Smombie	361 (10.9)
Scenarios, n (%)	
City center	2158 (65.4)
Residential areas	623 (18.9)
University campus	520 (15.7)

Gender Influence

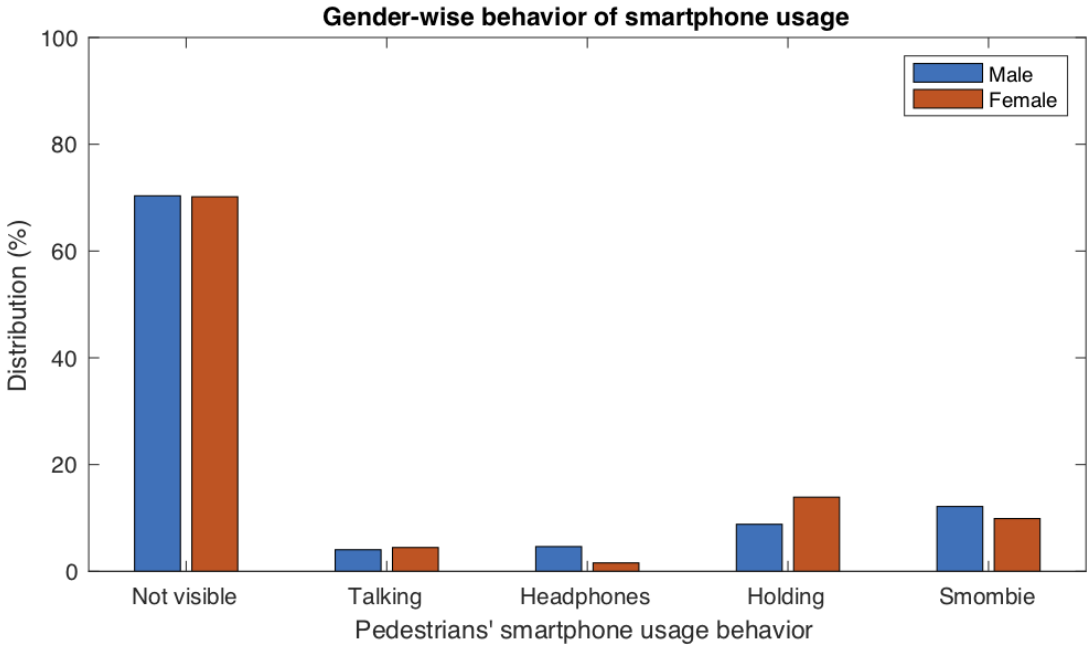
First, we analyzed the influence of the pedestrians' gender on their smartphone usage behavior (Table 2). Overall, the influence was found to be statistically significant (P<.001), although with a small effect size (Cramér's V=0.12).

Table 2 shows the complete distribution of the data collected according to gender, as well as additional results: the "Headphones" and "Holding" behaviors were particularly influenced by gender. Women showed the "Holding" behavior (P<.001, V=0.09) to a greater extent, while men showed the "Headphones" behavior (P<.001, V=0.08) to a greater extent. For a better understanding of the results, Figure 5 shows the differences in the smartphone usage behavior according to gender.

Table 2. Influence of the pedestrians' gender on smartphone usage behavior.

Behavior	Females, n=1770, n (%)	Males, n=1531, n (%)	*P* value	Cramér's V
Not visible	1242 (70.2)	1077 (70.4)	.94	0.00
Talking	79 (4.4)	62 (4.1)	.62	0.01
Headphones	28 (1.6)	71 (4.6)	<.001	0.09
Holding	246 (13.9)	135(8.8)	<.001	0.08
Smombie	175 (9.9)	186 (12.1)	<.001	0.04

Figure 5. Smartphone usage behavior of pedestrians across different genders. This figure shows the observational results (%) from Table 2.



Age Influence

Second, we analyzed the influence of the user's age on their smartphone usage behavior. Overall, a statistically significant influence was also detected (P<.001), although the effect size was small (Cramér's V=0.18). Table 3 shows the complete distribution according to age and additional results, which are consistent with the expectations. In this case, the behaviors most correlated with age were "Not visible" (ie, the pedestrian was not using the smartphone at all) (P<.001, V=0.30) and "Smombie" (P<.001, V=0.22). In the first case, the results showed that the use of the smartphone while walking was inversely related to age, and in the second case, an opposite behavior was observed.

Table 3. Influence of age on smartphone usage behavior in pedestrians.

Behavior	Teenagers[a], n=260, n (%)	Young people[b], n=1179, n (%)	Adults[c], n=1338, n (%)	Older people[d], n=524, n (%)	*P* value	Cramér's V
Not visible	157 (60.4)	647 (54.9)	1037 (77.5)	478 (91.2)	<.001	0.30
Talking	6 (2.3)	67 (5.7)	58 (4.3)	10 (1.9)	.001	0.07
Headphones	16 (6.2)	55 (4.7)	24 (1.8)	4 (0.8)	<.001	0.10
Holding	43 (16.5)	182 (15.4)	133 (10.0)	23 (4.4)	<.001	0.13
Smombie	38 (14.6)	228 (19.3)	86 (6.4)	9 (1.7)	<.001	0.22

[a]P<.001; Cramér's V=0.09.

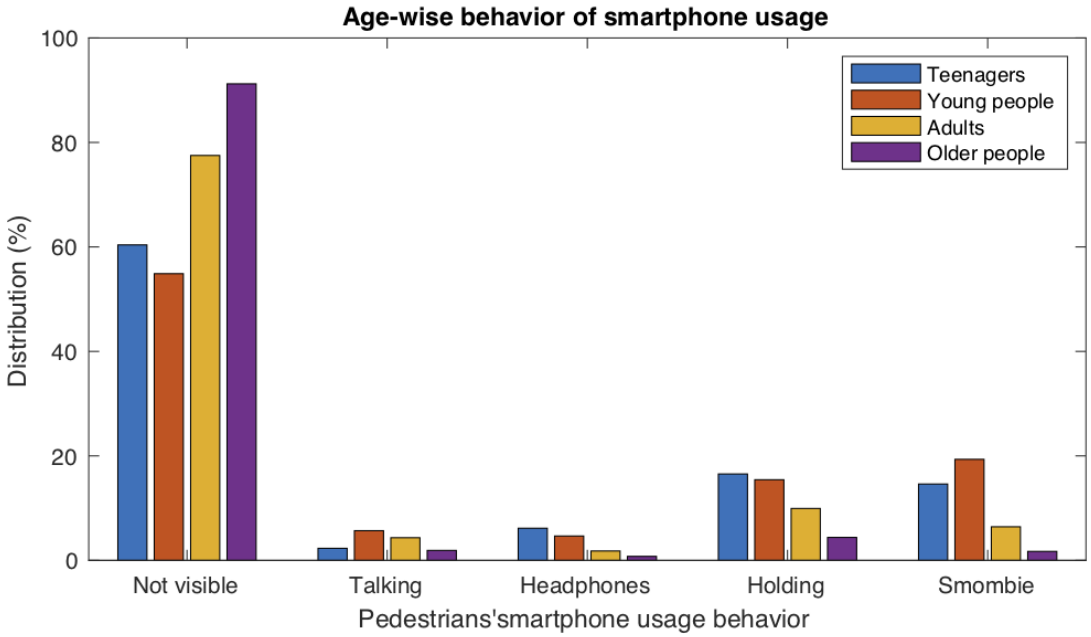
[b]P<.001; Cramér's V=0.26.

[c]P<.001; Cramér's V=0.15.

[d]P<.001; Cramér's V=0.20.

Adolescents and, to a greater extent, young people were the most likely to be in the "Smombie" category, while the older people were the least likely. Regarding the age ranges in which the pedestrians showed a different behavior from the average, the behavior of the young people (P<.001, V=0.26) and the older people (P<.001, V=0.20) was notable. Young pedestrians showed the highest values in the "smombie" and "talking" categories, while old pedestrians showed the highest value in the "not visible" category. Figure 6 shows the differences in the smartphone usage behavior according to age.

Figure 6. Smartphone usage behavior of pedestrians across different age ranges. This figure shows the observational results (%) from Table 3.



Zone Influence

We also analyzed whether different behaviors were associated with different zones or city areas (city center, residential areas, and University campus). Globally, the zone was statistically significant for behavior (P<.001), with a relatively small effect size (Cramér's V=0.16). Table 4 shows the complete distribution and all the analyses performed. The behaviors more dependent on the zone were "Not visible vs other behaviors" (P<.001, V=0.20) and "smombie vs other behaviors" (P<.001, V=0.18). The zone that was clearly different from the others was the University campus (P<.001, V=0.22), where the median age of the pedestrians was different from that of the pedestrians in other city areas. In the University campus, the proportion of "smombies" was as high as 24.0% (125/520) and the proportion of people walking without using the phone at all was as low as 48.5% (252/520), which were very different from the data in other city areas. Figure 7 shows the differences in the smartphone usage behavior of the pedestrians according to the city area.

Table 4. Influence of the city area on pedestrian smartphone usage behavior.

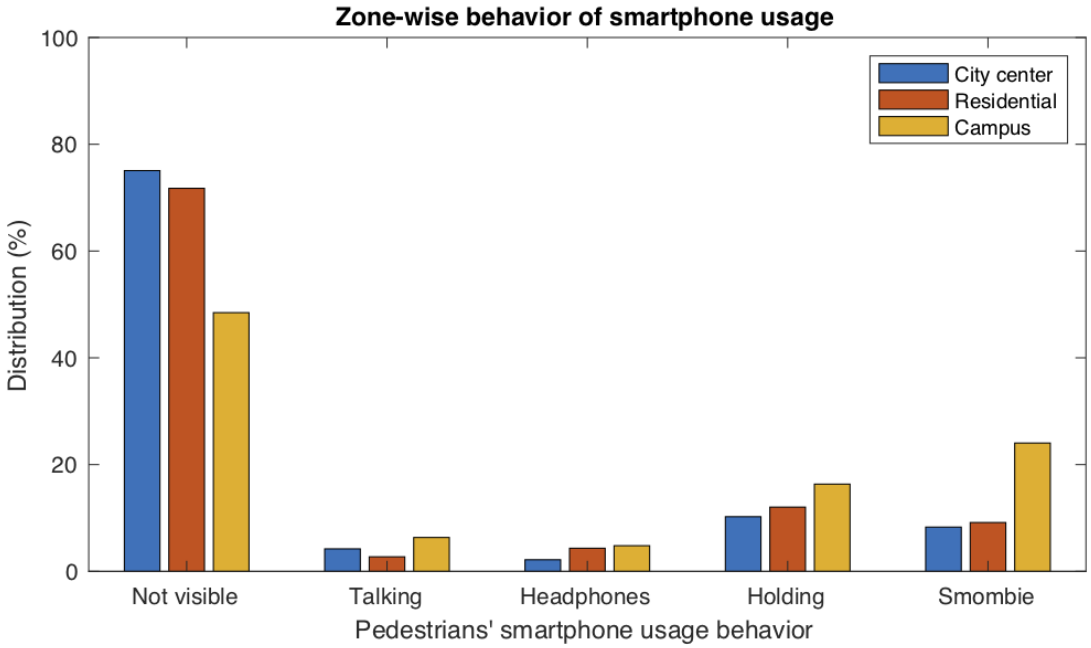
Behavior	Campus,[a] n=520, n (%)	City center,[b] n=2158, n (%)	Residential,[c] n=623, n (%)	*P* value	Cramér's V
Not visible	252 (48.5)	1620 (75.1)	447 (71.8)	<.001	0.21
Talking	33 (6.4)	91 (4.2)	17 (2.7)	.01	0.05
Headphones	25 (4.8)	47 (2.2)	27 (4.3)	<.001	0.07
Holding	85 (16.3)	221 (10.2)	75 (12.0)	<.001	0.07
Smombie	125 (24.0)	179 (8.3)	57 (9.2)	<.001	0.18

[a]P<.001; Cramér's V: 0.22.

[b]P<.001; Cramér's V: 0.16.

[c]P=.02; Cramér's V: 0.06.

Figure 7. Smartphone usage behavior of pedestrians in different areas of a city. This figure shows the observational results (%) from Table 4.



Gender and Age Influence

A deep analysis of the results allowed us to determine in which age range the behavior was most affected by gender. Table 5, Table 6, Table 7, and Table 8 show the grouped data, and we observed that the greatest differences in the behavior according to gender occurred in the young age range (P<.001, V=0.19). An additional analysis was carried out to specify the influence of gender on smartphone usage behavior in young people. The effect of gender on each of the behaviors (not visible, talking, headphones, holding, smombie) was studied for all age ranges. In particular, for the young age range (Table 6), the behavior most affected by gender was the "holding" behavior, which was much more common among women (P<.001, V=0.16).

Figures 8-9 show graphically the combined effects of gender and age on smartphone usage behavior.

Table 5. Influence of gender on smartphone usage behavior in teenagers (n=260, globally P=.12, V=0.17).

Behavior	Females, n=163, n (%)	Males, n=97, n (%)	*P* value	Cramér's V
Not visible	93 (57.1)	64 (66)	.19	0.080
Talking	6 (3.7)	0 (0)	.14	0.092
Headphones	8 (4.9)	8 (8)	.41	0.051
Holding	31 (19.0)	12 (12)	.22	0.076
Smombie	25 (15.3)	13 (13)	.81	0.015

Table 6. Influence of gender on smartphone usage behavior in young people (n=1179, globally P<.001, V=0.19).

Behavior	Females, n=559, n (%)	Males, n=620, n (%)	*P* value	Cramér's V
Not visible	293 (52.4)	354 (57.1)	.12	0.045
Talking	33 (5.9)	34 (5.5)	.85	0.005
Headphones	11 (2.0)	44 (7.1)	<.001	0.117
Holding	120 (21.5)	62 (10.0)	<.001	0.156
Smombie	102 (18.2)	126 (20.3)	.41	0.024

Table 7. Influence of gender on smartphone usage behavior in adults (n=1338, globally *P*=.11, V=0.074).

Behavior	Females, n=783, n (%)	Males, n=555, n (%)	*P* value	Cramér's V
Not visible	607 (77.5)	430 (77.5)	>.99	0.000
Talking	35 (4.5)	23 (4.1)	.88	0.004
Headphones	9 (1.1)	15 (2.7)	.06	0.052
Holding	86 (11.0)	47 (8.5)	.15	0.039
Smombie	46 (5.9)	40 (7.2)	.39	0.024

Table 8. Influence of gender on smartphone usage behavior in older people (n=524, globally *P*=.07, V=0.13).

Behavior	Females, n=265, n (%)	Males, n=259, n (%)	*P* value	Cramér's V
Not visible	249 (94.0)	229 (88.4)	.04	0.091
Talking	5 (1.9)	5 (1.9)	>.99	0.000
Headphones	0 (0)	4 (1.6)	.13	0.067
Holding	9 (3.4)	14 (5.4)	.36	0.040
Smombie	2 (0.7)	7 (2.7)	.17	0.060

Figure 8. Smartphone usage behavior of males across different age ranges.

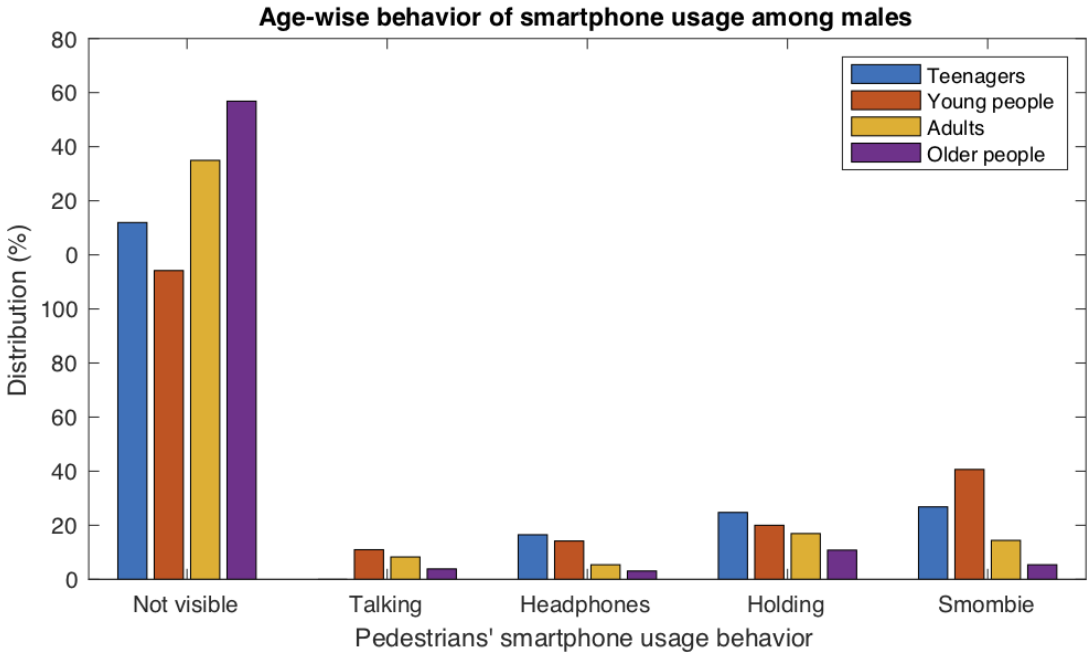
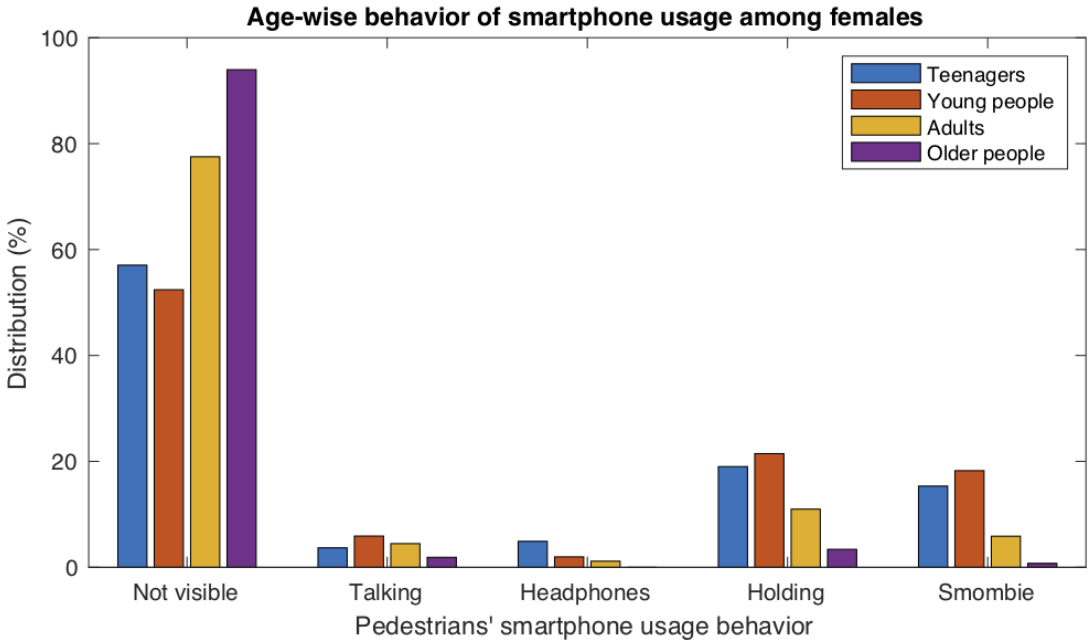


Figure 9. Smartphone usage behavior of females across different age ranges.



Effect of Pokemon Go Community Day

Our results showed that smombie behavior was clearly more frequent during the *Pokemon Go Community Day*. Table 9 shows the data distribution for this particular observation session. The proportion of smombies was 15.5% (21/136) on the *Pokemon Go Community Day* compared to 10.9% (361/3301) on normal days. Among teenagers, the proportion of smombies reached an impressive 61% (8/13) on the *Pokemon Go Community Day* compared to 14.6% (38/260) on normal days. This effect was also relevant in the young age range wherein the proportion of smombies reached 39% (12/31) during the *Pokemon Go Community Day* compared to 19.3% (228/1179) on the normal days. As expected, the chi-squared test showed that the effect was only statistically significant for teenagers (P<.001, V=0.244) and young people (P=.01, V=0.07).

Table 9. Demographic characteristics of the observed pedestrians on the Pokemon Go Community Day (N=136).

Characteristics	Value
Gender, n (%)	
Male	60 (44.1)
Female	76 (55.9)
Age (years), n (%)	
Teenagers (10-18 years)	13 (9.6)
Young people (18-35 years)	31 (22.8)
Adults (35-65 years)	65 (47.8)
Older people (more than 65 years)	27 (19.9)
Smartphone use, n (%)	
Not visible	105 (77.2)
Talking	5 (3.7)
Headphones	1 (0.7)
Holding	4 (2.9)
Smombie	21 (15.5)
Scenarios, n (%)	
City center	136 (100)
Residential areas	0 (0)
University campus	0 (0)

Table 10 shows the global influence of *Pokemon Go CommunityDay* on the smartphone usage behavior.

Tables 11-14 show the influence of the *Pokemon Go CommunityDay* on the smartphone usage behavior in individuals of each age range. For a clearer interpretation, Figure 10 shows graphically the same results.

Table 10. Influence of the Pokemon Go Community Day on the smartphone usage behavior (globally *P*=.006, V=0.065).

Behavior	Normal day, N=3301, n (%)	Pokemon day, N=136, n (%)	*P* value	Cramér's V
Not visible	2319 (70.3)	105 (77.2)	.10	0.028
Talking	141 (4.3)	5 (3.7)	.90	0.002
Headphones	99 (3.0)	1 (0.7)	.20	0.021
Holding	381 (11.5)	4 (2.9)	.003	0.051
Smombie	361 (10.9)	21 (15.5)	.13	0.025

Figure 10. Smartphone usage behavior of pedestrians across different age ranges on the Pokemon Go Community Day. This figure shows graphically the observational results (%) from Tables 11-14.

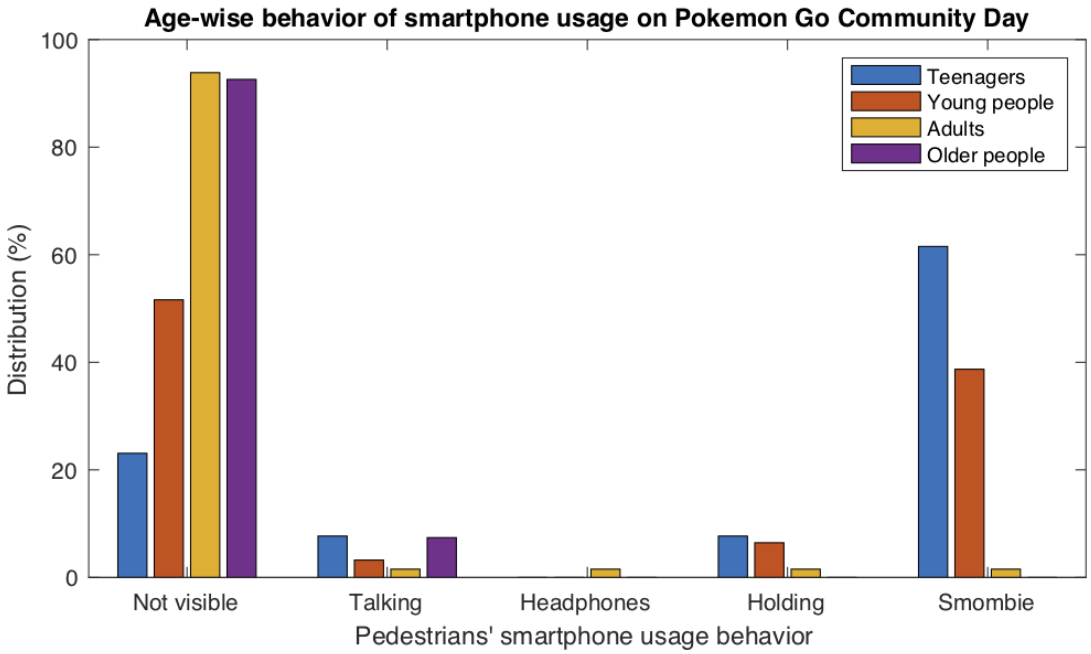


Table 11. Influence of the Pokemon Go Community Day on the smartphone usage behavior in teenagers (globally *P*<.001, V=0.283).

Behavior	Normal day, n=260, n (%)	Pokemon day, n=13, n (%)	*P* value	Cramér's V
Not visible	157 (60.4)	3 (23)	.02	0.144
Talking	6 (2.3)	1 (8)	.76	0.018
Headphones	16 (6.2)	0 (0)	.75	0.019
Holding	43 (16.5)	1 (8)	.64	0.028
Smombie	38 (14.6)	8 (61)	<.001	0.244

Table 12. Influence of the Pokemon Go Community Day on the smartphone usage behavior in young people (globally *P*=.06, V=0.087)

Behavior	Normal day, n=1179, n (%)	Pokemon day, n=31, n (%)	*P* value	Cramér's V
Not visible	647 (54.9)	16 (52)	.86	0.005
Talking	67 (5.7)	1 (3)	.85	0.005
Headphones	55 (4.7)	0 (0)	.43	0.023
Holding	182 (15.4)	2 (6)	.26	0.032
Smombie	228 (19.3)	12 (39)	.01	0.070

Table 13. Influence of the Pokemon Go Community Day on the smartphone usage behavior in adults (globally *P*=.04, V=0.085).

Behavior	Normal day, n=1338, n (%)	Pokemon day, n=65, n (%)	*P* value	Cramér's V
Not visible	1037 (77.5)	61 (94)	.003	0.079
Talking	58 (4.3)	1 (1)	.43	0.021
Headphones	24 (1.8)	1 (1)	>.99	0
Holding	133 (10.0)	1 (1)	.04	0.054
Smombie	86 (6.4)	1 (1)	.18	0.036

Table 14. Influence of the Pokemon Go Community Day on the smartphone usage behavior in older people (globally *P*=.25, V=0.099).

Behavior	Normal day, n=524, n (%)	Pokemon day, n=27, n (%)	*P* value	Cramér's V
Not visible	478 (91.2)	25 (93)	>.99	0
Talking	10 (1.9)	2 (7)	.22	0.053
Headphones	4 (0.8)	0 (0)	>.99	0
Holding	23 (4.4)	0 (0)	.54	0.026
Smombie	9 (1.7)	0 (0)	>.99	0

Discussion

Principal Results

According to the results obtained, the incidence of smartphone usage among pedestrians was high, with almost one-third of the observed pedestrians belonging to "talking," "headphones," "holding," or "smombie" categories. All these behaviors represent serious attention loss while walking. The most extreme situation, "smombie," was observed in 1 of every 10 pedestrians. Regarding age groups, the data clearly show that young people are more likely to have smombie-like behavior. Considering gender, almost half of the young women observed showed "holding" or "smombie" behavior, and they comprised the highest proportion among all the age and gender groups.

The use of smartphones while walking on the street, including smombie behavior, should be analyzed from a cognitive perspective. It is known that the human information processing capacity is limited but learning and practice make it possible to automate many of the usual daily behaviors. These behaviors are characterized because they are developed with extensive practice, performed smoothly and efficiently, are resistant to modification, "unaffected" by other activities, do not interfere with other activities, and do not require mental effort [22]. A clear example of this behavioral automation is walking.

The automation of actions such as walking releases attentional resources that can be used to perform other tasks simultaneously. Human beings can divide their attention between different tasks simultaneously and execute them successfully as long as the attentional demands of these tasks do not exceed their attentional capacity. Otherwise, the cognitive system is overloaded and the performance decreases [23]. In practice, this implies that the processing of stimuli and the emission of responses can be done both automatically through a process of memory retrieval and in a conscious and controlled way [22]. These 2 mechanisms are activated for habitual behavior depending on the level of mastery of the task and the situation. If we are walking down the street, we do not require full awareness of how we should move our feet, legs, or arms and we can perform the walking behavior while thinking about our next task or watching the traffic. However, when a new stimulus arises, the behavior moves from automatic control to a more conscious level where we make some decision, for example, to stop. Attention can be consciously and voluntarily controlled and focused on a particular stimulus, but it can also be unconsciously captured by an external stimulus such as a loud noise [23]. The latter is what happens when a smombie stops on the road when he hears the horn of a car that is about to hit him.

The immersion of technology in people's daily lives has enhanced the multitasking operation mode. The concept of media multitasking is defined as "engaging in one medium along with other media or nonmedia activities" [24]. Smombie behavior could fit within this definition. Multitasking implies the absence of total automation of tasks [23]. While it is true that walking and typing on the smartphone are fully automated activities for much of the population, identifying and avoiding obstacles or crafting meaningful messages requires active and controlled information processing. Studies on the influence of multitasking media in the educational environment show a significant reduction in student performance [23]. These results applied to the smombie phenomenon would explain the slowing down of walking, erratic wandering, and the increased risk of accidents (falls, running over, etc) [25].

The reduction of performance in multitasking media situations can not only be explained from the cognitive approach but should also be considered from the postphenomenological perspective that puts the focus on embodied habits and technical mediation in body-technology interaction [23]. For example, driving is more affected when talking on a hands-free phone than when talking to a codriver [26]. Similarly, the comprehension of information received during a lecture is greater when taking longhand notes than when writing on a computer. This is explained by the acquired modes of interaction with technology. Thus, when we write on computer keyboards,

we tend to transcribe the words heard automatically, while when we take notes with pencil and paper, we are forced to reinterpret and synthesize the information perceived [23,27].

In summary, although behaving like a smombie sometimes goes well, it entails significant risks that, in addition to reducing performance in tasks that are carried out simultaneously, put the physical integrity of the individual at risk. Decreased attention, reduced peripheral vision when looking at the smartphone, and the activation of embodied habits in the interaction with technology are the main factors underlying the risk associated with the smombie phenomenon.

According to our results, the "holding" behavior was particularly common among young women. Further analyses are needed to clarify the reasons for this effect. Possible reasons include the lack of pockets in clothes, the social pressure to be available, theft prevention, psychological dependence, or for display of the smartphones as status symbols. We are currently designing survey-based experiments to gain insight into this effect.

When gender effect was analyzed specifically for each age range, it turned out that the higher difference between male and female behavior was found in old and young age ranges, while in teenagers and adults, the effect was not statistically significant. The main reason for gender differences in the old age range (older than 65 years) can be found in the highly differentiated roles of Spanish men and women in such a population. Concerning the young age range, the main reason may be due to the different use each gender makes of smartphones—young women are more attracted to social networks, while young men are more attracted to video games. Social networks urge the user to be ready for notifications and this may justify the high proportion of young women showing the "holding" behavior.

The extremely high proportion of smombies among teenagers during the *Pokemon Go Community Days* (8/13, 61%) suggests the need for specific interventions on such days. As *Pokemon Go* players usually concentrate on certain city areas, traffic and pedestrian crossings should be specially monitored in these areas.

Comparison With Previous Studies

Previous observational studies have shown quantitative values of the number of pedestrians who use the smartphone while walking. In 2005, the study of Bungum et al [28] in Las Vegas, Nevada, already confirmed that 5.7% of the observed pedestrians (N=866) crossed the street while wearing headphones or while conversing on the phone. Some recent studies such as the one carried out in 2013 in Seattle, Washington, found that 29.8% of the observed pedestrians (N=1102) showed a distracting activity such as talking on the phone, texting, or listening to music [29].

An observational study in Paris in 2018 on the concept of "phone walkers" [20] surprisingly found that there were more female than male "phone walkers" (33.3% females, 19.7% males; N=3038). The statistical data of Schaposnik and Unwin [20], which are higher than those observed in our study, may be due to the mean age of the observed pedestrians who were younger since their estimated mean age was 35 years and they did not even consider people older than 65 years. Besides, their observational sessions were performed both on working days as well as weekends.

The observational data of the work of Ropaka et al [19] in Athens, Greece, in 2019 with a sample population of 2280 people is similar to our findings but with lower problematic smartphone use values: 83.4% of the pedestrians were classified as nondistracted (comparable to our "not visible" category with 70.3%), 5.0% were classified as distracted talking (comparable to our "talking" category with 4.3%), 5.4% as distracted listening to music (comparable to our "headphones" category with 3%), and 6.2% as distracted texting (comparable to our "smombie" category with 10.9%). In the study of Ropaka et al [19], they did not consider the "holding" behavior; further, the age ranges considered in their study were similar to that reported in our study, and the data were extracted from video analysis unlike our observation sessions in real time.

The obvious physical risk to which smombies are exposed is a problem with a solution that can be multifactorial and complex. Some cities have already taken structural measures to address this problem, such as the installation of a system of beacons at the edge of the pavement that function as a traffic light for pedestrians in Spain and Germany, the creation of phone lanes for smombies in China, the broadcasting of verbal messages on the railway in Hong Kong, or the writing of warning messages on the road ("Stop-Look-Cross. Answer Later" or "Heads Up, Phones Down") [30-32]. However, the positive effects of some of these interventions seem to be diluted after the novelty period [33]. There are also technology initiatives such as Smombie Guardian, which is a smartphone app that uses the device's camera to detect obstacles and alerts smombies through a red border and a vibration first and then an image of the obstacle with a colored border to prevent potential collisions [34].

Limitations

The categorization of pedestrian behavior is not universal. In our study, we considered 5 categories of people with their smartphones: smartphone not visible, talking on smartphone, using headphones, holding a smartphone, and smombie. A different categorization may yield different results. Although previous studies by other researchers used a variety of different categorizations [13,35-41], this categorization was based on recent studies by Ropaka et al [19] and Schaposnik and Unwin [20]. Our study adopted Ropaka's categories (not visible, talking, headphones, and smombie) and we added an extra category from the study of Schaposnik and Unwin (ie, holding or "phone walker"). Combined behaviors were not accounted for, for example, smombies who also wear headphones; such behaviors may cause a higher attention loss and may need specific consideration. Behaviors were ordered according to the increasing problematic smartphone use. Our ordering of problematic smartphone use is not universal and is possibly arguable. However, a different ordering would not affect the statistical results obtained. As stated in the Methods section, pedestrians checking their watches are not accounted for, because it was impossible to determine whether they were wearing a watch or a smartwatch. Not accounting for smartwatch smombies introduced a bias in the measures; the actual

proportion of the smombies may be higher than that accounted for. Age categorization did not reflect the exact age of each pedestrian. Besides, observers could make wrong estimations of the age ranges. However, the kappa analysis performed comparing the age estimations of 2 different observers yielded correct results (=0.703; P<.001). Perhaps, the method used for data collection (the quick annotation app) can be somewhat dystopian because when the observer was walking as the pedestrians do, he or she was just a smombie counting other smombies. Therefore, the typical distractions of the smartphone (notifications and calls) together with having to coordinate their movements with the observation of pedestrians could produce errors in data collection. Data were gathered in a particular Spanish city from April to November 2019 only during weekdays (except for the *Pokemon Go Community Day*) and only during rush hours; therefore, the results obtained are restricted to such circumstances and may not extrapolate correctly to reflect global pedestrian smartphone usage. For example, the results may have been different during the winter months or in cities with a different culture or wealth status. Given the quick evolution of smartphone technology and user behavior, this study represents a snapshot of pedestrians' smartphone usage in late 2019. Future studies are required to verify the results and analyze the trends in this topic. On the *Pokemon Go Community Day*, the study only covered 136 pedestrians; therefore, more data should be registered to confirm our results.

Future Research

A modified version of the quick annotation app used for the experiment is currently under development. The goal is to perform quick surveys and gather extra information. We are interested in determining why the "holding" behavior is increasingly common, particularly among young women. The survey will offer several answers to the question: why are you carrying your smartphone in your hand? Future work also includes an observational study of indoor walking while using the smartphone. The goal is to analyze the behavior of employees with their smartphones when they are moving around in office environments through the aisles and stairs.

Implications

Different actions can be carried out. First, an extra educational effort needs to be taken to raise the awareness about the risks of using the smartphone while walking. Second, cities need to be redesigned, thereby making them safer for smartphone users, by creating specific lanes and adding visual and sound signals in street crossings. Third, smartphones should be developed with prevention tools. Among these prevention tools, the simplest ones may just warn their users when the device is being used while the smartphone sensors detect the walking activity. More complex tools can alert the user when an obstacle is detected and a collision is imminent. Finally, other tools may be capable of blocking highly distractive apps while the user is walking.

Conclusions

The incidence of smartphone usage among pedestrians is high. Our study registered almost one-third of the pedestrians interacting with the smartphone in different ways, and more than 1 of each 10 pedestrians behaving as a smombie. According to the data gathered, the groups of greatest risk and, therefore, the groups that the interventions should be directed to, are the groups of adolescents and young people.

Besides, this study offers quantitative data about an increasingly common behavior with the smartphone—holding it while walking or, in other words, keeping it in the hand to immediately respond to any notification. According to the data collected, this behavior is more common in females, particularly among female adolescents and young women.

Acknowledgments

Aurora Mula, José Gómez, and César Junior assisted in many phases of this study. This work was partially financed by APPANDABOUT S.L from the Science Park of the Miguel Hernández University. In addition, this study is part of the research of the Excellence Groups Promeeteo/2017/173 financed by the Conselleria de Educación, Generalitat Valenciana.

Authors' Contributions

CF and MAV conceived the study. CF, IC, MAV, and MG participated in its design. CF and MAV designed the quick annotation app and performed the statistical analysis. CF, IC, JJM, MAV, and MG coordinated the qualitative research. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Confusion matrices obtained in the Cohen kappa analysis for each category of the study.
[DOCX File , 16 KB - jmir_v22i8e19350_app1.docx]

References

1. Silver L. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. Pew Research Center. 2019 Feb 05. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2020-01-14]
2. The Mobile Economy 2019. SM Association (GSMA). URL: https://www.gsmaintelligence.com/research/?file=b9a6e6202ee1d5f787cfebb95d3639c5&download [accessed 2020-01-14]
3. Informe ditrendia: Mobile en España y en el Mundo 2019. Ditrendia. URL: https://ditrendia.es/informe-mobile-espana-mundo-2019/ [accessed 2020-01-14]
4. Lopez-Fernandez O, Kuss DJ, Romo L, Morvan Y, Kern L, Graziani P, et al. Self-reported dependence on mobile phones in young adults: A European cross-cultural empirical survey. J Behav Addict 2017 Jun 01;6(2):168-177 [FREE Full text] [doi: 10.1556/2006.6.2017.020] [Medline: 28425777]
5. Toh SH, Howie EK, Coenen P, Straker LM. "From the moment I wake up I will use it…every day, very hour": a qualitative study on the patterns of adolescents' mobile touch screen device use from adolescent and parent perspectives. BMC Pediatr 2019 Jan 24;19(1):30 [FREE Full text] [doi: 10.1186/s12887-019-1399-5] [Medline: 30678720]
6. Bianchi A, Phillips JG. Psychological predictors of problem mobile phone use. Cyberpsychol Behav 2005 Feb;8(1):39-51. [doi: 10.1089/cpb.2005.8.39] [Medline: 15738692]
7. Sohn S, Rees P, Wildridge B, Kalk NJ, Carter B. Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people: a systematic review, meta-analysis and GRADE of the evidence. BMC Psychiatry 2019 Nov 29;19(1):356 [FREE Full text] [doi: 10.1186/s12888-019-2350-x] [Medline: 31779637]
8. Lennon A, Oviedo-Trespalacios O, Matthews S. Pedestrian self-reported use of smart phones: Positive attitudes and high exposure influence intentions to cross the road while distracted. Accid Anal Prev 2017 Jan;98:338-347. [doi: 10.1016/j.aap.2016.10.028] [Medline: 27825043]
9. Schaposnik LP, Unwin J. The phone walkers: a study of human dependence on inactive mobile devices. Behav 2018;155(5):389-414. [doi: 10.1163/1568539x-00003496]
10. Appel M, Krisch N, Stein J, Weber S. Smartphone zombies! Pedestrians' distracted walking as a function of their fear of missing out. Journal of Environmental Psychology 2019 Jun;63:130-133. [doi: 10.1016/j.jenvp.2019.04.003]
11. Duke E, Montag C. Smartphone addiction, daily interruptions and self-reported productivity. Addict Behav Rep 2017 Dec;6:90-95 [FREE Full text] [doi: 10.1016/j.abrep.2017.07.002] [Medline: 29450241]
12. Schwebel DC, Stavrinos D, Byington KW, Davis T, O'Neal EE, de Jong D. Distraction and pedestrian safety: how talking on the phone, texting, and listening to music impact crossing the street. Accid Anal Prev 2012 Mar;45:266-271 [FREE Full text] [doi: 10.1016/j.aap.2011.07.011] [Medline: 22269509]
13. Chen P, Pai C. Pedestrian smartphone overuse and inattentional blindness: an observational study in Taipei, Taiwan. BMC Public Health 2018 Dec 31;18(1):1342 [FREE Full text] [doi: 10.1186/s12889-018-6163-5] [Medline: 30595132]
14. Horberry T, Osborne R, Young K. Pedestrian smartphone distraction: Prevalence and potential severity. Transportation Research Part F: Traffic Psychology and Behaviour 2019 Jan;60:515-523. [doi: 10.1016/j.trf.2018.11.011]
15. Lim J, Chang SH, Lee J, Kim K. Effects of smartphone texting on the visual perception and dynamic walking stability. J Exerc Rehabil 2017 Feb;13(1):48-54 [FREE Full text] [doi: 10.12965/jer.1732920.460] [Medline: 28349033]
16. Kong X, Xiong S, Zhu Z, Zheng S, Long G. Development of a Conceptual Framework for Improving Safety for Pedestrians Using Smartphones While Walking: Challenges and Research Needs. Procedia Manufacturing 2015;3:3636-3643 [FREE Full text] [doi: 10.1016/j.promfg.2015.07.749]
17. Haga S, Sano A, Sekine Y, Sato H, Yamaguchi S, Masuda K. Effects of using a Smart Phone on Pedestrians' Attention and Walking. Procedia Manufacturing 2015 Sep;3(3):2574-2580. [doi: 10.1016/j.promfg.2015.07.564]
18. Kim HJ, Min JY, Kim HJ, Min KB. Accident risk associated with smartphone addiction: A study on university students in Korea. J Behav Addict 2017 Dec 01;6(4):699-707 [FREE Full text] [doi: 10.1556/2006.6.2017.070] [Medline: 29099234]
19. Ropaka M, Nikolaou D, Yannis G. Investigation of traffic and safety behavior of pedestrians while texting or web-surfing. In: Traffic Injury Prevention. 2020 Jun 05 Presented at: Proceedings of the Transportation Research Board (TRB) 99th Annual Meeting; January 2020; Washington, DC p. 389-394. [doi: 10.1080/15389588.2020.1770741]
20. Schaposnik LP, Unwin J. The phone walkers: a study of human dependence on inactive mobile devices. Behaviour 2018;155(5):389-414. [doi: 10.1163/1568539x-00003496]
21. Pokemon Go Community Day (October 2019): Start Times, Shiny Trapinch, And More. www.gamespot.com. 2019. URL: https://www.gamespot.com/articles/pokemon-go-community-day-october-2019-start-times-/1100-6470427/ [accessed 2020-03-17]
22. Underwood G, Everatt J. Chapter 6 Automatic and controlled information processing: The role of attention in the processing of novelty. In: Neumann O, Sanders AF, editors. Handbook of Perception and Action. Vol 3. London: Academic Press; 1996:185-227.
23. Aagaard J. Media multitasking, attention, and distraction: a critical discussion. Phenom Cogn Sci 2014 Jul 10;14(4):885-896. [doi: 10.1007/s11097-014-9375-x]
24. Zhang W, Zhang L. Explicating multitasking with computers: Gratifications and situations. Comput Human Behav 2012 Sep;28(5):1883-1891. [doi: 10.1016/j.chb.2012.05.006]

25. Gary CS, Lakhiani C, DeFazio MV, Masden DL, Song DH. Smartphone use during ambulation and pedestrian trauma: A public health concern. J Trauma Acute Care Surg 2018 Dec;85(6):1092-1101. [doi: 10.1097/TA.0000000000002051] [Medline: 30124625]
26. Rosenberger R. Embodied technology and the dangers of using the phone while driving. Phenom Cogn Sci 2011 Oct 15;11(1):79-94. [doi: 10.1007/s11097-011-9230-2]
27. Mueller PA, Oppenheimer DM. The pen is mightier than the keyboard: advantages of longhand over laptop note taking. Psychol Sci 2014 Jun;25(6):1159-1168. [doi: 10.1177/0956797614524581] [Medline: 24760141]
28. Bungum TJ, Day C, Henry LJ. The association of distraction and caution displayed by pedestrians at a lighted crosswalk. J Community Health 2005 Aug;30(4):269-279. [doi: 10.1007/s10900-005-3705-4] [Medline: 15989209]
29. Thompson LL, Rivara FP, Ayyagari RC, Ebel BE. Impact of social and technological distraction on pedestrian crossing behaviour: an observational study. Inj Prev 2013 Aug;19(4):232-237 [FREE Full text] [doi: 10.1136/injuryprev-2012-040601] [Medline: 23243104]
30. Una empresa valenciana desarrolla un sistema para prevenir los atropellos de 'peatones zombi'. laSexta.com, Europa Press. Madrid: Atresmedia; 2019. URL: https://www.lasexta.com/noticias/sociedad/empresa-valenciana-desarrolla-sistema-prevenir-atropellos-peatones-zombi_201905295cee925e0cf22ed9c59dc251.html [accessed 2020-01-15]
31. Levanta la cabeza. Compromiso Atresmedia, Internet. Madrid: Atresmedia; 2019. URL: https://compromiso.atresmedia.com/levanta-la-cabeza/especiales/text-neck/ [accessed 2020-01-15]
32. Robles P. Why your Smartphone is causing you "text neck" syndrome. South China Morning Post. 2019. URL: https://multimedia.scmp.com/lifestyle/article/2183329/text-neck/index.html [accessed 2020-01-15]
33. Barin EN, McLaughlin CM, Farag MW, Jensen AR, Upperman JS, Arbogast H. Heads Up, Phones Down: A Pedestrian Safety Intervention on Distracted Crosswalk Behavior. J Community Health 2018 Aug;43(4):810-815. [doi: 10.1007/s10900-018-0488-y] [Medline: 29492825]
34. Kim D, Han K, Sim JS, Noh Y. Smombie Guardian: We watch for potential obstacles while you are walking and conducting smartphone activities. PLoS One 2018;13(6):e0197050 [FREE Full text] [doi: 10.1371/journal.pone.0197050] [Medline: 29944656]
35. Johnson GM, Kulpa A. Dimensions of online behavior: toward a user typology. Cyberpsychol Behav 2007 Dec;10(6):773-779. [doi: 10.1089/cpb.2007.0043] [Medline: 18085964]
36. Livingstone S, Helsper E. Gradations in digital inclusion: children, young people and the digital divide. New Media Soc 2016 Jun 30;9(4):671-696. [doi: 10.1177/1461444807080335]
37. Lu JM, Lo YC. Investigation of smartphone use while walking and its influences on one's behavior among pedestrians in Taiwan. Communications in Computer and Information Science 2017:469-475. [doi: 10.1007/978-3-319-58753-0_67]
38. Yoshiki S, Tatsumi H, Tsutsumi K, Miyazaki T, Fujiki T. Effects of Smartphone Use on Behavior While Walking. URPR 2017;4:138-150. [doi: 10.14398/urpr.4.138]
39. Jiang K, Ling F, Feng Z, Ma C, Kumfer W, Shao C, et al. Effects of mobile phone distraction on pedestrians' crossing behavior and visual attention allocation at a signalized intersection: An outdoor experimental study. Accid Anal Prev 2018 Jun;115:170-177. [doi: 10.1016/j.aap.2018.03.019] [Medline: 29604515]
40. Caramia C, Bernabucci I, D'Anna C, De Marchis C, Schmid M. Gait parameters are differently affected by concurrent smartphone-based activities with scaled levels of cognitive effort. PLoS ONE 2017 Oct 12;12(10):e0185825. [doi: 10.1371/journal.pone.0185825]
41. Mourra GN, Sénécal S, Fredette M, Lepore F, Faubert J, Bellavance F, et al. Using a smartphone while walking: The cost of smartphone-addiction proneness. Addict Behav 2020 Jul;106:106346 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106346] [Medline: 32114216]

Edited by G Eysenbach; submitted 14.04.20; peer-reviewed by J Unwin, A Kulakli; comments to author 12.06.20; revised version received 19.06.20; accepted 25.06.20; published 14.08.20.

Please cite as:
Fernández C, Vicente MA, Carrillo I, Guilabert M, Mira JJ
Factors Influencing the Smartphone Usage Behavior of Pedestrians: Observational Study on "Spanish Smombies"
J Med Internet Res 2020;22(8):e19350
URL: http://www.jmir.org/2020/8/e19350/
doi:10.2196/19350
PMID:32667896

Original Paper

Novel Toilet Paper–Based Point-Of-Care Test for the Rapid Detection of Fecal Occult Blood: Instrument Validation Study

Hsin-Yao Wang[1,2*], MD; Ting-Wei Lin[1*], MD; Sherry Yueh-Hsia Chiu[3,4], PhD; Wan-Ying Lin[5], MD; Song-Bin Huang[6], PhD; Jason Chia-Hsun Hsieh[7,8], MD, PhD; Hsieh Cheng Chen[6], MSc; Jang-Jih Lu[1,9,10], MD, PhD; Min-Hsien Wu[2,7,11,12], PhD

[1]Department of Laboratory Medicine, Chang Gung Memorial Hospital at Linkou, Taoyuan City, Taiwan
[2]PhD Program in Biomedical Engineering, Chang Gung University, Taoyuan City, Taiwan
[3]Department of Health Care Management, College of Management, Chang Gung University, Taoyuan City, Taiwan
[4]Division of Hepatogastroenterology, Department of Internal Medicine, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung City, Taiwan
[5]Syu Kang Sport Clinic, Taipei City, Taiwan
[6]Sigknow Biomedical Co Ltd, Taipei City, Taiwan
[7]Division of Oncology, Department of Internal Medicine, Chang Gung Memorial Hospital at Linkou, Taoyuan City, Taiwan
[8]Division of Oncology, Department of Internal Medicine, New Taipei Municipal TuCheng Hospital, New Taipei City, Taiwan
[9]School of Medicine, Chang Gung University, Taoyuan City, Taiwan
[10]Department of Medical Biotechnology and Laboratory Science, Chang Gung University, Taoyuan City, Taiwan
[11]Graduate Institute of Biomedical Engineering, Chang Gung University, Taoyuan City, Taiwan
[12]Department of Chemical Engineering, Ming Chi University of Technology, New Taipei City, Taiwan
[*]these authors contributed equally

Corresponding Author:
Min-Hsien Wu, PhD
PhD Program in Biomedical Engineering
Chang Gung University
No. 259, Wenhua 1st Rd, Guishan Dist
Taoyuan City
Taiwan
Phone: 886 32118800
Email: mhwu@mail.cgu.edu.tw

Abstract

Background: Colorectal cancer screening by fecal occult blood testing has been an important public health test and shown to reduce colorectal cancer–related mortality. However, the low participation rate in colorectal cancer screening by the general public remains a problematic public health issue. This fact could be attributed to the complex and unpleasant operation of the screening tool.

Objective: This study aimed to validate a novel toilet paper–based point-of-care test (ie, JustWipe) as a public health instrument to detect fecal occult blood and provide detailed results from the evaluation of the analytic characteristics in the clinical validation.

Methods: The mechanism of fecal specimen collection by the toilet-paper device was verified with repeatability and reproducibility tests. We also evaluated the analytical characteristics of the test reagents. For clinical validation, we conducted comparisons between JustWipe and other fecal occult blood tests. The first comparison was between JustWipe and typical fecal occult blood testing in a central laboratory setting with 70 fecal specimens from the hospital. For the second comparison, a total of 58 volunteers were recruited, and JustWipe was compared with the commercially available Hemoccult SENSA in a point-of-care setting.

Results: Adequate amounts of fecal specimens were collected using the toilet-paper device with small day-to-day and person-to-person variations. The limit of detection of the test reagent was evaluated to be 3.75 μg of hemoglobin per milliliter of reagent. Moreover, the test reagent also showed high repeatability (100%) on different days and high reproducibility (>96%) among different users. The overall agreement between JustWipe and a typical fecal occult blood test in a central laboratory setting was 82.9%. In the setting of point-of-care tests, the overall agreement between JustWipe and Hemoccult SENSA was 89.7%. Moreover, the usability questionnaire showed that the novel test tool had high scores in operation friendliness (87.3/100), ease of reading results (97.4/100), and information usefulness (96.1/100).

Conclusions: We developed and validated a toilet paper–based fecal occult blood test for use as a point-of-care test for the rapid (in 60 seconds) and easy testing of fecal occult blood. These favorable characteristics render it a promising tool for colorectal cancer screening as a public health instrument.

(J Med Internet Res 2020;22(8):e20261) doi:10.2196/20261

KEYWORDS

fecal occult blood test; point-of-care diagnostics; paper-based analytical devices; diagnostic; testing; detection; validation; cancer; public health

Introduction

Colorectal cancer ranks globally as the third-leading and second-leading cancer type in terms of incidence and mortality rates, respectively. The incidence rate of colorectal cancer continues to rise because of aging populations and lifestyle changes. The continuous increase in its incidence rate has brought considerable health burdens worldwide [1-3]. Clinically, early screening of colorectal cancer in average-risk persons has been shown to reduce both mortality and incidence [4-7]. A variety of screening methods, including fecal occult blood testing, fecal immunochemical-based testing, sigmoidoscopy, digital rectal examination, colonoscopy, and computed tomographic (CT) colonography, have been developed for colorectal cancer screening [8]. However, the overall screening rate is still low [2], especially in areas with limited resources [9,10] or in socioeconomic minority groups [11-13]. In the United States, for example, the overall colorectal cancer screening rate was approximately 67.3% in 2016 [14]. In Taiwan, the screening rate was approximately 52.3% to 56.6% [15]. The low rate of colorectal cancer screening can be mainly attributed to the low availability, poor usability, or high cost of screening tools [9]. It is generally believed that the keys to successful implementation of colorectal cancer screening are cost-effectiveness, patient preference, and related professional medical resources [4,16,17]. Furthermore, it is important to provide flexibility and different choices for patients to conduct the examination themselves [18,19].

Among the screening methods described earlier, flexible sigmoidoscopy and colonoscopy are the gold standard primary screening methods [9,20,21] and can visually inspect the internal lining of the intestine [22]. By using endoscopy-based techniques, suspicious lesions or tumors can be detected, removed, and confirmed by pathology examination. However, scopy-based techniques require highly trained medical staff and specialized instruments. The operation of scopy-based instruments and the interpretation of the results are also highly operator dependent [23]. Moreover, the risks of scopy-based techniques include lower gastrointestinal bleeding, perforation, myocardial infarction, and ischemic stroke (approximately 5.3 per 10,000 persons; without biopsy or intervention, approximately 4.3 per 10,000 persons) [24,25]. CT colonography is a noninvasive exam for colorectal cancer screening with high sensitivity; it is comparable to colonoscopy [26,27]. Nevertheless, its widespread application is also limited by the availability of CT instruments and related facilities. High radiation exposure is another unfavorable feature of CT colonography [9,28]. Fecal occult blood tests and fecal immunochemical-based tests are both stool-based, noninvasive tests for colorectal cancer screening. Fecal occult blood testing targets heme, while fecal immunochemical-based testing detects hemoglobin in stool specimens. Studies have reported fecal immunochemical-based testing to have slightly higher performance than fecal occult blood testing in colorectal cancer screening [6,29]. However, the cost of fecal immunochemical-based testing is higher than that of fecal occult blood testing due to the expensive antihemoglobin antibody used in fecal immunochemical-based tests [30]. In terms of the successful implementation of colorectal cancer screening, fecal occult blood testing is the only screening method endorsed by the American Cancer Society to have the technical features of low cost and low medical profession requirements compared to other methods [6,31]. However, for current fecal occult blood testing, it is normally required that the test be performed by professional staff, restricting its application for point-of-care test or even home use. Moreover, current fecal occult blood tests generally require the user to collect fecal specimens using sticks to scoop stool after defecation. This unpleasant process could affect the widespread utilization of fecal occult blood tests for colorectal cancer screening [32-39].

To address this issue, we herein proposed a toilet paper–based fecal occult blood test (JustWipe) encompassing a toilet paper designed for fecal specimen collection and the reagents required for fecal occult blood test. For fecal specimen collection, we evaluated the working performance of the specially designed toilet paper. Moreover, we also developed and verified the analytical reagents used in the toilet paper–based tool. Based on the toilet paper design and analytical reagents, comparisons between the toilet paper–based tool and commercially available Hemoccult SENSA, as well as routine hospital fecal occult blood tests, were conducted to validate the new tool's clinical utility and show that by using the novel toilet paper–based fecal occult blood test, we could easily collect fecal specimens in a regular buttocks-wiping move after defecation and detect fecal occult blood rapidly and accurately.

Methods

Design of the Toilet Paper-Based Fecal Specimen Collection Device

The toilet paper–based fecal occult blood test (JustWipe; Sigknow Biomedical Co Ltd) contained the testing paper for fecal specimen collection and the necessary reagents, as illustrated in Figure 1. Briefly, the specimen collection device was designed as a toilet paper–based tool. The user sticks his or her fingers onto the specific adhesive area on the front side

of the paper and collects fecal specimens on the back side of the toilet paper during an ordinary buttocks-wiping move. The reagent part contains the two developers required for testing occult blood in the fecal specimen. Specifically, the toilet paper for fecal specimen collection (width: 100 mm, height: 140 mm) was composed of leaf bleached kraft pulp, which is commonly used for ordinary toilet paper. The front side (ie, hand handle side) was designed as a circular adhesive area (diameter: 50 mm) with a visual mark to facilitate correct handling of the tool (Figure 1 and Figure 2). The back side (ie, specimen collection side) was designed with a circular stool collection area (diameter: 50 mm) composed of a water repellent polyester cloth for fecal specimen collection (Figure 1 and Figure 2). Developer A contained 3,3',5,5'-tetramethylbenzidine, and developer B contained hydrogen peroxide and ethanol. For JustWipe, the steps are to stick fingers onto the circular area of the front side, wipe buttocks, and collect fecal specimens on the circular area on the back side, fold the device, and apply reagents (developer A and developer B) to develop the test reaction, whereas for typical fecal occult blood tests, the steps are to defecate, scoop stool with stick, spread the collected stool on the test zone, and apply reagent to develop the test reaction. The entire process for JustWipe is schematically illustrated in comparison with that of the conventional fecal occult blood test counterpart in Figure 3. When occult blood exists in the specimen, a blue-green color develops. In contrast, a typical guaiac-based fecal occult blood test (eg, Hemoccult Sensa; Beckman Coulter) would require additional feces collection and spreading using sticks.

Figure 1. Design of the toilet paper-based fecal occult blood point-of-care test (JustWipe). TMB: 3,3',5,5'-tetramethylbenzidine; H_2O_2: hydrogen peroxide.

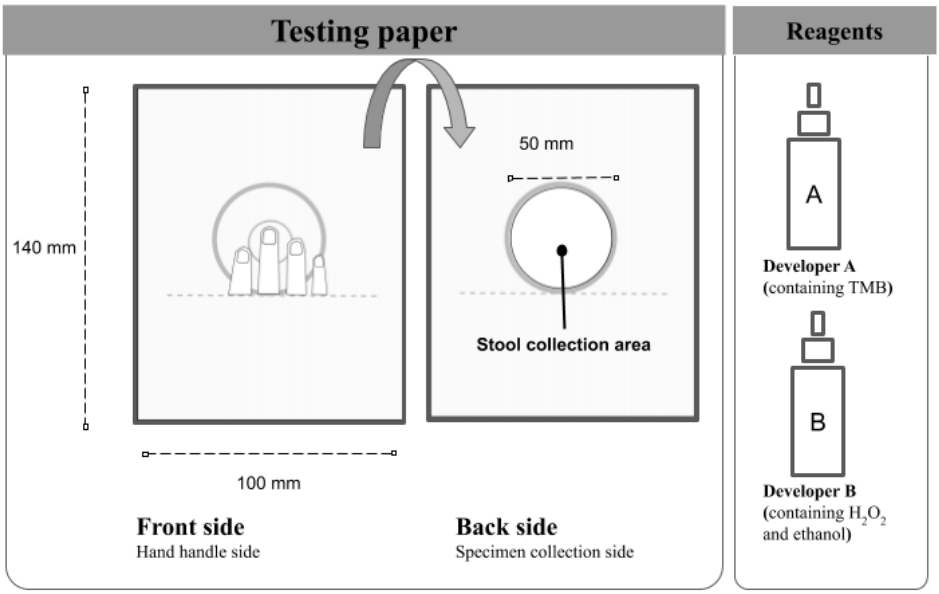


Figure 2. The appearance of the JustWipe: (A) The front side of the toilet testing paper with the hand handle side is clearly marked. (B) The backside of the toilet testing paper has a stool collection area with water repellent polyester cloth.

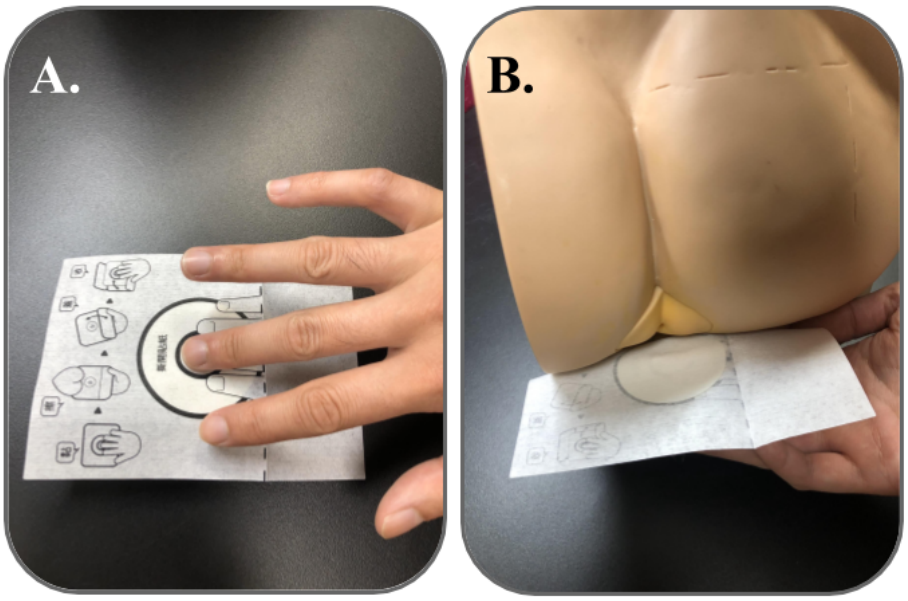
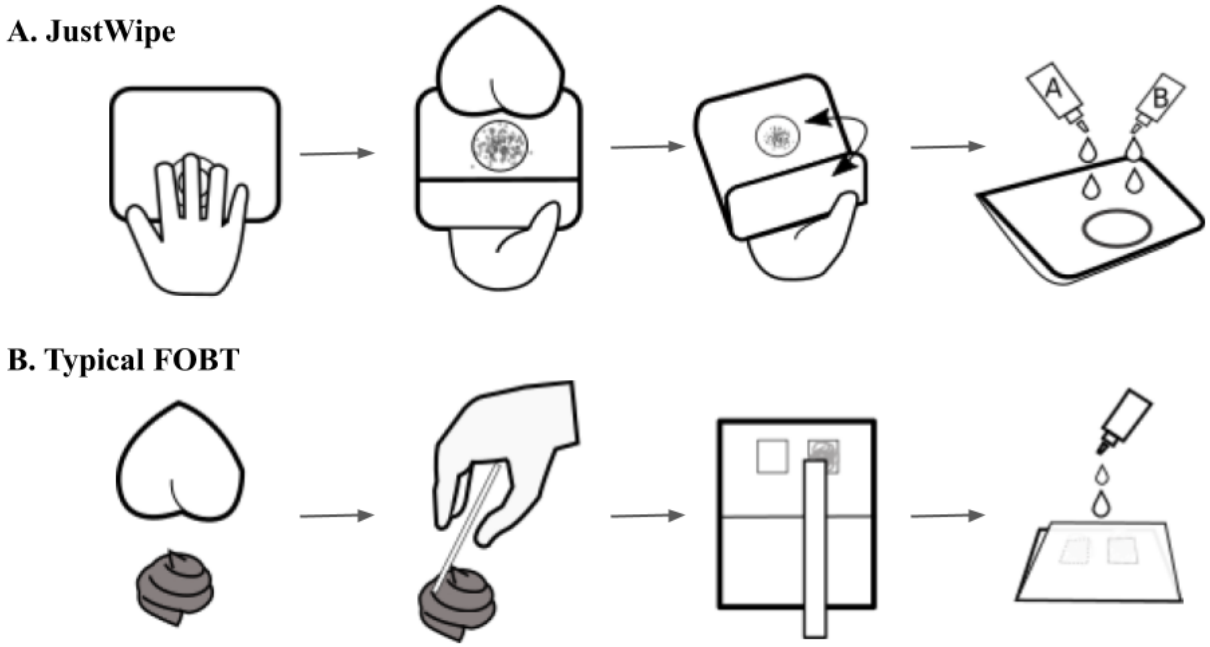


Figure 3. Comparison between JustWipe and typical fecal occult blood testing processes.



Specimen Collection Performance of the Toilet Paper-Based Fecal Specimen Collection Device

To examine the variation in specimen collection using the toilet paper–based fecal specimen collection, 18 volunteers were recruited for the evaluation. For three consecutive days, each volunteer collected a specimen after defecation using the JustWipe method directly with a testing paper device. The weight of the testing paper was measured before and after collecting the specimen. The specimen weight from the testing paper was calculated by subtracting the testing paper weight before wiping from the testing paper weight after wiping.

Limit of Detection of the Testing Reagents

We conducted a qualitative performance evaluation to determine the limit of detection of the reagents used in JustWipe according to Clinical and Laboratory Standards Institutes (CLSI) EP12-A2 User Protocol for Evaluation of Qualitative Test Performance [40]. Serial concentrations of hemoglobin solutions (0 μg/mL, 1.88 μg/mL, 2.26 μg/mL, and 3.75 μg/mL) consisting of human hemoglobin powder (Sigma-Aldrich) dissolved in double distilled water (ddH_2O; Sigma-Aldric) were prepared. For each concentration, we applied 10 μL of hemoglobin solution to the circular collection area on the back side of the specimen collection device and then covered the circular collection area by folding the paper. Subsequently, we added 2 drops (80 μL) of reagent A and 2 drops (80 μL) of reagent B to develop the reaction for 60 seconds. Tests for each hemoglobin concentration were replicated 160 times (40 times by n=4 operators).

Intraassay and Interassay Repeatability of the Testing Reagents

We evaluated both intraassay and interassay repeatability based on CLSI EP12-A2 [40]. We measured the performance of the testing reagents on different hemoglobin concentrations (0 μg/mL, 3.75 μg/mL, and 15 μg/mL) on three different days. Each hemoglobin concentration was tested 9 times on each independent day. We applied 10 μL of hemoglobin solution to the circular collection area on the back side of the specimen collection device and then covered the circular collection area by folding the paper. Subsequently, we added 2 drops (80 μL) of reagent A and 2 drops (80 μL) of reagent B to develop the reaction for 60 seconds. Tests of the same hemoglobin concentration performed on the same day were used to calculate the intraassay repeatability. In contrast, interassay repeatability was evaluated using the test results collected on different days.

Reproducibility of the Testing Reagents in Untrained Users and Medical Staff

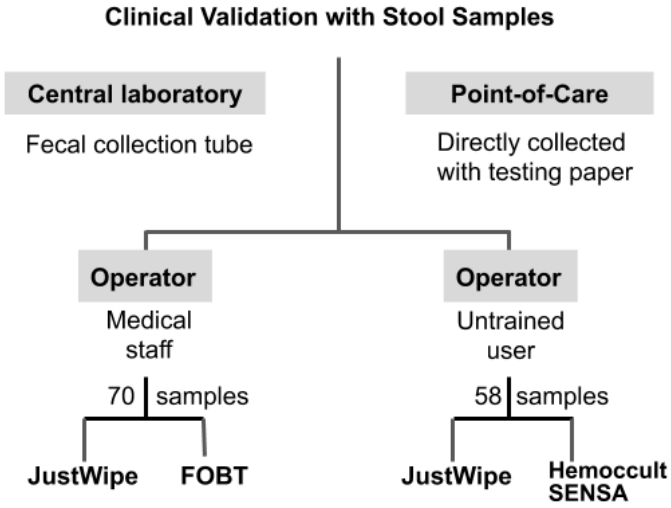
The reproducibility between untrained users (n=50) and trained medical staff (n=2) using the test reagents was evaluated based on CLSI EP5-A3 Evaluation of Precision Performance of Quantitative Measurement Methods guidance [41]. Two concentrations of hemoglobin solution (0 μg/mL and 3.75 μg/mL) were prepared and tested by the untrained users and trained medical staff. Each concentration of hemoglobin solution was tested twice.

Performance Comparison Between JustWipe and the Typical Fecal Occult Blood Test in a Central Medical Laboratory

We collected 70 convenience nonduplicate stool specimens from the central medical laboratory of Chang Gung Memorial Hospital Linkou branch between September 11, 2019 and March 24, 2020. The study was approved by the Institutional Review Board of Chang Gung Medical Foundation (No. 201901287B0). The stool specimens were collected in wards by nurses and placed into routine stool collection tubes [42]. The volume of the specimens was visually estimated to be greater than the size of a thumb. The specimens were deidentified before analytical

measurements. An aliquot of stool was tested using a typical fecal occult blood test (O-tolidine test; Shin-Yung Medical Instruments Co Ltd). Another aliquot of stool was tested using JustWipe. The study design is illustrated in Figure 4.

Figure 4. Study design flowchart. FOBT: fecal occult blood test.



Performance Comparison Between JustWipe (as a Point-of-Care Test) and Hemoccult Sensa

We recruited 58 volunteers to use JustWipe as a point-of-care test in the period between March 1, 2019 and January 31, 2020. The volunteers were recruited from the Chang Gung Memorial Hospital and Chang Gung University. The study was approved by the Institutional Review Board of Chang Gung Medical Foundation (No. 201900133B0). The volunteers were asked to follow the instructions of JustWipe after defecation: the volunteer wiped his or her buttocks to collect stool specimens with the toilet paper–based fecal specimen collection device, followed by the steps of folding the device and applying the reagents. The volunteers interpreted the results themselves after 60 seconds of reaction. Subsequently, for testing with Hemoccult SENSA, the volunteers collected the rest of the specimens (a volume greater than that of a thumb) and sent the specimens to the medical staff within 20 hours which were kept at room temperature or stored between 2 ℃ to 8 ℃ and shipped to the laboratory next day. For Hemoccult SENSA (Beckman Coulter), the stool specimens received by the laboratory were then tested by medical staff according to the instructions of Hemoccult SENSA.

Usability and User Preference Evaluation of JustWipe

The questionnaire was designed to assess the usability and user preference of the JustWipe. The questionnaire contained 12 questions about the user's experience of the toilet paper–based fecal occult blood test (Table 1). Each question had 5 different response options (strongly agree, agree, neutral, disagree, and strongly disagree), which reflected the volunteers' feedback on the usability of the test. The 58 recruited volunteers from the above study were requested to fill in the questionnaire by themselves after using JustWipe.

Table 1. Questionnaire used for usability evaluation. There were 12 questions on the questionnaire to assess the three aspects of the test, namely, operation friendliness, ease of reading the results, and information usefulness.

Group	Number	Items	Average agreement, %
Information usefulness	1	Did you know that the detection target is fecal occult blood?	100
	2	Did you know that the testing result is not an indicator of cancer?	98.6
	3	Did you know that the testing result is only for physical conditions?	100
	4	Do you know what to do after testing?	88.3
	5	If the test result is positive, would you go to the hospital for further examination?	93.8
Operational friendliness	6	Are the two bottle designs easy to identify?	89.0
	7	Do you understand all of the items in the device using the instruction manual?	86.2
	8	Is detailed company information provided in the instruction manual?	87.2
	9	Are all cautions clearly presented to the user?	87.6
	10	Are the operational procedures clearly presented to the user through icons and words?	86.9
Ease of result reading	11	Is your interpretation of the circle window the same as the medical staff's result?	95.2
	12	Is your interpretation of the control area the same as the medical staff's result?	99.7

Statistical Analysis

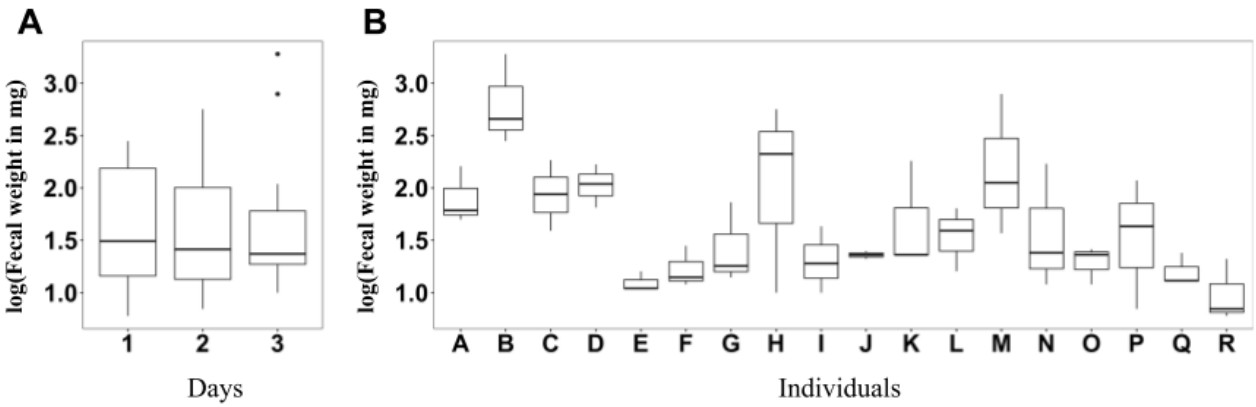
We used a one-way analysis of variance (ANOVA) to test weight differences among specimens collected on different days. Kruskal-Wallis one-way ANOVA was used to test weight differences among specimens collected by different individuals. The statistical analysis of the two diagnostic test evaluation studies followed the statistical guidance on reporting results from studies evaluating diagnostic tests [43]. The 2×2 tables for each study were produced for comparisons between the index test method (ie, JustWipe) and the comparative methods. The estimation of the agreements included positive agreement, negative agreement, and overall agreement. Approximate 95% confidence limits for the true overall, positive, and negative agreement were calculated as the estimated value ± 2 standard error. The method for standard error calculation followed EP12-P from National Committee for Clinical Laboratory Standards [44].

Results

Variation in Specimen Collection is Acceptable Among Different Individuals and Different Days

We evaluated the variation of specimen collection when using the toilet paper–based fecal specimen collection device (Figure 5). Both between-day variation and between-individual variation were evaluated in 18 individuals. In total, the median specimen weight was 25 mg, and the interquartile range was 89 mg. The minimum specimen weight was 6 mg, and the maximum weight was 1897 mg. Regarding the between-day variation, there was no difference in the weight of specimens collected on the 3 different days (P=.12). In contrast, a significant specimen weight difference was noted for different individuals. Specifically, 3 out of the 18 individuals collected a significantly greater amount (as a specimen) than the others did. When these 3 individuals were excluded, the statistical test showed no specimen weight differences among individuals. In summary, between-day variation was not significant. Between-individual variation was significant though, the outliers collected more specimen than the averages.

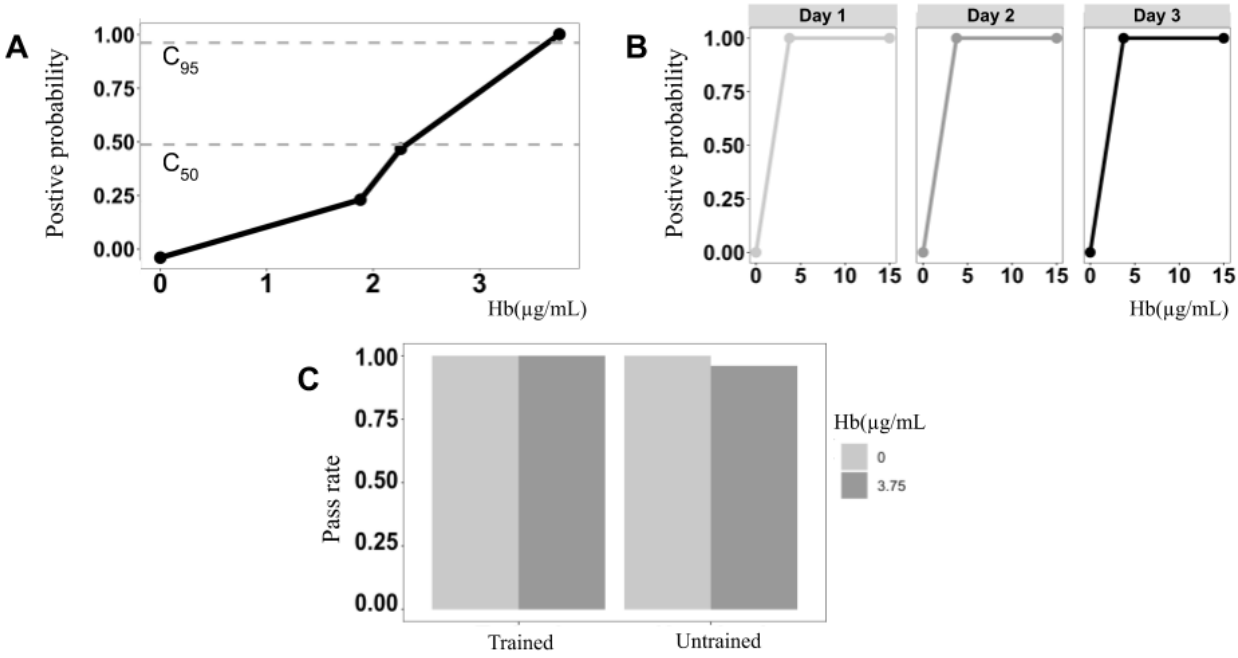
Figure 5. Weight distribution and variation of fecal specimens that are collected among (A) different days and (B) different users.



Analytical Characteristics of the Test Reagents

An imprecision curve of the analytical reagents used in JustWipe was used to illustrate the analytical characteristics (Figure 6). The probabilities of positive results for 2.26 μg/mL and 3.75 μg/mL were 48.8% and 98.8%, respectively. The concentration of 2.26 μg/mL could be defined as C50 (at which yielded 50% positive result and 50% negative result) for the test reagents [43]. The concentration of 3.75 μg/mL could be defined as the limit of detection (equal to C95) at which over 95% of the samples tested positive. The repeatability of testing different hemoglobin levels on different days is demonstrated in Figure 6. The positive probabilities for 5 μg/mL and 15 μg/mL were 100% for each over the three days; for 0 μg/mL, the positive probability was 0% over the three days. The repeatability for a longer period (20 days) is shown in Multimedia Appendix 1. Moreover, the reproducibility across untrained individuals and medical staff is illustrated in Figure 6. The trained users correctly operated all the samples whose concentrations were either 0 μg/mL or 3.75 μg/mL; the untrained users correctly operated the samples of 0 μg/mL, but the pass rate for 3.75 μg/mL was 96.0% (48/50).

Figure 6. Analytical characteristics of the test reagents: (A) limit of detection, (B) positive probabilities for 5 and 15 μg/mL, and (C) reproducibility among trained (medical staff) or untrained users.



Clinical Validation of JustWipe in a Central Laboratory Setting

A comparison of the performance between JustWipe and a typical fecal occult blood test (O-tolidine–based test) was conducted using 70 clinical specimens collected in a tertiary referral hospital. The overall agreement was 82.9% (52/70); the positive agreement and negative agreement were 83.9% (26/31) and 82.1% (32/39), respectively (Table 2). The qualitative test results of both methods can be found in Multimedia Appendix 2.

Table 2. Performance comparison between JustWipe and a typical fecal occult blood test, and between JustWipe and Hemoccult SENSA.

Comparison	Hospital fecal occult blood test			Hemoccult SENSA		
	Positive	Negative	Total	Positive	Negative	Total
JustWipe, n						
Positive	26	7	33	16	2	18
Negative	5	32	37	4	36	40
Total	31	39	70	20	38	58
Agreement, %						
Estimate (95% CL[a])	83.9 (70.7, 97.1)	82.1 (69.8, 94.3)	82.9 (73.9, 91.9)	80.0 (62.1, 80.0)	94.7 (87.5, 102)	89.7 (81.7, 97.7)

[a]95% confidence limits calculated as (estimate – 2 standard error, estimate + 2 standard error).

Validation of JustWipe as a Point-Of-Care Test

The intended use of JustWipe is testing occult blood in the stool at home or at a point of care. Thus, we evaluated the performance of JustWipe when it was used as a point-of-care test according to the manufacturer's instructions. The overall agreement was 89.7% (52/58); the positive agreement and negative agreement were 80.0% (16/20) and 94.7% (36/38), respectively (Table 2). The qualitative test results of both methods can be found in Multimedia Appendix 3.

Usability and User Preference Evaluation of JustWipe

Of the 58 volunteers, there were 32 (55.2%) women and 26 (44.9%) men. The average age of the volunteers was 59.1 (SD 12.7) years. We asked the volunteers to complete the questionnaire (12 questions included) after using JustWipe as a point-of-care test. We summarized the results of the questionnaire into three categories: information usefulness (Q1, Q2, Q3, Q4, and Q5), operational friendliness (Q6, Q7, Q8, Q9, and Q10), and ease of reading results (Q11 and Q12). In terms of information usefulness, average agreement percentages for Q1, Q2, Q3, Q11, and Q12 were 100%, 98.6%, 100%, 88.3%, and 93.8%, respectively. In the category of operational friendliness, average agreement percentages for Q6, Q7, Q8, Q9, and Q10 were 89%, 86.2%, 87.2%, 87.6%, and 86.9%, respectively. For the category of ease of reading results, Q11 had an average of 95.2% and Q12 had an average of 99.7%.

Discussion

Principal Findings

In this study, we developed a toilet paper–based point-of-care test (JustWipe) for the rapid detection of fecal occult blood. An ordinary buttocks-wiping move on the toilet was adapted as the mechanism of stool specimen collection. Specimen collection by buttocks-wiping was evaluated to be useful and stable. In addition, a set of test reagents was developed and characterized. We compared the performance of JustWipe with that of a typical fecal occult blood test in a central laboratory setting. Moreover, we also demonstrated the performance of JustWipe as a point-of-care test. Based on the toilet paper–based device, stool specimens can be collected with a regular buttocks-wiping move. A rapid test result is available within 60 seconds using the test reagents. An ordinary user can operate the test easily, rapidly, and with reproducibility. The ease of operation and reliability render the novel fecal occult blood test a promising tool for colorectal cancer screening.

Troublesome operation of stool specimen collection is considered one of the obstacles affecting participation in colorectal cancer screening [13,32,34,36,37,39]. In a typical fecal occult blood test, fecal immunochemical-based test, or other stool-based occult blood tests, stool specimen collection is difficult. Users, typically with a nonmedical background, must sample stool by themselves. Several steps, including flushing the toilet bowl and floating tissue paper on the surface of the toilet bowl water, are required to be followed and performed correctly to ensure the quality of the specimen [45]. The users must also collect stool samples before it comes into contact with the toilet bowl water [45]. The demanding requirements not only reduce willingness to use the device but also result in some analytical errors when some of the steps are not performed correctly. In contrast, the fecal specimen collection using JustWipe requires a regular buttocks-wiping move only. The number of steps in fecal specimen collection is reduced so that the errors occurring in specimen collection can be largely mitigated. Moreover, the toilet paper–based stool specimen collection was shown to be stable between days and among different users (Figure 5). The mean and median weights of the stool specimens were 120.00 mg and 25.50 mg, respectively. The amount of specimen collected was higher than that collected via typical methods. The relationship between the amount of stool specimen to test sensitivity is not clear. However, a larger amount collected as a specimen is thought to be an advantageous feature for a test [46].

The analytical evaluation of the test reagents used in JustWipe demonstrated several favorable characteristics, including high sensitivity, high repeatability, and high reproducibility. Low-concentration hemoglobin could be detected with minimal day-to-day variation (high repeatability) and person-to-person variation (high reproducibility). Regarding the analytical sensitivity of the test reagents, the test reagents were found to have adequate analytical sensitivity to detect occult fecal blood in colorectal cancer patients. In a population study including 5.8 million individuals, a hemoglobin concentration of approximately 20 μg/mL could detect colorectal cancer with a detection rate of 5.2% in men and 2.2% in females. The hemoglobin concentration for the colorectal cancer patients identified in that study was between 172.8 and 231 μg hemoglobin/g feces [47]. In another study, a cut-off of 80-90 μg hemoglobin/g feces was sufficient for clinical application [48-50]. Based on the reported values (ie, 80-90 μg hemoglobin/g feces), a cut-off of 114.3-128.6 μg hemoglobin/mL is clinically useful when the water content in stool is 70% [51]. In contrast, the hemoglobin concentration in the feces of healthy individuals without colorectal cancer is 0.519 μg hemoglobin/mL (90% CI 0.468-0.575) in men and 0.283 μg hemoglobin/mL (90% CI 0.257-0.316) in women [50,52,53]. In brief, the limit of detection (3.75 μg hemoglobin/mL) of the test reagents of toilet paper–based tool was sensitive and stable enough for detecting fecal occult blood in colorectal cancer patients (Figure 6).

We validated JustWipe in the settings of a central laboratory and point of care. In both settings, JustWipe showed high agreement with typical test methods. In the setting of the central laboratory, both JustWipe and the comparative test (O-tolidine–based fecal occult blood test) were performed by medical staff. The positive agreement (83.9%) and negative agreement (82.1%) were quite balanced (Table 2). In contrast, the positive agreement (80.0%) was significantly lower than the negative agreement (94.7%) when we validated JustWipe used as a point-of-care test (Table 2). To validate JustWipe as a point-of-care test, the comparative method (Hemoccult SENSA) was operated by medical staff, while JustWipe was operated by nonmedical individuals. Regarding the association between the usability and the discordance between positive and negative agreement of the test results, we used Fisher exact test for the analysis. The results in Multimedia Appendix 4 showed that the usability indicators (Table 1) were not associated with the discordance of the test results. The specific usability indicators should have been associated with the discordant results. However, based on our data, the significant association was not detected in the study. The nonsignificant association could be attributed to the relatively small sample size used in the proof-of-concept validation. Yet, the association is worthy of further investigation as a key for improving the proposed device. Furthermore, the possible cause for the suboptimal positive agreement could be attributed to false negative interpretation of the weak positive reaction. Users who are not trained medical professionals may tend to ignore the weak signal on the toilet paper. To address the limitation of interpretation, especially in the weak positive case, we plan to develop an artificial intelligence–aided interpretation tool. By using the artificial intelligence–aided interpretation tool, images of the test result can be interpreted with a standardized approach. The interpretative error resulting from insufficient interpretation experience would be largely mitigated. We illustrate the approach in Multimedia Appendix 5.

The major aim of designing JustWipe was to improve the usability of fecal occult blood testing. We assessed the usability of JustWipe using a questionnaire (Table 1). The categories of operational friendliness and ease of reading results are important indicators for a nonmedical professional using a point-of-care test [37,39]. Regarding operational friendliness, the agreement

of all the questions (Q6, Q7, Q8, Q9, and Q10) was greater than 80% (range 81.9%-84.5%). For ease of reading results (Q11 and Q12), the agreement was greater than 94%. In brief, the high agreement in operational friendliness and ease of reading results indicated that users without professional medical training can easily operate the tool and interpret it following the instructions.

Limitations

Although the volunteers from our trial in the point-of-care test setting would be representative of the target population with respect to the range of age, sampling bias could still exist in the validation setting for the point-of-care test trial. All volunteers were recruited from the urban region in northern Taiwan. The average age of the recruited volunteers in the point-of-care test was 59.1 (SD 12.7) years. The range of age was approximately the age of the target population for most cancer screening programs. In the majority of Europe, the colorectal cancer prevention program recommends screening for people above 50 years old [54]. The US Preventive Services Task Force also recommend screening the population above 50 years old [6]. The colorectal cancer screening program in Taiwan also recommends screening for the population above 50 years old [55]. In this study, the volunteers were recruited from one tertiary medical center (ie Chang Gung Memorial Hospital, Linkou branch) and one university (ie, Chang Gung University). The volunteers may have had chronic diseases and subclinical conditions. The preliminary validation results reported in the study could not be easily applied to other populations. The performance using the proposed device in general population needs further investigation in a larger cohort. The demographic characteristics of the 58 recruited volunteers are listed in Multimedia Appendix 6.

Conclusions

We developed and validated a toilet paper–based point-of-care test for detecting fecal occult blood. The result showed that the toilet paper–based collection of fecal specimens was stable. The test reagents of the point-of-care test also showed high repeatability and reproducibility. The novel toilet paper–based point-of-care test revealed high agreement with the comparative methods in both central laboratory and point-of-care test settings. The usability evaluation of the point-of-care test showed high operation friendliness and high ease of reading results. The favorable characteristics render the proposed novel point-of-care test a promising colorectal cancer screening tool.

Acknowledgments

This work was edited by American Journal Experts. This work was supported by Sigknow Biomedical Co Ltd (SCRPD2I0011) and Chang Gung Memorial Hospital (CMRPD2J0031-33).

Authors' Contributions

H-YW conceptualized the study. H-YW and T-WL wrote the manuscript, analyzed the data, plotted the figures, and created the tables. S-BH and HCC performed the experiments. SY-HC and J-JL reviewed and edited the manuscript for important intellectual content. M-HW obtained funding and supervised the study. All authors discussed the results and revised the manuscript.

Conflicts of Interest

This work was supported by Sigknow Biomedical Co Ltd (SCRPD2I0011). S-BH and H-CC were the employees of the Sigknow Biomedical Co Ltd.

Multimedia Appendix 1
Repeatability for a longer period (20 days) for the analytical characteristics of the test reagents.
[DOCX File , 69 KB - jmir_v22i8e20261_app1.docx]

Multimedia Appendix 2
Qualitative test results for both methods in the central laboratory.
[DOCX File , 54 KB - jmir_v22i8e20261_app2.docx]

Multimedia Appendix 3
Qualitative test results for both methods in the point-of-care test setting.
[DOCX File , 60 KB - jmir_v22i8e20261_app3.docx]

Multimedia Appendix 4
Association between individual usability indicators and result discordancy. The table showed the contingency table for each question on recruited volunteers with discordance and concordance result.
[DOCX File , 292 KB - jmir_v22i8e20261_app4.docx]

Multimedia Appendix 5

Using the mobile device for interpretation of test results. The workflow demonstrates how to use a mobile phone to assist interpretation with artificial intelligence model prediction.
[PNG File , 38 KB - jmir_v22i8e20261_app5.png]

Multimedia Appendix 6
This table contains the demographic characteristics of age, gender, education, and occupation of the recruited volunteers.
[DOCX File , 291 KB - jmir_v22i8e20261_app6.docx]

References

1. Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi R, Corcione F. Worldwide burden of colorectal cancer: a review. Updates Surg 2016 Mar 11;68(1):7-11. [doi: 10.1007/s13304-016-0359-y] [Medline: 27067591]
2. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. Nat Rev Gastroenterol Hepatol 2019 Dec;16(12):713-732. [doi: 10.1038/s41575-019-0189-8] [Medline: 31455888]
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Sep 12;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
4. Levin TR, Corley DA, Jensen CD, Schottinger JE, Quinn VP, Zauber AG, et al. Effects of organized colorectal cancer screening on cancer incidence and mortality in a large community-based population. Gastroenterology 2018 Nov;155(5):1383-1391.e5 [FREE Full text] [doi: 10.1053/j.gastro.2018.07.017] [Medline: 30031768]
5. Winawer SJ. The history of colorectal cancer screening: a personal perspective. Dig Dis Sci 2015 Mar;60(3):596-608. [doi: 10.1007/s10620-014-3466-y] [Medline: 25599958]
6. Lin JS, Piper MA, Perdue LA, Rutter CM, Webber EM, O'Connor E, et al. Screening for colorectal cancer: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 2016 Jun 21;315(23):2576-2594. [doi: 10.1001/jama.2016.3332] [Medline: 27305422]
7. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. N Engl J Med 1993 May 13;328(19):1365-1371. [doi: 10.1056/nejm199305133281901]
8. PDQ Screening and Prevention Editorial Board. Colorectal cancer screening (PDQ). PDQ Cancer Information Summaries Internet National Cancer Institute (US).: National Cancer Institute; 2020 Mar 17. URL: https://www.ncbi.nlm.nih.gov/books/NBK65825/ [accessed 2020-03-19]
9. Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJY, Young GP, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015 Oct;64(10):1637-1649 [FREE Full text] [doi: 10.1136/gutjnl-2014-309086] [Medline: 26041752]
10. Beeker C, Kraft JM, Southwell BG, Jorgensen CM. Colorectal cancer screening in older men and women: qualitative research findings and implications for intervention. J Community Health 2000 Jun;25(3):263-278. [doi: 10.1023/a:1005104406934] [Medline: 10868818]
11. Klabunde C, Blom J, Bulliard J, Garcia M, Hagoel L, Mai V, et al. Participation rates for organized colorectal cancer screening programmes: an international comparison. J Med Screen 2015 Sep;22(3):119-126. [doi: 10.1177/0969141315584694] [Medline: 25967088]
12. de Klerk CM, Gupta S, Dekker E, Essink-Bot ML, Expert Working Group 'Coalition to reduce inequities in colorectal cancer screening' of the World Endoscopy Organization. Socioeconomic and ethnic inequities within organised colorectal cancer screening programmes worldwide. Gut 2018 Apr;67(4):679-687. [doi: 10.1136/gutjnl-2016-313311] [Medline: 28073892]
13. Dawidowicz S, Le Breton J, Moscova L, Renard V, Bercier S, Brixi Z, et al. Predictive factors for non-participation or partial participation in breast, cervical and colorectal cancer screening programmes. Fam Pract 2020 Feb 19;37(1):15-24. [doi: 10.1093/fampra/cmz031]
14. Joseph DA, King JB, Richards TB, Thomas CC, Richardson LC. Use of colorectal cancer screening tests by state. Prev Chronic Dis 2018 Jun 14;15:E80 [FREE Full text] [doi: 10.5888/pcd15.170535] [Medline: 29908051]
15. Wang Y, Chen H, Wu M, Chiu H, Taiwanese Nationwide Colorectal Cancer Screening Program. Current status and future challenge of population-based organized colorectal cancer screening: Lesson from the first decade of Taiwanese program. J Formos Med Assoc 2018 May;117(5):358-364 [FREE Full text] [doi: 10.1016/j.jfma.2017.09.010] [Medline: 28988890]
16. Lieberman D, Ladabaum U, Cruz-Correa M, Ginsburg C, Inadomi JM, Kim LS, et al. Screening for colorectal cancer and evolving issues for physicians and patients: a review. JAMA 2016 Nov 22;316(20):2135-2145. [doi: 10.1001/jama.2016.17418] [Medline: 27893135]
17. Dougherty MK, Brenner AT, Crockett SD, Gupta S, Wheeler SB, Coker-Schwimmer M, et al. Evaluation of interventions intended to increase colorectal cancer screening rates in the united states: a systematic review and meta-analysis. JAMA Intern Med 2018 Dec 01;178(12):1645-1658 [FREE Full text] [doi: 10.1001/jamainternmed.2018.4637] [Medline: 30326005]

18. Denters MJ, Deutekom M, Bossuyt PM, Fockens P, Dekker E. A feces collection paper does not enhance participation in a fecal immunochemical test-based colorectal cancer screening program. European Journal of Cancer Prevention 2013;22(4):299-304 [FREE Full text] [doi: 10.1097/cej.0b013e32835b3882]
19. Quick BW, Hester CM, Young KL, Greiner KA. Self-reported barriers to colorectal cancer screening in a racially diverse, low-income study population. J Community Health 2013 Apr;38(2):285-292. [doi: 10.1007/s10900-012-9612-6] [Medline: 22976770]
20. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA Cancer J Clin 2018 Jul;68(4):250-281 [FREE Full text] [doi: 10.3322/caac.21457] [Medline: 29846947]
21. Kaminski MF, Robertson DJ, Senore C, Rex DK. Optimizing the quality of colorectal cancer screening worldwide. Gastroenterology 2020 Jan;158(2):404-417 [FREE Full text] [doi: 10.1053/j.gastro.2019.11.026]
22. Pluta RM. Colonoscopy. JAMA 2011 Mar 16;305(11):1154 [FREE Full text] [doi: 10.1001/jama.305.16.1154]
23. Coriat R, Lecler A, Lamarque D, Deyra J, Roche H, Nizou C, et al. Quality indicators for colonoscopy procedures: a prospective multicentre method for endoscopy units. PLoS One 2012;7(4):e33957 [FREE Full text] [doi: 10.1371/journal.pone.0033957] [Medline: 22509267]
24. Wang L, Mannalithara A, Singh G, Ladabaum U. Low rates of gastrointestinal and non-gastrointestinal complications for screening or surveillance colonoscopies in a population-based study. Gastroenterology 2018 Feb;154(3):540-555.e8. [doi: 10.1053/j.gastro.2017.10.006] [Medline: 29031502]
25. Levin TR, Zhao W, Conell C, Seeff LC, Manninen DL, Shapiro JA, et al. Complications of colonoscopy in an integrated health care delivery system. Ann Intern Med 2006 Dec 19;145(12):880-886. [doi: 10.7326/0003-4819-145-12-200612190-00004] [Medline: 17179057]
26. Pickhardt PJ, Choi JR, Hwang I, Butler JA, Puckett ML, Hildebrandt HA, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. N Engl J Med 2003 Dec 04;349(23):2191-2200. [doi: 10.1056/NEJMoa031618] [Medline: 14657426]
27. Johnson CD, Chen MH, Toledano AY, Heiken JP, Dachman A, Kuo MD, et al. Accuracy of CT colonography for detection of large adenomas and cancers. N Engl J Med 2008 Sep 18;359(12):1207-1217 [FREE Full text] [doi: 10.1056/NEJMoa0800996] [Medline: 18799557]
28. Spada C, Stoker J, Alarcon O, Barbaro F, Bellini D, Bretthauer M, et al. Clinical indications for computed tomographic colonography: European Society of Gastrointestinal Endoscopy (ESGE) and European Society of Gastrointestinal and Abdominal Radiology (ESGAR) Guideline. Eur Radiol 2015 Feb;25(2):331-345 [FREE Full text] [doi: 10.1007/s00330-014-3435-z] [Medline: 25278245]
29. Levi Z, Birkenfeld S, Vilkin A, Bar-Chana M, Lifshitz I, Chared M, et al. A higher detection rate for colorectal cancer and advanced adenomatous polyp for screening with immunochemical fecal occult blood test than guaiac fecal occult blood test, despite lower compliance rate. A prospective, controlled, feasibility study. Int J Cancer 2011 May 15;128(10):2415-2424 [FREE Full text] [doi: 10.1002/ijc.25574] [Medline: 20658527]
30. Van Ballegooijen M, Habbema JDF, Boer R, Zauber AG, Brown ML. A comparison of the cost-effectiveness of fecal occult blood tests with different test characteristics in the context of annual screening in the medicare population. In: Technology Assessment Report. Rockville(MD): Agency for Healthcare Research and Quality(US); Aug 9, 2003:A.
31. Wolf A, Fontham E, Church T, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA Cancer J Clin 2018 Jul;68(4):250-281 [FREE Full text] [doi: 10.3322/caac.21457] [Medline: 29846947]
32. Phillipson L, Pitts L, Hall J, Tubaro T. Factors contributing to low readiness and capacity of culturally diverse participants to use the Australian national bowel screening kit. Public Health Res Pract 2019 Mar 06;29(1):e28231810 [FREE Full text] [doi: 10.17061/phrp28231810] [Medline: 30972408]
33. Deding U, Torp-Pedersen C, Bøggild H. The association between immigration status and ineligible stool samples for colorectal cancer screening. Cancer Epidemiol 2018 Dec;57:74-79. [doi: 10.1016/j.canep.2018.10.005] [Medline: 30326395]
34. de Klerk CM, Wieten E, van der Steen A, Ramakers CR, Kuipers EJ, Hansen BE, et al. Participation and ease of use in colorectal cancer screening: a comparison of 2 fecal immunochemical tests. Am J Gastroenterol 2019 Mar;114(3):511-518. [doi: 10.14309/ajg.0000000000000148] [Medline: 30789417]
35. Shin HY, Suh M, Choi KS, Hwang S, Jun JK, Han DS, et al. Higher satisfaction with an alternative collection device for stool sampling in colorectal cancer screening with fecal immunochemical test: a cross-sectional study. BMC Cancer 2018 Apr 02;18(1):365 [FREE Full text] [doi: 10.1186/s12885-018-4290-0] [Medline: 29609647]
36. Ellis RJB, Wilson S, Holder RL, McManus RJ. Different faecal sampling methods alter the acceptability of faecal occult blood testing: a cross sectional community survey. Eur J Cancer 2007 Jun;43(9):1437-1444. [doi: 10.1016/j.ejca.2007.03.019] [Medline: 17475476]
37. Koo JH, Arasaratnam MM, Liu K, Redmond DM, Connor SJ, Sung JJY, et al. Knowledge, perception and practices of colorectal cancer screening in an ethnically diverse population. Cancer Epidemiol 2010 Oct;34(5):604-610. [doi: 10.1016/j.canep.2010.05.013] [Medline: 20580631]

38. Le Breton J, Journy N, Attali C, Le Corvoisier P, Brixi Z, Bastuji-Garin S, et al. Improving participation in colorectal cancer screening: targets for action. Prev Med 2012 Nov;55(5):488-492. [doi: 10.1016/j.ypmed.2012.08.004] [Medline: 22926013]
39. Gwede CK, Koskan AM, Quinn GP, Davis SN, Ealey J, Abdulla R, Tampa Bay Community Cancer Network (TBCCN). Patients' perceptions of colorectal cancer screening tests and preparatory education in federally qualified health centers. J Cancer Educ 2015 Jun;30(2):294-300 [FREE Full text] [doi: 10.1007/s13187-014-0733-8] [Medline: 25249181]
40. Garrett PE, Lasky FD, Meier KL. EP12-A2, user protocol for evaluation of qualitative test performance. In: Approved Guideline-Second Edition. Wayne, Pennsylvania, USA: Clinical and Laboratory Standards Institute; Jan 2008.
41. McEnroe RJ, Magari R, Durham AP, Middle JG, Goldford MD, Pierson-Perry JF, et al. Evaluation of Precision of Quantitative Measurement Procedures. In: Approved Guideline - Third Edition. Wayne, Pennsylvania, USA: Clinical and Laboratory Standards Institute; Oct 2014.
42. Wahba N. An ortho-tolidine hydrochloride test for the detection of occult blood in faeces without dietary restrictions. J Clin Pathol 1965 Sep;18(5):687-688 [FREE Full text] [doi: 10.1136/jcp.18.5.687] [Medline: 5835456]
43. FDA, Diagnostic Devices Branch, Division of Biostatistics, Office of Surveillance and Biometrics. Reporting results from studies evaluating diagnostic tests. Guidance for Industry and FDA Staff. 2007 Mar 13. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-results-studies-evaluating-diagnostic-tests-guidance-industry-and-fda [accessed 2020-04-01]
44. Clark LW, Garrett PE, Martin R, Meyer K. User protocol for evaluation of qualitative test performance. In: Proposed Guideline. Wayne, Pennsylvania, USA: National Committee for Clinical Laboratory Standards; 2000.
45. Beckman Coulter. Hemoccult SENSA. Product Instructions. Brea, CA, USA: Beckman Coulter, Inc; 2015 Jun. URL: https://www.beckmancoulter.com/download/file/wsr-116766/462489EF?type=pdf [accessed 2020-05-14]
46. Lee E, Lee Y. Is it necessary to repeat fecal occult blood tests with borderline results for colorectal cancer screening? Ann Lab Med 2018 Jan;38(1):51-53 [FREE Full text] [doi: 10.3343/alm.2018.38.1.51] [Medline: 29071819]
47. Arana-Arri E, Idigoras I, Uranga B, Pérez R, Irurzun A, Gutiérrez-Ibarluzea I, EUSKOLON Group. Population-based colorectal cancer screening programmes using a faecal immunochemical test: should faecal haemoglobin cut-offs differ by age and sex? BMC Cancer 2017 Aug 29;17(1):577 [FREE Full text] [doi: 10.1186/s12885-017-3555-3] [Medline: 28851318]
48. Libby G, Fraser CG, Carey FA, Brewster DH, Steele RJC. Occult blood in faeces is associated with all-cause and non-colorectal cancer mortality. Gut 2018 Dec;67(12):2116-2123 [FREE Full text] [doi: 10.1136/gutjnl-2018-316483] [Medline: 30012724]
49. Chen L, Yen AM, Fraser CG, Chiu SY, Fann JC, Wang P, et al. Impact of faecal haemoglobin concentration on colorectal cancer mortality and all-cause death. BMJ Open 2013 Nov 07;3(11):e003740 [FREE Full text] [doi: 10.1136/bmjopen-2013-003740] [Medline: 24202058]
50. McDonald PJ, Strachan JA, Digby J, Steele RJC, Fraser CG. Faecal haemoglobin concentrations by gender and age: implications for population-based screening for colorectal cancer. Clin Chem Lab Med 2011 Dec 07;50(5):935-940. [doi: 10.1515/CCLM.2011.815] [Medline: 22149740]
51. Rose C, Parker A, Jefferson B, Cartmell E. The characterization of feces and urine: a review of the literature to inform advanced treatment technology. Crit Rev Environ Sci Technol 2015 Sep 02;45(17):1827-1879 [FREE Full text] [doi: 10.1080/10643389.2014.1000761] [Medline: 26246784]
52. Fraser CG, Rubeca T, Rapi S, Chen L, Chen H. Faecal haemoglobin concentrations vary with sex and age, but data are not transferable across geography for colorectal cancer screening. Clin Chem Lab Med 2014 Aug;52(8):1211-1216. [doi: 10.1515/cclm-2014-0115] [Medline: 24637000]
53. Fraser CG, Auge JM, PROCOLON Group. Faecal haemoglobin concentrations do vary across geography as well as with age and sex: ramifications for colorectal cancer screening. Clin Chem Lab Med 2015 Aug;53(9):e235-e237. [doi: 10.1515/cclm-2014-1172] [Medline: 25544746]
54. Navarro M, Nicolas A, Ferrandez A, Lanas A. Colorectal cancer population screening programs worldwide in 2016: An update. World J Gastroenterol 2017 May 28;23(20):3632-3642. [doi: 10.3748/wjg.v23.i20.3632] [Medline: 28611516]
55. Cancer Prevention Group. Taiwan breast, oral cancer, and colorectal cancer screening programs. Cancer Prevention and Control. Taipei, Taiwan: The Health Promotion Administration; 2016 Jan 26. URL: https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=1051&pid=5957 [accessed 2020-06-17]

Abbreviations

ANOVA: analysis of variance
CLSI: Clinical and Laboratory Standards Institutes
CT: computed tomographic

Edited by G Eysenbach; submitted 19.05.20; peer-reviewed by J Lau, I Zachary; comments to author 10.06.20; revised version received 22.06.20; accepted 25.06.20; published 07.08.20.

Please cite as:
Wang HY, Lin TW, Chiu SYH, Lin WY, Huang SB, Hsieh JCH, Chen HC, Lu JJ, Wu MH
Novel Toilet Paper–Based Point-Of-Care Test for the Rapid Detection of Fecal Occult Blood: Instrument Validation Study
J Med Internet Res 2020;22(8):e20261
URL: https://www.jmir.org/2020/8/e20261
doi:10.2196/20261
PMID:32763879

Original Paper

Toward Detecting Infection Incidence in People With Type 1 Diabetes Using Self-Recorded Data (Part 1): A Novel Framework for a Personalized Digital Infectious Disease Detection System

Ashenafi Zebene Woldaregay[1], MSc; Ilkka Kalervo Launonen[2], PhD; Eirik Årsand[1], PhD; David Albers[3,4], PhD; Anna Holubová[5,6], MSc; Gunnar Hartvigsen[1], PhD

[1]Department of Computer Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway

[2]Department of Clinical Research, University Hospital of North Norway, Tromsø, Norway

[3]Department of Pediatrics, Informatics and Data Science, University of Colorado, Aurora, CO, United States

[4]Department of Biomedical Informatics, Columbia University, New York, NY, United States

[5]Department of ICT in Medicine, Faculty of Biomedical Engineering, Czech Technical University, Prague, Czech Republic

[6]Spin-off Company and Research Results Commercialization Center of the First Faculty of Medicine, Charles University, Prague, Czech Republic

Corresponding Author:
Ashenafi Zebene Woldaregay, MSc
Department of Computer Science
University of Tromsø – The Arctic University of Norway
Hansine Hansens veg 54, Science building Realfagbygget, office A124
Tromsø
Norway
Phone: 47 46359333
Email: ashenafi.z.woldaregay@uit.no

Abstract

Background: Type 1 diabetes is a chronic condition of blood glucose metabolic disorder caused by a lack of insulin secretion from pancreas cells. In people with type 1 diabetes, hyperglycemia often occurs upon infection incidences. Despite the fact that patients increasingly gather data about themselves, there are no solid findings that uncover the effect of infection incidences on key parameters of blood glucose dynamics to support the effort toward developing a digital infectious disease detection system.

Objective: The study aims to retrospectively analyze the effect of infection incidence and pinpoint optimal parameters that can effectively be used as input variables for developing an infection detection algorithm and to provide a general framework regarding how a digital infectious disease detection system can be designed and developed using self-recorded data from people with type 1 diabetes as a secondary source of information.

Methods: We retrospectively analyzed high precision self-recorded data of 10 patient-years captured within the longitudinal records of three people with type 1 diabetes. Obtaining such a rich and large data set from a large number of participants is extremely expensive and difficult to acquire, if not impossible. The data set incorporates blood glucose, insulin, carbohydrate, and self-reported events of infections. We investigated the temporal evolution and probability distribution of the key blood glucose parameters within a specified timeframe (weekly, daily, and hourly).

Results: Our analysis demonstrated that upon infection incidence, there is a dramatic shift in the operating point of the individual blood glucose dynamics in all the timeframes (weekly, daily, and hourly), which clearly violates the usual norm of blood glucose dynamics. During regular or normal situations, higher insulin and reduced carbohydrate intake usually results in lower blood glucose levels. However, in all infection cases as opposed to the regular or normal days, blood glucose levels were elevated for a prolonged period despite higher insulin and reduced carbohydrates intake. For instance, compared with the preinfection and postinfection weeks, on average, blood glucose levels were elevated by 6.1% and 16%, insulin (bolus) was increased by 42% and 39.3%, and carbohydrate consumption was reduced by 19% and 28.1%, respectively.

Conclusions: We presented the effect of infection incidence on key parameters of blood glucose dynamics along with the necessary framework to exploit the information for realizing a digital infectious disease detection system. The results demonstrated that compared with regular or normal days, infection incidence substantially alters the norm of blood glucose dynamics, which are quite significant changes that could possibly be detected through personalized modeling, for example, prediction models and

anomaly detection algorithms. Generally, we foresee that these findings can benefit the efforts toward building next generation digital infectious disease detection systems and provoke further thoughts in this challenging field.

(J Med Internet Res 2020;22(8):e18911) doi:10.2196/18911

KEYWORDS

type 1 diabetes; self-recorded health data; infection incidence; decision making; infectious disease outbreaks; public health surveillance

Introduction

The incidence of infectious disease outbreaks can create panic in society and is a threat to local and global health security. Such outbreaks require immediate detection and appropriate response during the initial phase of the incidence to reduce fatality and save lives [1]. The timeliness of outbreak detection defines the success of the appropriate response by the concerned bodies. The state-of-the-art syndromic surveillance systems have been improved compared with the traditional surveillance system, which is generally passive and dependent on laboratory confirmation [2]. Syndromic surveillance makes use of features that come before diagnosis, including different activities triggered by the onset of symptoms, such as Google search, Twitter, school and work absenteeism, pharmacy drug sells, and other sources as a signal of change in individual and population health [2]. These signals are mainly acquired from the secondary source of information, typically built for other purposes. However, to keep up the pace with the rapidly changing social and biological dynamics, novel outbreak detection mechanisms are highly sought [2].

The advancement and omnipresence of smartphones, Internet of Things (IoT) devices, wearables, and sensors have enabled individuals to easily self-record health-related events often for self-tracking or self-managing their disease [3,4]. The recent movement known as quantified self and lifelogging is the result of such technological advancement, where people collect various kinds of health-related events and data for personal informatics purposes, that is, self-surveillance and self-management [5-8]. To this end, people with diabetes are not an exception, where they self-record detailed information as part of their self-management, including blood glucose levels, diet and insulin intake, physical activity, medication, and other information [4,9,10]. Consequently, a huge amount of self-recorded, personal health-related data is generated each day that have great potential to be used as a secondary source of information for other purposes such as digital epidemiology [11,12]. According to recent reports, personal health data or self-collected health-related data have provided an enormous opportunity to enhance the possibility of detecting infection incidence during the presymptomatic stage (improved sensitivity and timeliness), specifically during the incubation period, where most of the existing systems neglect from their process [13].

Type 1 diabetes is a chronic condition of blood glucose metabolic disorder caused by lack of insulin secretion from pancreas cells [14]. These patient groups are recommended to maintain their blood glucose levels within a specified range through self-management practice [14,15]. Blood glucose levels are controlled by balancing insulin and meal intake along with other contexts such as physical activity, medications, and others. Blood glucose dynamics are affected by various factors that can be categorized as common, individual, and unpredictable factors [16]. These factors could be further categorized as patient-controllable and patient-uncontrollable parameters [17]. Patient-controllable parameters incorporate factors on which the patient has direct control and can roughly understand their immediate effect on blood glucose dynamics. However, patient-uncontrollable parameters include factors in which the patient does not have direct control and faces a challenge to understand their immediate effect on blood glucose levels. From the patient perspective, usually patient-controllable parameters induce reasonable deviations on blood glucose levels; however, patient-uncontrollable parameters induce unreasonable blood glucose deviations and usually differ from the usual norm of blood glucose dynamics [18]. The total number of people living with diabetes is increasing worldwide. According to recent reports [14], there were 415 million people between the ages of 20 and 79 years in 2015, and this value is projected to increase by 54% in 2040. From this figure, 5% are believed to have type 1 diabetes. In these patient groups, infection incidence often results in complications and difficulties in controlling blood glucose levels within the recommended range [19-21]. As a result, early detection of infection incidence among these patient groups could provide a way to assist the individual and at the same time can be used to realize a digital infectious disease detection system.

Currently, with the advancement of technology, the need to have a system that is able to detect infection incidence at the presymptomatic stage is highly sought [13]. In this regard, there are some previous investigations that have showcased the use of self-recorded data from people with diabetes as surveillance events (indicators) by uncovering the effect of infection incidence on blood glucose levels and glycemic control in real-life settings [18,22-36]. These studies reported the presence of prolonged hyperglycemia episodes as a result of infection incidence, thereby revealing the potential of self-recorded data as a secondary source of information for realizing a digital infectious disease detection system. For instance, Botsis et al [22] conducted a proof-of-concept study based on daily glycemic control data of 248 people with type 2 diabetes and concluded that blood glucose levels, insulin dosage, diet (carbohydrate consumption), physical activity, and other physiological parameters could be used as potential event indicators of infection incidence but calls for further investigations. Furthermore, Botsis et al [18] also reported elevated glycated hemoglobin (HbA_{1c}) levels after infections regardless of tight blood glucose control, which only settled down to normal levels after the patient recovered. Moreover, other studies conducted in hospital settings also reported similar results in this direction

[37,38]. Despite reporting the potential of using self-recorded data as a surveillance event indicator, none of these studies demonstrated the extent to which each parameter is affected at an individual level as a result of infection incidence. Therefore, the purpose of this study was to retrospectively analyze the effect of infection incidence at an individual level and pinpoint optimal parameters that can effectively be used as input variables for developing an infection detection algorithm, thereby illustrating how these patient groups can assist in detecting infectious disease outbreaks. Moreover, this study provides a general framework regarding how a digital infectious disease detection system can be designed using self-recorded data from people with type 1 diabetes as a secondary source of information. Furthermore, this sheds light on the possibility of assisting the individual during such an incident. To this end, we analyzed temporal trends and probability distributions of different diabetes profile parameters (ie, blood glucose, insulin, carbohydrate, and others) to uncover the effect of infection incidence on the blood glucose dynamics, thereby identifying parameters that can effectively be used as potential events (indicators) of infection incidence. In addition, a framework is presented depicting the necessary structure to properly exploit self-recorded data from these patient groups to realize a real-time digital infectious disease detection system. This paper is structured as follows: the Methods section describes the materials and methods used to analyze the data sets. The Results section presents the results depicting the effect of acute infection incidence in comparison with regular or normal situations. The Discussion section presents the overall findings and proposes a framework for designing and developing a real-time digital infectious disease detection system using self-recorded data from these patient groups. The final section of *Discussion* presents our concluding remarks.

Methods

Materials

High precision self-recorded data of 10 patient years collected from 3 real subjects (2 males and 1 female) with type 1 diabetes were used. The patients were free from any other chronic or other form of disease, except the self-reported acute infection incidence throughout the entire data collection period. The data sets consisted of blood glucose measurements (self-monitoring of blood glucose [SMBG] and continuous glucose monitoring [CGM]), injected insulin (basal and bolus), diet (carbohydrate in grams), and self-reported events of acute infection. The patients used different diabetes self-management technologies throughout the data collection period to gather these data sets including the Diabetes Diary app (Norwegian Centre for E-health Research) [39], the Spike app [40], the xDrip with app, Dexcom CGM, insulin pens, and insulin pumps, as shown in Table 1. The data sets consist of both normal years, without any significant acute infection incidence, and years with at least one or more acute infection incidence. The normal (without infection) patient years were used as a baseline to compare the effect of all patient-controllable parameters and patient-uncontrollable parameters against the self-reported incidence of acute infection. The self-reported incidences of acute infections were a case of influenza (flu) and mild and light common cold without fever. All the experiments and analyses were conducted using MATLAB version 2018a (Mathworks).

Table 1. Equipment used in diabetes self-management.

Patients	Self-management		
	BG[a]	Insulin administration	Diet
Subject 1	SMBG[b]—finger pricks recorded in the Diabetes Diary mobile app and Dexcom CGM[c]	Insulin pen (multiple bolus and one-time basal in the morning) recorded in the Diabetes Diary mobile app	Carbohydrate in grams recorded in the Diabetes Diary mobile app
Subject 2	SMBG—finger pricks recorded in the Spike mobile app and Dexcom G4 CGM	Insulin pen (multiple bolus [Humalog] and one-time basal [Toujeo] before bed) recorded in the Spike mobile app	Carbohydrate in grams recorded in the Spike mobile app
Subject 3	Enlite (Medtronic) CGM and Dexcom G4 CGM	Medtronic MinMed G640 insulin pump (basal rates profile [Fiasp] and multiple bolus [Fiasp])	Carbohydrate in grams recorded in pump information

[a]BG: blood glucose.

[b]SMBG: self-monitoring of blood glucose.

[c]CGM: continuous glucose monitoring.

Patient Characteristics

The participants were highly motivated individuals with type 1 diabetes who had advanced knowledge and understanding of several diabetes-related technologies. Hence, the self-recorded data can be regarded as highly precise and accurate. All the participants had advanced knowledge of carbohydrate counting, which can be considered as level 3 (advanced) [41]. The long-term average HbA_{1c} and characteristics of the participants are given in Table 2.

Table 2. Participants characteristics.

Variables	Values
Gender, n	
Male	2
Female	1
Age (years), mean (SD)	34 (13.2)
Body weight (kg)	
Subject 1	83
Subject 2	77
Subject 3	70
HbA_{1c}[a](%)	
Subject 1	6.0
Subject 2	7.3
Subject 3	6.2
Carbohydrate counting	Level 3 (advanced)

[a]HbA_{1c}: glycated hemoglobin.

Data Collection and Ethics

The study protocol has been submitted to the Norwegian Regional Committees for Medical Health Research Ethics Northern Norway (REK) for evaluation and was found exempted from regional ethics review because it resides outside of the scope of medical research (reference number: 108435). Written consent was obtained and the participants donated the data sets. All data from the participants were anonymized.

Approaches

We retrospectively assessed and analyzed the diabetes profile (blood glucose, insulin, carbohydrate, and insulin-to-carbohydrate ratio) to uncover the nature, size, and shape of the infection-induced shift in the operating region of the blood glucose dynamics. A data size of 10 patient years incorporating blood glucose levels (SMBG and CGM), insulin (bolus and basal), diet (carbohydrate in grams), and self-reported events of acute infection was used. The analysis was performed based on specified timeframes (weekly, daily, and hourly) to reveal the effect of acute infection development on blood glucose dynamics. The data set incorporates 5 normal patient years without any infection incidence and 5 patient years each with at least one case of self-reported incidence of acute infection. Normal patient years were used as a baseline for comparison purposes. We analyzed the temporal evolution and probability distribution of blood glucose levels, injected insulin, carbohydrate intake (grams), and insulin-to-carbohydrate ratio within the stated timeframe. For the daily and hourly timeframes, a moving-average filter and nonparametric density estimation techniques, the kernel density estimator, were used to analyze the trend and data distribution before, during, and after the infection incidence. A moving-average filter with a window size of 2 days was employed to remove fast timescale features through smoothing. The window size includes $N-1$ observations from the previous data points and the current data point, where N is the window size. Generally, the window size of a moving-average filter is determined based on complementary issues of better smoothing and the cost of significant delay (shift) incurred [42,43]. A small window size often generates less delay (shift) but at the cost of more short-term features and having a larger window size will smoothen the data in a better manner but at the cost of significant delay in the timeliness of detecting the infection incidence. Therefore, the window size was determined based on these complementary issues, and more importance was given to minimize the inherent delay (shift) incurred due to the window size. To this end, window sizes of 1, 2, 3, and 4 days were applied and tested to choose the optimal size of the window, and as a result, a window size of 2 days was found to be satisfactory. The preinfection, infection, and postinfection week analyses were carried out on the raw data set based on the week's daily average and SD of blood glucose levels and daily sum and SD of insulin and carbohydrate. A statistical boxplot was used to depict the comparison during preinfection, infection, and postinfection weeks.

Data Resampling, Imputation, and Preprocessing

The features of the self-collected data from individuals with type 1 diabetes are shown in Table 3. The raw data were resampled at a uniform rate by assigning each measurement into the nearest time-bin based on its time stamp. Generally, whenever there is more than one measurement within each time-bin, the measurements are combined into a single measurement by either summing or averaging the elements. For blood glucose levels (both CGM and SMBG), the measurements were averaged into their respective sampling time-bins. However, regarding carbohydrate consumption and insulin injections, the sum of the elements in their respective sampling time-bin was computed, as shown in Table 4. In each time-bin, the effect of total insulin and total carbohydrate on the average blood glucose level was considered. The resampled data were further preprocessed using a moving-average filter with a 2-day (48-hour) window size to capture only the important patterns—long-term variation, while filtering and smoothing

local and short-term variations. Moreover, for narrower time-bin resampling, for example, an hour, there are more frequent zeros of measurement, especially for carbohydrate and insulin measurements, which poses a significant challenge to compute the insulin-to-carbohydrate ratio as the ratio goes to infinity given that the carbohydrate amount is zero. Therefore, in such cases of a narrower time-bin, the ratio was computed only after computing the moving-average value of insulin and carbohydrate based on a window size of 48 hours. Regarding the missing blood glucose values during the hourly computations, a cubic spline interpolation was used to estimate the missing values.

Table 3. Self-collected user data.

Variable names	Subject's record variables	
	Description	Units
	Continuous glucose reading	mg/dL
	Self-management blood glucose reading	mg/dL
	Injected insulin (bolus)	Units
	Injected insulin (basal)	Units
	Ingested carbohydrate	Grams

Table 4. Data preprocessing.

Variable name	Preprocessed variables	
	Description	Units
	Average continuous glucose reading	mg/dL
	Average self-management blood glucose reading	mg/dL
	Sum injected insulin (bolus)	Units
	Sum injected insulin (basal)	Units
	Sum ingested carbohydrate	Grams
	Ratio of insulin (bolus) to carbohydrate	Units/grams
	Ratio of insulin (basal) to carbohydrate	Units/grams

Kernel Density Estimation

Nonparametric density estimation is an alternative to the parametric approach, which involves specifying a model using a number of parameters that can be estimated through the likelihood principle [44,45]. In this study, we used kernel density estimation techniques [46-48] to estimate the probability distribution of the diabetes profile key parameters to uncover the deviation incurred by the acute infection incidence. In this regard, both univariate and bivariate kernel density estimators are used to assess and analyze the insulin-to-carbohydrate ratio (univariate) and blood glucose levels along with the insulin-to-carbohydrate ratio (bivariate), respectively. An adaptive kernel density estimator with a Gaussian kernel was used in both cases**.** For the univariate kernel density estimator [49], bandwidth selection is based on the suggestion from Botev et al [44], which is a data-driven and plug-in bandwidth selector that does not use normal reference rules. For the bivariate estimator, a rule-of-thumb bandwidth selection suggested by Bowman et al [50,51] was used to determine the appropriate bandwidth [52]. These computations are carried out based on the procedures given in Textboxes 1 and 2.

Textbox 1. One-dimensional adaptive kernel density estimation.

Approach: one-dimensional adaptive kernel density estimation

- Given: time series data sets of the insulin-to-carbohydrate ratio X ε D and an adaptive kernel density estimator *M – one – dimensional*
- Remove the reported days of infection from the time series data sets D and form a new data set *X* ε Q
- Compute the one dimensional density based on the kernel density estimator *M* using D and Q
- Compare the distribution from *M*

Textbox 2. Two-dimensional adaptive kernel density estimation.

Approach: two-dimensional adaptive kernel density estimation

- Given: time series data sets of blood glucose level and the insulin-to-carbohydrate ratio X, Y ε D and an adaptive kernel density estimator N – *two - dimensional*
- Remove the reported days of infection from the time series data sets D and form a new data set *X, Y* ε Q
- Compute the two-dimensional density based on the kernel density estimator N using D and Q
- Compare the distribution from N

Results

Overview

The analysis was conducted based on an hourly, daily, and weekly basis to reveal the deviations incurred due to the infection incidence. A total of 10 patient years were analyzed, and 5 of these years were found to include at least one incidence of acute infection lasting around 1-2 weeks. The proposed approach is designed to smooth out short-duration variations and include the 2 major patient-controllable factors, insulin and diet intake. Normal patient years were used to compare the effect of all patient-controllable parameters and patient-uncontrollable parameters against the self-reported incidence of acute infection. The trend analysis for both the normal patient years and patient years with acute infections using the proposed approach is presented below along with the nonparametric probability distribution. The weekly mean deviations of key diabetes parameters (blood glucose, insulin, and diet) during the preinfection, infection, and postinfection weeks are given in Multimedia Appendix 1.

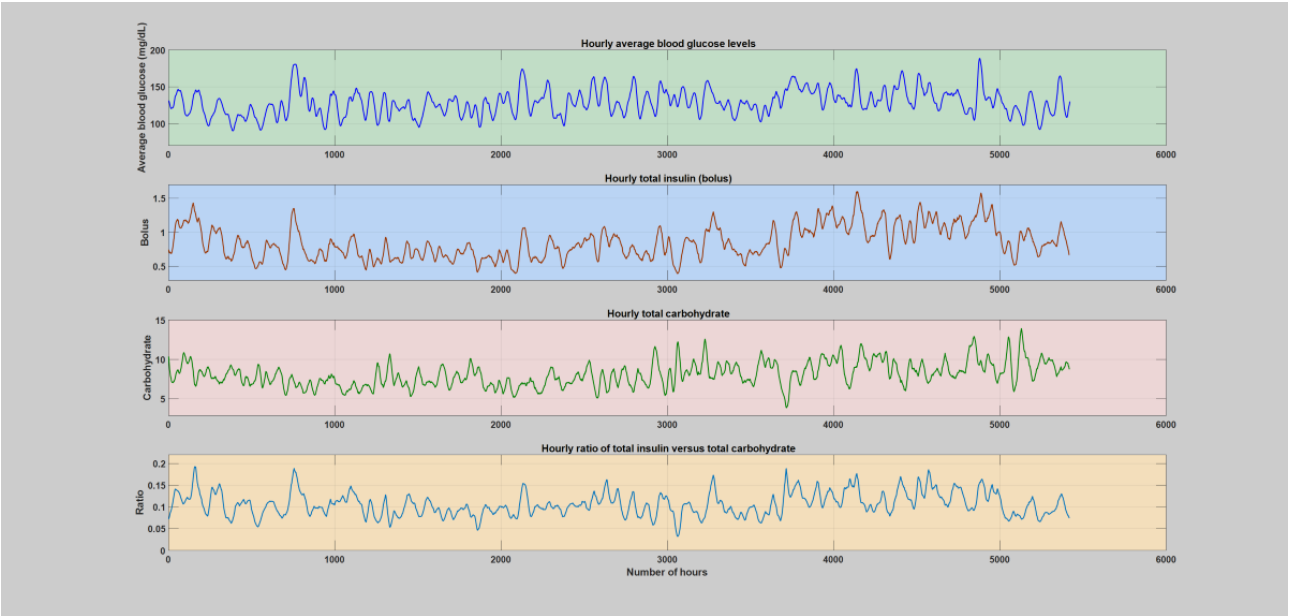
Trend Analysis

Trend Comparison for Normal Patient Years

During normal years when patients do not have any significant illness or infections (Multimedia Appendix 2), the insulin-to-carbohydrate ratio follows a similar trend in all the subjects, where the insulin-to-carbohydrate ratio lies between 0.05 and 0.2. An elaborate analytical plot of a typical patient year without infection incidence showing the phenomena is depicted in Figures 1 and 2. A detailed analytical plot of the 5 patient years depicting the same phenomena can be found in Multimedia Appendix 2. The insulin-to-carbohydrate ratio conveys interesting information about the usual operating point of the patient, depicting the necessary amount of insulin (bolus) required for every gram of carbohydrate consumed to maintain the blood glucose levels within a healthy range (typically recommended to be between 70 and 180 mg/dL). As can be seen from the yearlong trend analysis of the regular or normal patient years (Multimedia Appendix 2), despite the presence of various factors that are known to disturb blood glucose dynamics, both patient-controllable parameters and patient-uncontrollable parameters except infection incidence, the insulin-to-carbohydrate ratio remains to be relatively stable.

Figure 1. The first patient year, where there is no incidence of acute infections. The figure depicts the daily variation of average blood glucose levels, total insulin (bolus), total carbohydrate, and total insulin-to-total carbohydrate ratio. The operating point of the patient's insulin-to-carbohydrate ratio through these regular or normal days is between 0.05-0.2.

Figure 2. The first patient year, where there is no incidence of acute infections. The figure depicts variation of average blood glucose levels, total insulin (bolus), total carbohydrate, and total insulin-to-carbohydrate ratio during each hours of the day. The operating point of the patient's insulin-to-carbohydrate ratio through these regular or normal hours is between 0.05-0.2.



Trend Comparison of Patient Years With Acute Infection

The trend analysis of the key diabetes parameters, blood glucose, insulin, and carbohydrate, during acute infection suggests that there is a dramatic shift in the evolution of blood glucose, insulin, and carbohydrate (for detailed information, see Multimedia Appendices 1 and 3). Infection incidence brought about a dramatic increase in blood glucose levels, insulin intake, and reduction in carbohydrate consumption. The detailed analysis and the shift incurred on a weekly, daily, and hourly basis are presented in the following section.

Weekly Analysis

The weekly analysis of the patient years was conducted by analyzing the deviation incurred on the key parameters of the blood glucose dynamics during the infection week in comparison with before and after the infection incidence. The raw data were used to estimate the deviations incurred due to infection incidence. The mean and SD of blood glucose levels, total insulin (bolus), and total carbohydrate were computed and used for comparison of the infection-induced deviations. As shown in Figures 3-5 and Table 5, in all the infection cases, the weekly analysis demonstrated that blood glucose levels were elevated despite higher insulin injection and reduced carbohydrate consumption. In all of these cases, it is clear that the incidence

of infection has brought unreasonable deviation, with respect to the patient-controllable parameters, in the operation of the overall blood glucose dynamics as compared with the usual norm of the blood glucose dynamics. The presence of elevated blood glucose levels in the infection week, regardless of the high amount of insulin injections and lower carbohydrate consumption, clearly violated the norm of the blood glucose dynamics, where during normal situations the blood glucose levels are expected to drop with high insulin and reduced carbohydrate consumption. The fact that the blood glucose remains elevated during the infection incidence despite higher insulin injections and low carbohydrate consumption is highly associated with the infection phenomenon, which enhances the production of glucose and increased insulin resistance within the body to deliver more energy for the body to fight the pathogens. A more detailed description of the weekly analysis can be found in Multimedia Appendix 1.

Figure 3. Analysis of blood glucose levels during the preinfection week, infection week, and postinfection week based on the first case of infection (flu). The asterisk shows the mean value, and the red line depicts the median value for the week.

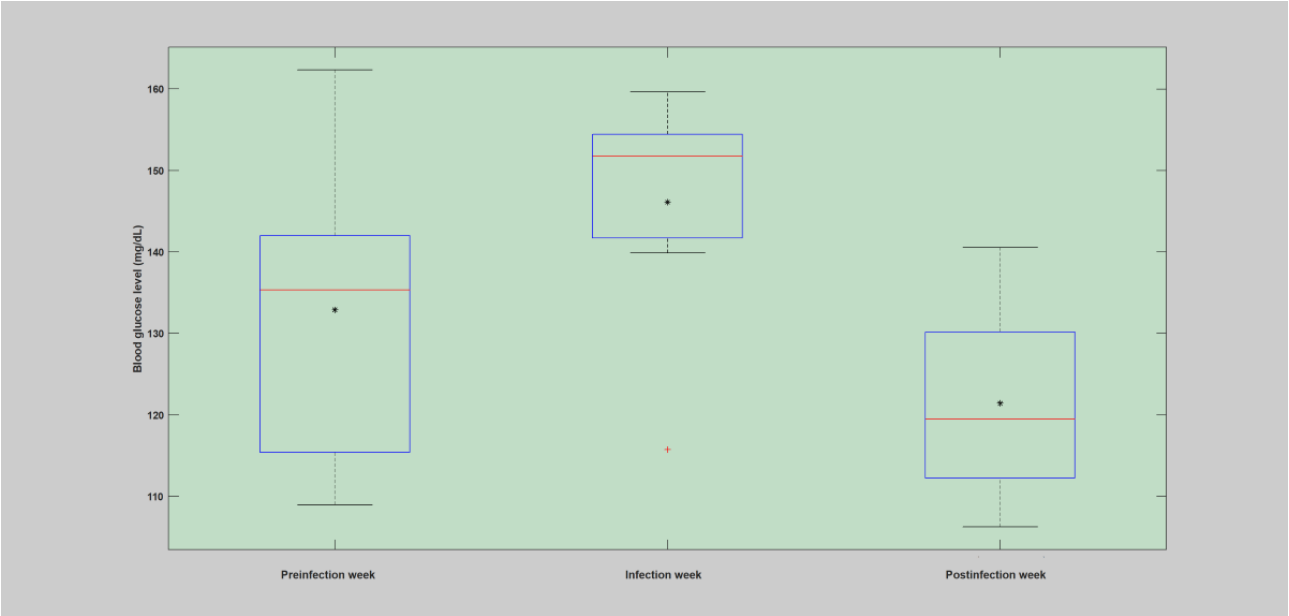


Figure 4. Analysis of total insulin (bolus) intake during preinfection week, infection week, and postinfection week based on the first case of infection (flu). The asterisk shows the mean value, and the red line depicts the median value for the week.

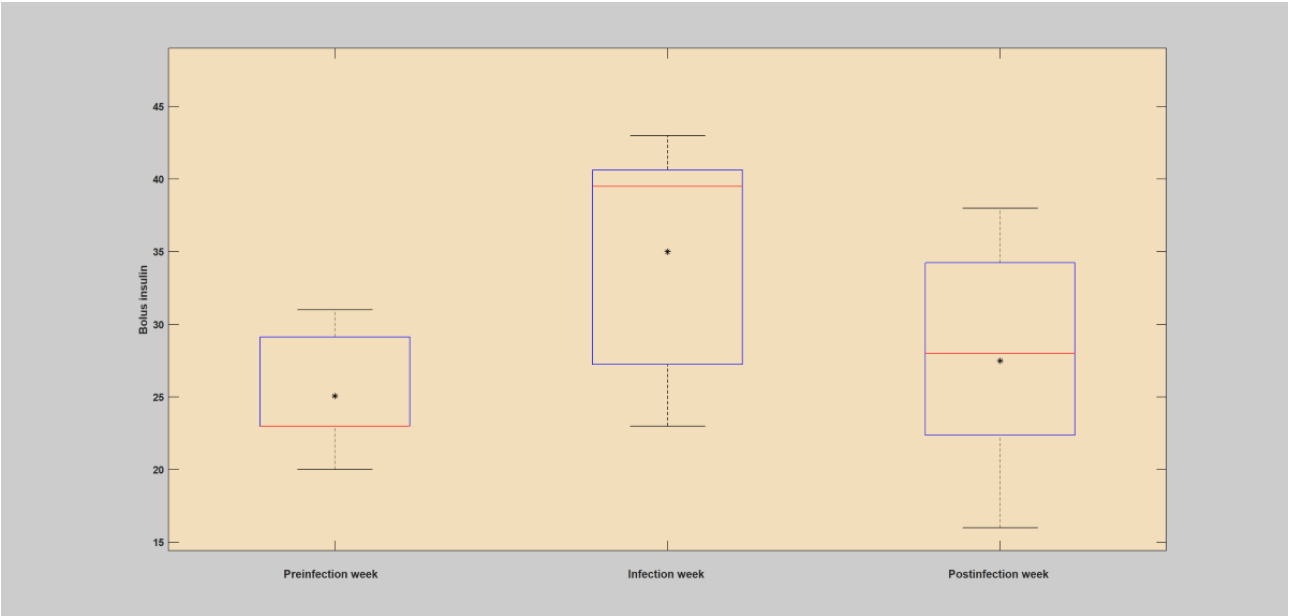


Figure 5. Analysis of total carbohydrate (grams) intake during preinfection week, infection week, and postinfection week based on the first case of infection (flu). The asterisk shows the mean value, and the red line depicts the median value for the week.

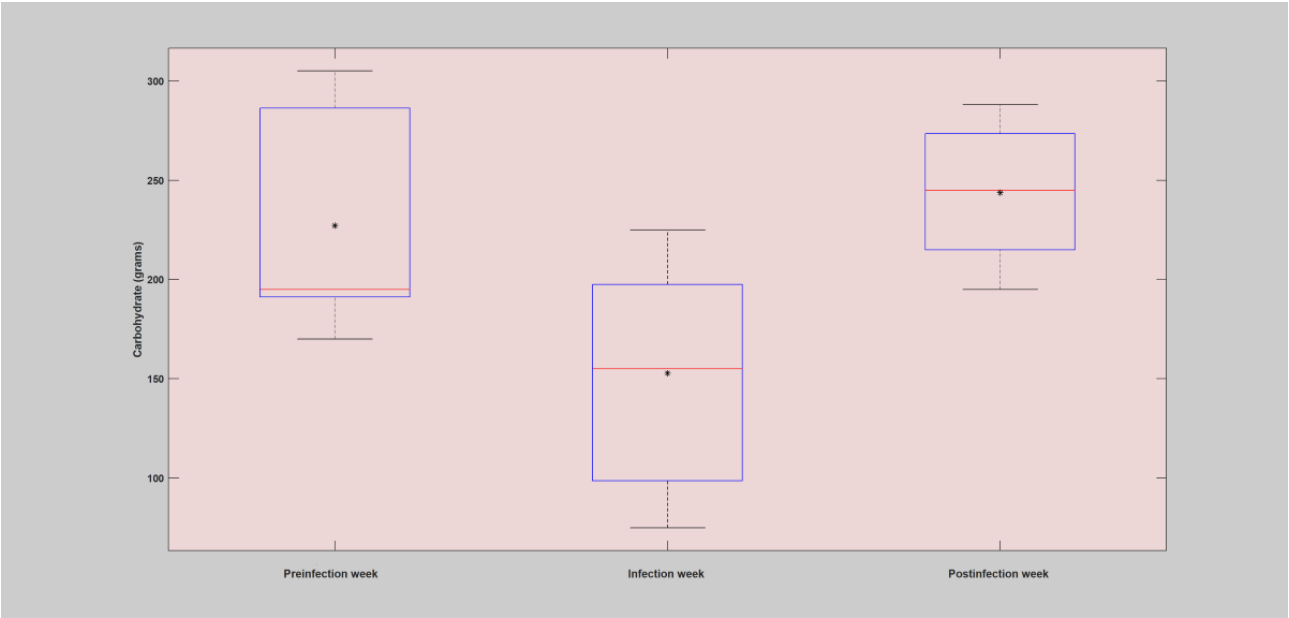


Table 5. Mean and standard deviation of blood glucose levels, total insulin (bolus), and total carbohydrate during the preinfection week, infection week, and postinfection week.

Parameters	Preinfection week, mean (SD)	Infection week, mean (SD)	Postinfection week, mean (SD)
The first case of infection (flu)			
BG[a] (mg/dL)	130.74 (16.89)	141.95 (14.37)	119.16 (7.39)
Total insulin (bolus)	23.39 (4.91)	35.30 (6.11)	21.32 (4.61)
Carbohydrate (grams)	241.11 (57.27)	178.80 (65.69)	241.18 (37.63)
The second case of infection (flu)			
BG (mg/dL)	143.01 (19.53)	155.36 (21.99)	126.17 (11.70)
Total insulin (bolus)	28.07 (8.85)	41.07 (9.44)	25.36 (6.93)
Carbohydrate (grams)	190.14 (43.93)	161.14 (58.43)	214.57 (34.66)
The third case of infection (flu)			
BG (mg/dL)	136.93 (18.58)	144.12 (20.30)	134.18 (11.96)
Total insulin (bolus)	20.08 (5.44)	31.50 (10.84)	22.83 (3.86)
Carbohydrate (grams)	178.0 (45.87)	144.83 (37.63)	195.83 (42.59)
The fourth case of infection (flu)			
BG (mg/dL)	157.74 (31.12)	161.34 (19.88)	138.57 (19.83)
Total insulin (bolus)	24.43 (5.26)	32.14 (7.01)	29.29 (5.22)
Carbohydrate (grams)	199.06 (53.45)	167.04 (44.94)	226.07 (18.23)
The fifth case of infection (flu)			
BG (mg/dL)	135.21 (14.58)	139.88 (15.54)	122.87 (14.49)
Insulin (bolus)	32.80 (4.59)	40.37 (8.31)	33.36 (7.94)
Insulin (basal)	19.20 (1.21)	20.42 (2.06)	18.68 (1.56)
Total insulin	52.33 (5.14)	61.21 (8.26)	52.46 (8.47)

[a]BG: blood glucose.

Blood Glucose Levels

In all these infection incidences, the individual blood glucose levels remain elevated for a prolonged period of time despite low carbohydrate consumption and increased insulin injections as compared with the regular or normal days. Blood glucose levels were elevated during the infection week as compared with the preinfection and postinfection weeks.

- During the first case of infection, the overall mean percentage increase in the infection week's blood glucose levels was 8.57% over the preinfection week and 19.12% over the postinfection week, as shown in Table 5.
- During the second case of infection, the overall mean percentage increase in the infection week's blood glucose levels was 8.63% over the preinfection week and 23.13% over the postinfection week, as shown in Table 5.
- During the third case of infection, the overall mean percentage increase in the infection week's blood glucose levels was 7.26% over the preinfection week and 7.41% over the postinfection week, as shown in Table 5.
- During the fourth case of infection, the overall mean percentage increase in the infection week's blood glucose levels was 2.28% over the preinfection week and 16.43% over the postinfection week, as shown in Table 5.
- During the fifth case of infection, the overall mean percentage increase in the infection week's blood glucose levels was 3.45% over the preinfection week and 13.84% over the postinfection week, as shown in Table 5.

Insulin Intake

The comparison of infection week insulin injections with preinfection and postinfection weeks revealed that there was a dramatic increase in the amount of insulin intake during the infection period.

- During the first case of infection, the overall mean percentage increase in the infection week's insulin (bolus) injection was 50.93% over the preinfection week and 65.59% over the postinfection week, as shown in Table 5.
- During the second case of infection, the overall mean percentage increase in the infection week's insulin (bolus) injection was 46.31% over the preinfection week and 61.94% over the postinfection week, as shown in Table 5.
- During the third case of infection, the overall mean percentage increase in the infection week's insulin (bolus) injection was 56.87% over the preinfection week and 37.98% over the postinfection week, as shown in Table 5.
- During the fourth case of infection, the overall mean percentage increase in the infection week's insulin (bolus) injection was 31.56% over the preinfection week and 9.7% over the postinfection week, as shown in Table 5.
- During the fifth case of infection, the overall mean percentage increase in the infection week's insulin (bolus) injection was 23.08% over the preinfection week and 21.01% over the postinfection week, as shown in Table 5.

Carbohydrate Consumption

Comparison of the amount of carbohydrate consumption during the infection week with the preinfection and postinfection weeks revealed that there was a significant reduction during the infection period.

- During the first case of infection, the overall mean percentage reduction in the infection week's carbohydrate consumption was 25.84% below the preinfection week and 25.87% below the postinfection week, as shown in Table 5.
- During the second case of infection, the overall mean percentage reduction in the infection week's carbohydrate consumption was 15.25% below the preinfection week and 24.90% below the postinfection week, as shown in Table 5.
- During the third case of infection, the overall mean percentage increase in the infection week's carbohydrate consumption was 18.63% below the preinfection week and 26.04% below the postinfection week, as shown in Table 5.
- During the fourth case of infection, the overall mean percentage increase in the infection week's carbohydrate consumption was 16.09% below the preinfection week and 35.34% below the postinfection week, as shown in Table 5.

Insulin-to-Carbohydrate Ratio

The insulin-to-carbohydrate ratio defines the amount of insulin a patient needs to take for every gram of carbohydrate consumed. The value of the insulin-to-carbohydrate ratio usually lies between 0.05 and 0.2 on normal occasions. However, it has dramatically increased upon the incidence of infection.

- During the first case of infection, the overall mean percentage increase in the infection week's insulin-to-carbohydrate ratio was around 125.84% above the normal operating point of the patient, as shown in Table 5.
- During the second case of infection, the overall mean percentage increase in the infection week's insulin-to-carbohydrate ratio was approximately 144.43% above the normal operating point of the patient, as shown in Table 5.
- During the first case of infection, the overall mean percentage increase in the infection week's insulin-to-carbohydrate ratio was around 93.75% above the normal operating point of the patient, as shown in Table 5.
- During the fourth case of infection, the overall mean percentage increase in the infection week's insulin-to-carbohydrate ratio was approximately 70.84% above the normal operating point of the patient, as shown in Table 5.

Daily and Hourly Analysis

Hourly and daily analyses were conducted by analyzing the deviations incurred on the key diabetes parameters, blood glucose levels, insulin, carbohydrate, and the insulin-to-carbohydrate ratio as a result of infection incidence in contrast to the whole patient year. The comparison was carried out based on the smoothed version of the data, that is, 2 days window moving-average filter. Similar to the weekly analysis, the infection-induced shift of the blood glucose dynamics, that

is, higher glucose production and increased insulin resistance, is clearly shown in both the daily and hourly analyses. As can be seen in Figures 6-11, the insulin-to-carbohydrate ratio of the patient has drastically shifted to a higher value to account for the effect of increased glucose production and insulin resistance (see Multimedia Appendix 3 for a detailed plot of the hourly analysis in all the infection cases). In all of these cases, the insulin-to-carbohydrate ratio increases from the usual values of 0.05 to 0.2 during the normal period to higher values reaching 0.6, depending on the degree of severity of the infection incidence, type of pathogens involved, and the individual immunity.

Figure 6. Daily analysis of the first infection case (flu). The figure depicts variation of average blood glucose levels, total insulin (bolus), total carbohydrate, and total insulin-to-total carbohydrate ratio. The operating point of the patient's insulin-to-carbohydrate ratio had dramatically shifted and raised above the regular or normal days and reach a top around 0.5 upon midinfection week.

Figure 7. Hourly analysis of the first infection case (flu). The figure depicts variation of average blood glucose levels, total insulin (bolus), total carbohydrate, and total insulin-to-total carbohydrate ratio. The operating point of the patient's insulin-to-carbohydrate ratio had dramatically shifted and raised above the regular or normal days and reach a top around 0.5 upon midinfection week.

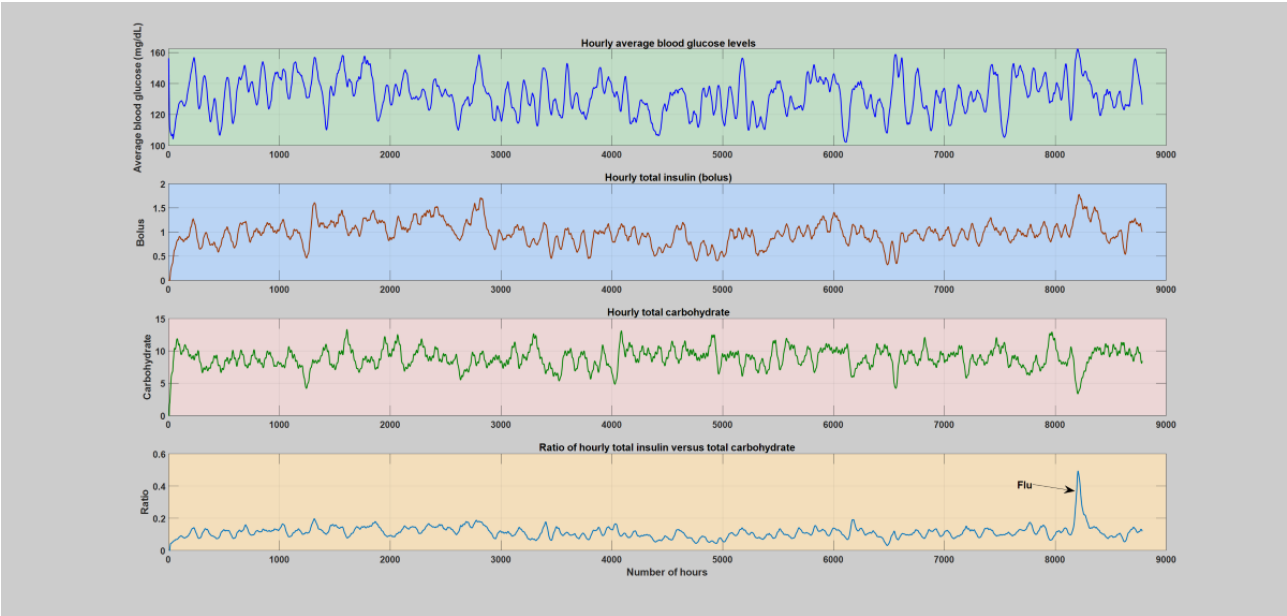


Figure 8. Daily analysis of the second infection case (flu). The figure depicts variation of average blood glucose levels, total insulin (bolus), total carbohydrate, and total insulin-to-total carbohydrate ratio. The operating point of the patient's insulin-to-carbohydrate ratio had dramatically shifted and raised above the regular or normal days and reach a top around 0.45 upon midinfection week.

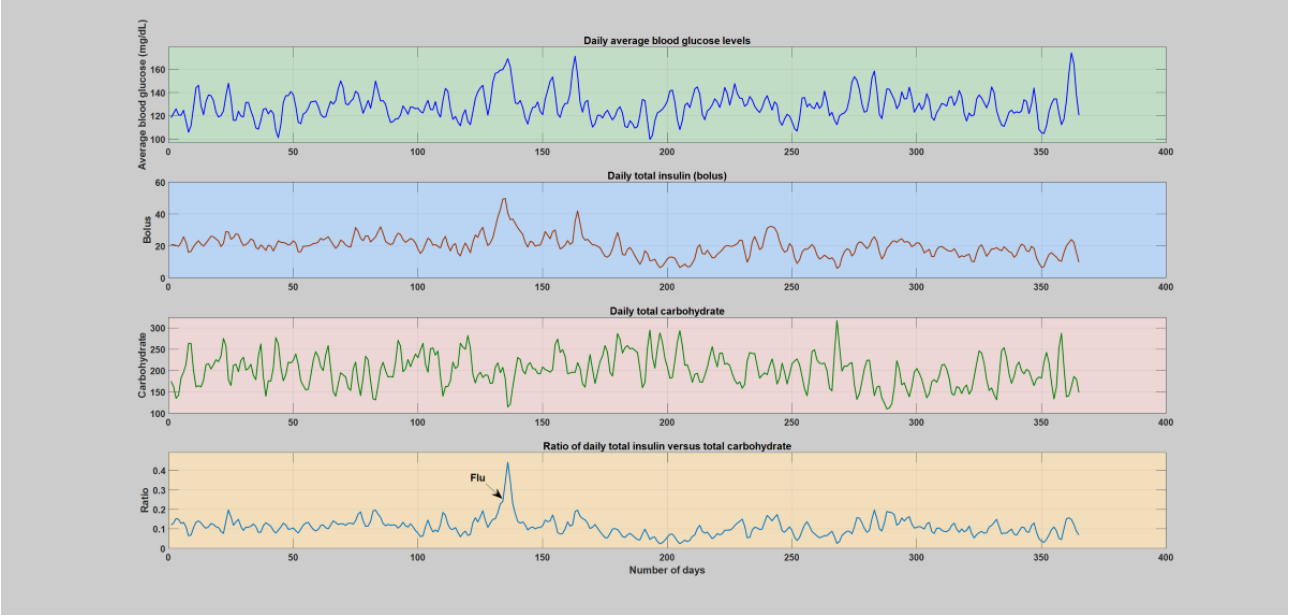


Figure 9. Daily analysis of the third infection case (flu). The figure depicts variation of average blood glucose levels, total insulin (bolus), total carbohydrate, and total insulin-to-total carbohydrate ratio. The operating point of the patient's insulin-to-carbohydrate ratio had dramatically shifted and raised above the regular or normal days and topped around 0.4 upon midinfection week.

Figure 10. Daily analysis of the fourth infection case (mild common cold without fever, light common cold without fever, and flu). The figure depicts variation of average blood glucose levels, total insulin (bolus), total carbohydrate, and total insulin-to-total carbohydrate ratio. The operating point of the patient's insulin-to-carbohydrate ratio had dramatically shifted and raised above the regular or normal days and reach a top around 0.28 upon midinfection week. A light common cold without fever seems to not significantly affect the operating point.

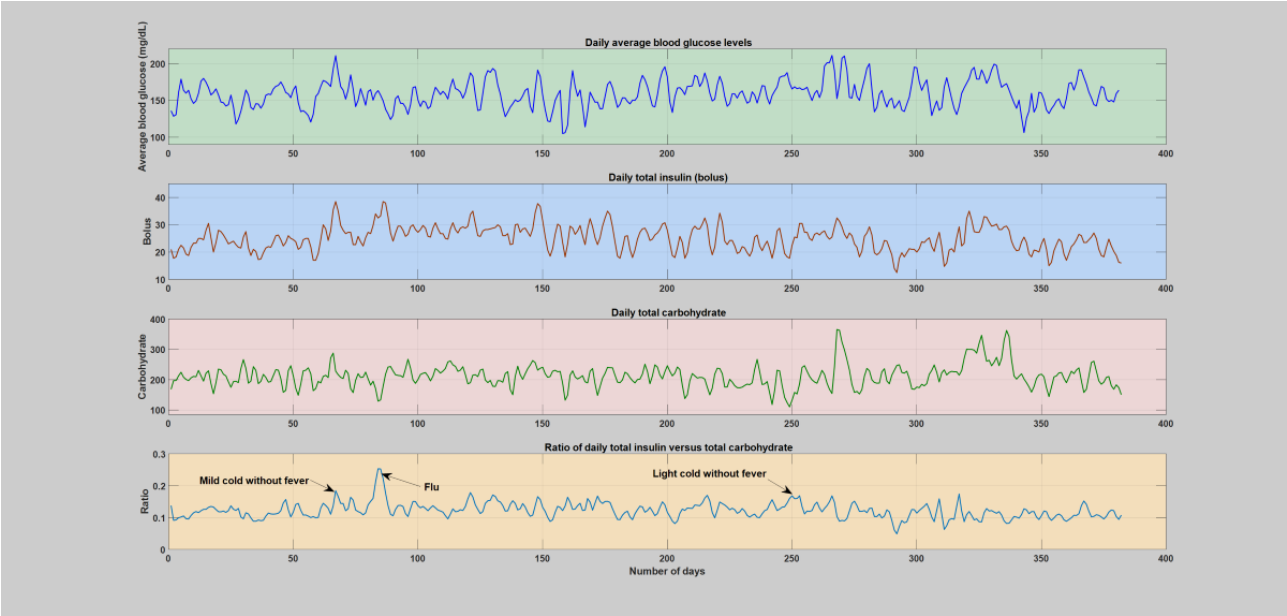
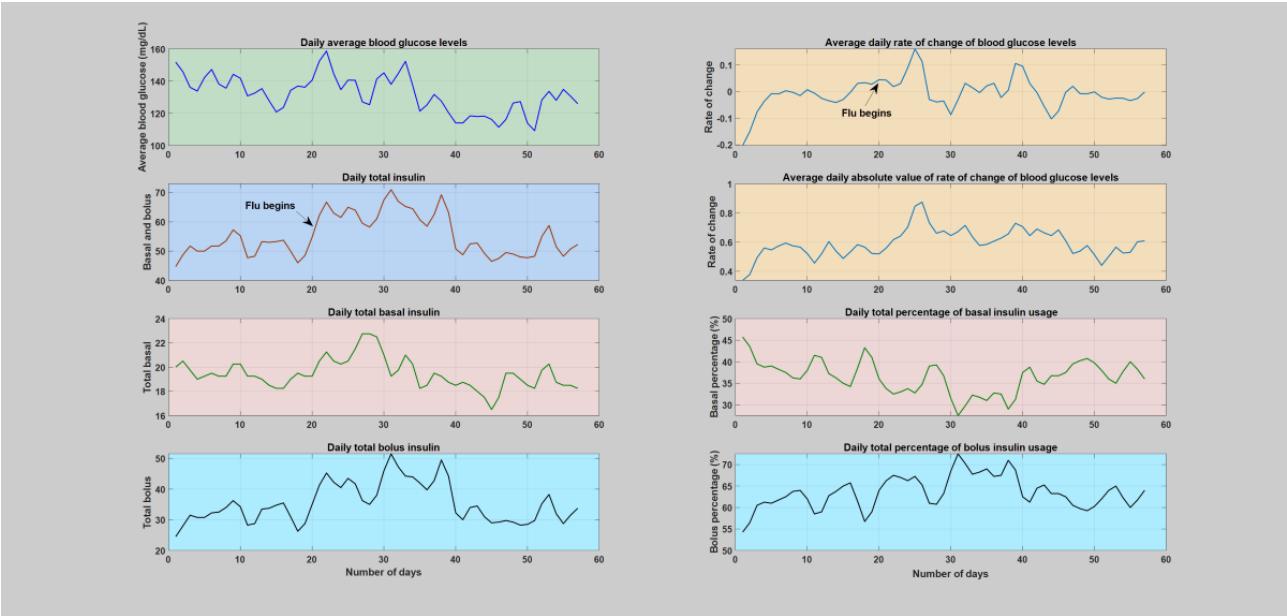


Figure 11. Daily analysis of the fifth infection case (flu). The figure depicts variation of average blood glucose levels, total insulin including both bolus and basal insulin, daily average rate of change of CGM and, absolute value of rate of change of CGM (computed based on CGM direction from the pump information), percentage of basal and bolus per total insulin units. As can be seen, as a result of the ongoing infection incidence, there is clear and dramatic rise in the amount of insulin, while blood glucose levels remain elevated.



Kernel Density Estimation–Probability Distribution

Kernel density was estimated to study and characterize the nature, shape, and degree of severity of the deviations incurred due to infection incidence by analyzing the probability distribution of the individual key parameters of the blood glucose dynamics. A univariate and bivariate kernel density estimation based on the insulin-to-carbohydrate ratio and blood glucose levels was carried out on the yearlong data, as shown in Figures 12 and 13 (a detailed plot for all the infection cases, both hourly and daily, can be found in Multimedia Appendix 1). As can be seen from the figures, the infection incidence has brought a significant change in the probability distribution. However, the nature, shape, and degree of outlierness depend on the type of pathogen involved, severity of infection, and individual immunity.

Figure 12. Univariate kernel density estimation of a patient year using the daily insulin-to-carbohydrate ratio. As can be seen from the tail of the distribution, during regular or normal days (the green shaded region), the yearly distribution of the patient's insulin-to-carbohydrate ratio lies within the values of 0.005 and 0.2. However, during infection incidence (the red shaded region), there is a clear deviation in the tail of the distribution, where the values reaches around 0.58.

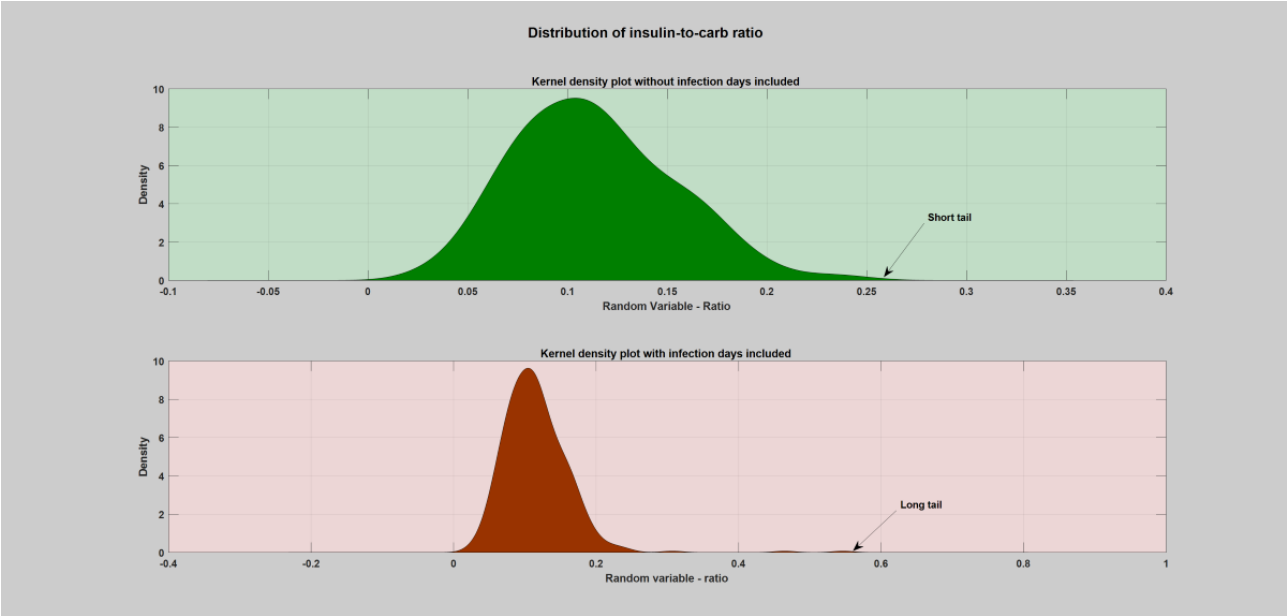
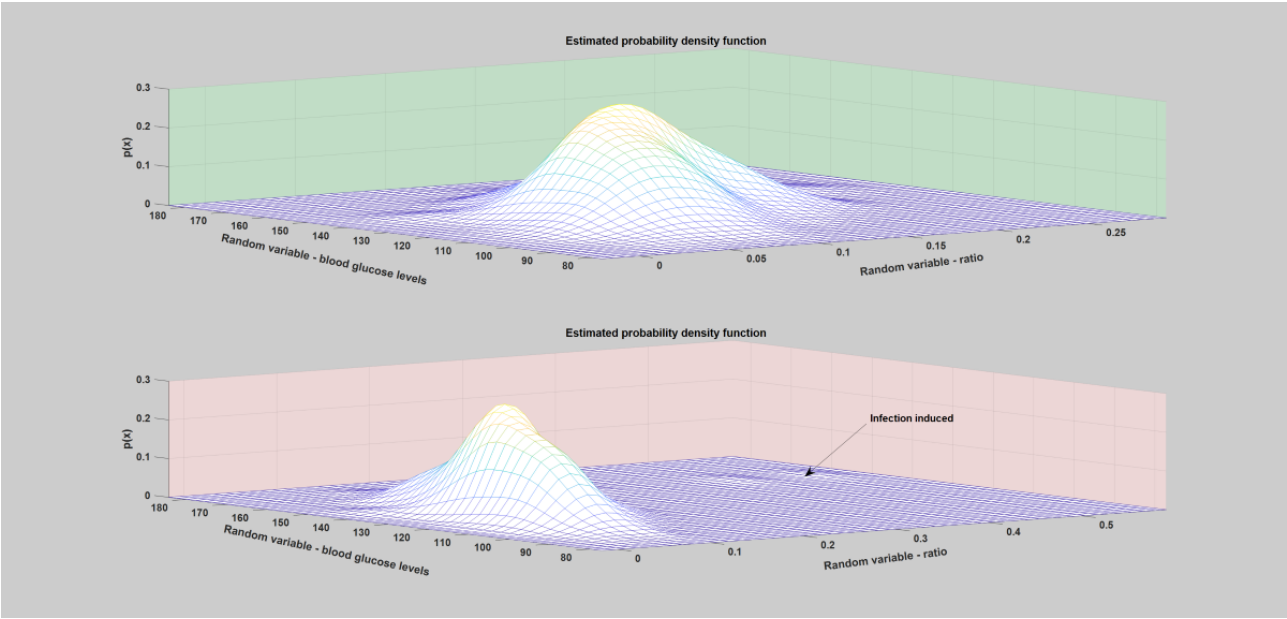


Figure 13. Bivariate kernel density estimation of a patient year using both the daily average blood glucose levels and insulin-to-carbohydrate ratio. As can be seen from the bivariate distribution, during regular or normal days (the top light green figure), the distributions are concentrated around the high density regions. However, during infection incidence (the lower figure), there is a clear bump far from the high density regions.



Discussion

Principal Findings

Presently, in relation to people's mobility and travel, there is a growing concern regarding an infectious disease outbreak. Such an incident can be a menace to our global health security, which calls for early detection and immediate response. Thus, there is a growing need for new approaches and technologies to upgrade the existing surveillance system for early detection of emerging infectious diseases [1]. Existing disease surveillance systems detect the incidence of outbreaks long after the incidence of the first symptoms. Therefore, the purpose of this study was to demonstrate how people with type 1 diabetes can assist in outbreak detection and further to shed light upon the possibility of assisting the individual during such an incident.

The advancement and omnipresence of smartphones, IoT devices, wearables, and sensors have enabled individuals to easily self-record health-related events often for self-tracking or self-managing their disease [5,6,53]. People with diabetes self-record detailed information including blood glucose levels, diet and insulin intake, physical activity, medication, and other parameters. The presence of such large self-recorded health data presents an opportunity to be used as a secondary source of information for other purposes such as digital epidemiology

and decision support applications. According to recent reports, the use of personal health information or self-collected data could mitigate the possibility of detecting infection incidence during the presymptomatic stage (improved sensitivity and timeliness), specifically during the incubation period, of which most of the current systems neglect from their process [13]. Our findings demonstrated that upon infection incidence, there is a dramatic shift in the operating point of the individual's blood glucose dynamics, which clearly violates the usual norm of blood glucose dynamics. During regular or normal days, blood glucose levels usually decrease when there is a significant increase in insulin injection and reduction in carbohydrate consumption. However, in all of the infection cases we analyzed, compared with the preinfection and postinfection weeks, the following were noticed:

- Blood glucose levels were elevated by an average of 6.1% and 16% over the preinfection and postinfection weeks, respectively.
- Insulin injection (bolus) increased by 42% and 39.3% over preinfection and postinfection weeks, respectively.
- Carbohydrate consumption was reduced by 19% and 28.1% compared with preinfection and postinfection weeks, respectively.
- The insulin-to-carbohydrate ratio increased by 108.7% on average in all cases.

In general, all of these findings confirm that during infection incidence, blood glucose levels are elevated despite injecting higher amounts of insulin and reduced carbohydrate consumption. The identified changes are quite significant anomalies compared with the regular or normal days and could potentially be detected with a dedicated personalized (individualized) computational health model. Various algorithms that span from prediction models to anomaly detection algorithms can be investigated to detect such infection-induced changes in blood glucose dynamics. Apart from the potential use of these findings in personalized digital infectious disease detection systems, it could also be used for decision support in self-management during infection and illness. As presented earlier, during the course of infection, individuals with diabetes usually struggle with severe hyperglycemia. Managing blood glucose levels during infection incidence is not an easy task, given the fact that it is caused by a mixed effect of both patient-controllable and patient-uncontrollable parameters. The patient can only estimate the disturbance caused by the amount of carbohydrate consumption, insulin injection, and physical activity load, which is not the case during infection incidence. Apart from these known major factors, that is, patient-controllable parameters, there is an underlying and unknown disturbance caused by the patient's uncontrollable parameters, such as counterregulatory hormones (CRHs), as a result of infection incidence. This unknown disturbance mainly increases glucose production from the liver and reduces insulin sensitivity. To this end, people with type 1 diabetes face a very difficult challenge to estimate the necessary amount of insulin for a given amount of carbohydrate consumption. In this regard, providing real-time decision support could reduce the burden during such a crisis. One possible approach could be characterizing the effect of different pathogens on blood glucose dynamics, mainly on insulin resistance and its sensitivity change over the course of infection. However, a large set of infection-related self-recorded data need to be analyzed for investigating how each pathogen affects the key parameters of blood glucose dynamics during the entire course of infection. This requires collecting and analyzing infection-related data, and estimating the overall changes each pathogen could bring on insulin sensitivity during the course of infection. To this end, the presented result reflects a promising result that can be geared toward decision support during infection or illness. For example, the change in insulin–to-carbohydrate ratio can be used to provide general information related to each pathogen on what to expect, such as the percentage of insulin resistance during the first days, in the middle, and at the final days of the infection.

Infection-Induced Shift of Operating Point in Blood Glucose Dynamics

During infection incidence, people with diabetes usually struggle with severe hyperglycemia and critical hypoglycemia if not properly managed. However, during regular or normal days, the patient can manage the incidence of hyperglycemia, which is mostly diet-induced, by properly controlling the patient-controllable parameters, for example, amount of carbohydrate consumption, insulin injection, and performing balanced physical activity or exercise. Yet, during infection incidence, it turns out to be very difficult to manage the hyperglycemia incidence due to the fact that it is caused by a mixed effect of both patient-controllable and patient-uncontrollable parameters. The patient's uncontrollable parameters define the action of hormonal effects such as CRH induced by either physiological stress or emotional stress. The hormonal effect is two-sided, which is a higher glucose production from the liver and inhibiting insulin production and reducing sensitivity [54,55]. A detailed study conducted by Waldhausl et al [56] demonstrated the significant effect of stress hormones on the production of glucose and insulin resistance. The study was conducted by infusing different stress hormones to investigate the effect of exposure to these hormones on blood glucose response [55]. The extent and degree of hyperglycemia events and insulin resistance during infection incidence directly correlate with the type of pathogen, the type of hormone involved and the severity of the infection [37,38,55]. Generally, the phenomenal effect of infection incidence on blood glucose dynamics in people with diabetes can be simply described using the following relationships:

Where ϕ is an insulin sensitivity factor, BG is the blood glucose level, CH is the amount of carbohydrate consumption, IN is the amount of insulin injection, PA is the amount of physical activity session or exercise load, and CRH is the effect of CRHs. The equation depicts the phenomena that occur during infection incidence, where blood glucose levels are raised by the action of both patient-controllable parameters (CH) and patient-uncontrollable parameters (CRHs, such as cortisol and adrenalin). Thus, consumption of any regular diet in an individual can induce severe hyperglycemia due to the added effect of glucose production from the liver as a result of the

CRH effect [55]. For this reason, the patient is expected to reduce the amount of carbohydrate intake to a certain extent to optimally manage the hyperglycemia crises and at the same time avoiding any critical hypoglycemia incidences (for more information, see Multimedia Appendix 1). By the same token, blood glucose levels can be lowered to euglycemia by the patient-controllable parameters (insulin [IN] and physical activity session or exercise load [PA]). However, due to the change in insulin sensitivity, the action of insulin is reduced (ϕ is affected by infection incidence), and the patient is expected to deliver more insulin injections to counterbalance the effect of insulin resistance [57]. According to our results, all these scenarios are reflected in the individual's blood glucose dynamic infected with flu (influenza), where a dynamic shift occurred from the usual operating point of the blood glucose dynamics. There are elevated blood glucose levels, despite injecting a higher amount of insulin and consuming less carbohydrate than the regular or normal days. These characteristics are clearly demonstrated on the shift incurred on the individual's insulin–to-carbohydrate ratio as compared with the regular or normal days. Therefore, blood glucose, amount of injected insulin, diet intake, and insulin-to-carbohydrate ratio and other supporting physiological parameters such as body temperature and blood pressure can be exploited to develop a personalized health model for detecting infection incidence among people with type 1 diabetes. Given the similarity, this result can also be translated to other types of diabetes, such as people with type 2 diabetes. It is worth mentioning that apart from infection incidence, other factors such as emotional stress could also result in similar variable episodes of elevated blood glucose levels [17]. This can obviously impact the detection performance of the model. However, our results based on yearlong patients' data demonstrated that the use of carbohydrate consumption, insulin injections, and insulin-to-carbohydrate ratio along with the blood glucose could solve this confounding nature. Moreover, acute emotional stress, other than the chronic ones, might have less influence on one's meal appetite compared with infection incidence to skew the insulin-to-carbohydrate ratio [17].

Relevance of the Data

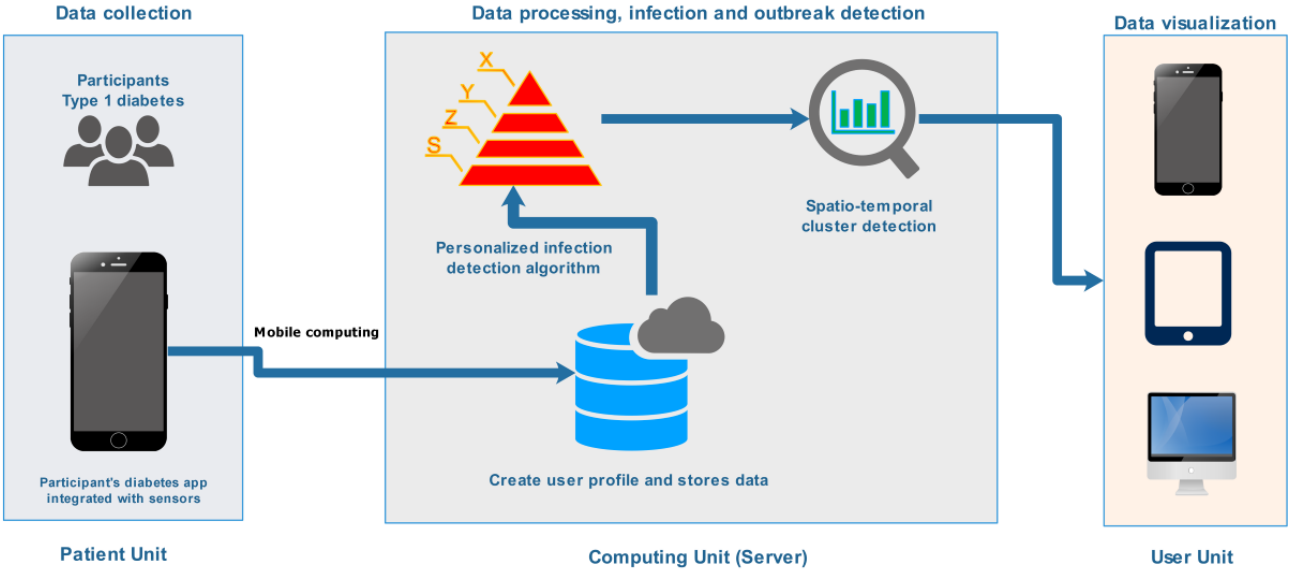
The informational values of the data, availability of the data, and cost of the data are the 3 key metrics necessary to evaluate the relevance of new surveillance data for a digital infectious disease detection system [58]. The informational value of the data assesses how informative the data are to facilitate the detection or characterization of infectious disease outbreaks. In this regard, the surveillance data must clearly indicate the absence or presence of infections either on an individual or population level or both in a timely manner. Furthermore, the rate of false alarms derived from the data is an important factor that dictates the acceptability of the surveillance data, which is in turn governed by the signal-to-noise ratio defining the signal's strength depicting the infection period as compared to the regular or normal period (baseline data) [58]. In this regard, our results demonstrated that the infection-induced signal exhibits high discriminative power from the baseline (normal or regular) patterns. The availability of surveillance data is another crucial indicator for screening potential types of data, which needs to be addressed [58]. In this regard, given the widespread and ubiquitous nature of mobile apps, and different sensors, people with type 1 diabetes collect far more data than ever. For example, many people with type 1 diabetes use continuous glucose monitors and insulin pumps, which are predicted to grow further in terms of both quality and quantity of data in the coming years. The most crucial challenge in this direction includes issues related to security, privacy, and confidentiality of user data if there is a necessity to collect user data into a central server than deploying the detection algorithm on the user's own mobile device. The cost of data delineates the associated cost in relation to acquiring the data in question, including the cost incurred for realizing the data collection system [58]. In this regard, the individual's self-recorded data are solely collected for their own use and used as a secondary source of information for disease surveillance purposes. Providing tailored and valuable feedback to the individual patient might further motivate them to participate on a large scale (for further details, see the section Ethical and Motivational Challenges).

Framework of a Personalized Digital Infectious Disease Detection System

Epidemic intelligence encompasses activities directed toward early detections, verification, and assessment of potential public health threats to notify and recommend necessary measures for the concerned bodies regarding the ongoing situation [56]. Early detection systems such as Google Flu Trends and other existing systems have certain limitations because they do not have the mechanisms to identify or track individual cases through diagnosis or screening based on a personalized health model. This limitation has a major impact and certainly introduces bias in disease outbreak prediction. Currently, a personalized health model, which resembles the way clinicians and epidemiologists classify an individual as normal, suspected, or confirmed case, for screening and case detection doesn't exist [58]. Having a personalized health model can provide information for both individual health-related decision support purposes and at the same time can be used for tracking infectious disease outbreaks among the public. The results of this study demonstrated that commencement of infection in people with type 1 diabetes significantly alters the individual blood glucose dynamics, and such a change can potentially be detected through modeling of the individual blood glucose dynamics. Moreover, incorporating various physiological parameters, for example, heart rate and body temperature, to a personalized health model will further enable the capture of infection incidence as early as possible, that is, incubation period. Therefore, the development of a personalized health model–based digital infectious disease detection system is vital for the success of next-generation public health surveillance systems. The data sources and signal exploited, outbreak detection algorithms employed, clustering approaches, and visualization techniques used to play a central role in any digital infectious disease detection systems by determining its accuracy (sensitivity) and timeliness (lead time) [56]. On the basis of the kind of data sources and signals exploited, infectious disease surveillance systems can be generally grouped into an indicator-based and event-based system [56,59,60]. Event-based systems mainly rely on

unstructured data collected through formal or informal sources and is characterized by quick detection, reporting, and assessments of public health events, including clusters of disease [56,60]. On the other hand, indicator-based systems mainly use structured data, which are collected following a standard case definition and is characterized by routine reporting of disease cases [56,60]. The proposed system [26,61], as shown in Figure 14, is categorized under event-based digital infectious disease detection systems, where the events are grouped under microevents and macroevents [56]. Under the umbrella of these events and the proposed system in general, a framework of several components such as infection detection algorithms (how to develop an algorithm to detect infection incidence at the individual level); clustering algorithms (how to group the infected individuals to form a cluster); visualization techniques (how to report and display the detected outbreak incidence) and further ethical and motivation challenges are briefly discussed below.

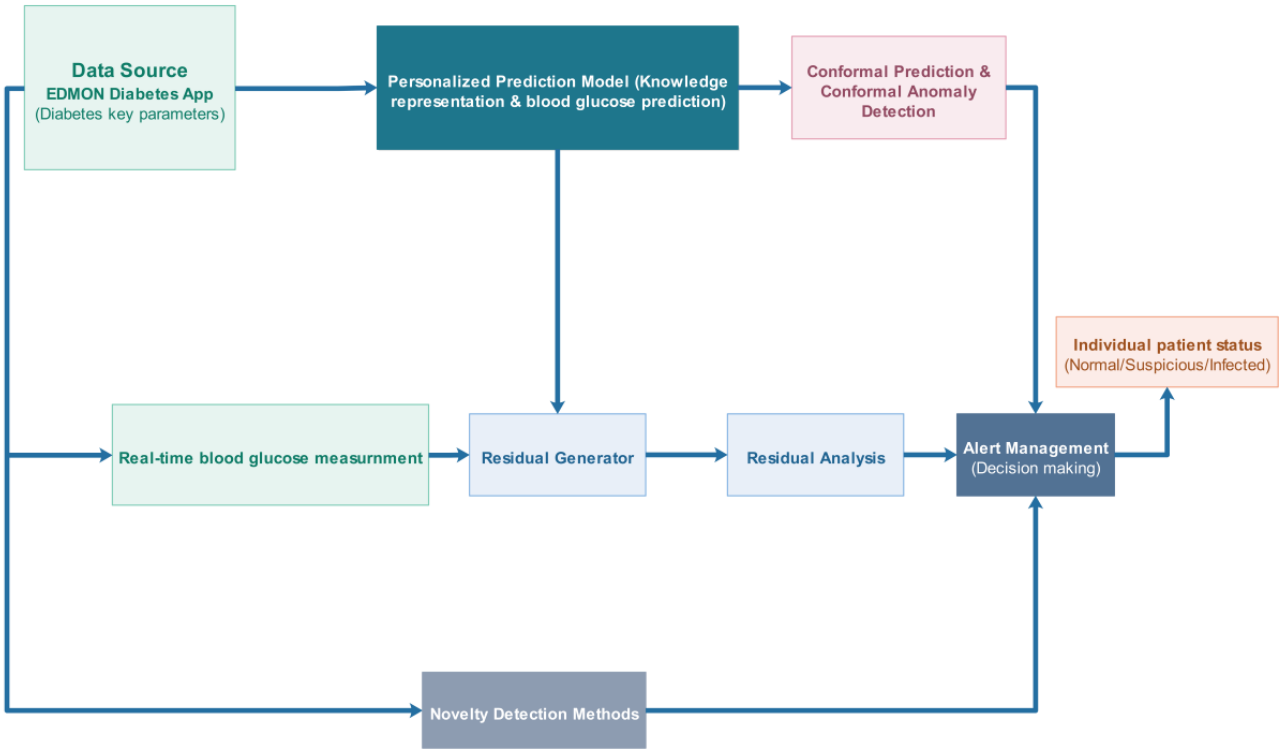
Figure 14. The Proposed System Architecture.



Microevent: Individual-Level Detection of Infection Incidence

The detection of microevents as the name suggests is carried out at an individual level by tracking the individual's diabetes profile including blood glucose levels, amount of insulin injections, carbohydrate consumption, physical activity or exercise sessions, and others. The presence of elevated blood glucose levels despite injecting higher amounts of insulin and consumption of less carbohydrates is regarded as a marker of an event of infection incidence and hence can be defined as a microevent for the event-based digital infectious disease detection system. Detecting the incidence of these kinds of deviation from the usual norm of blood glucose dynamics requires a proper personalized health model, which can learn from past history of the patient and judge whether the information conforms with the usual trend. Hence, the proposed personalized health model for detecting these types of microevents incorporates 3 components: a data source, personalized infection detection algorithm, and alarm management module, as shown in Figure 15. As can be seen from the figure, the personalized infection detection algorithm can be modeled using either a prediction model–based approach or a novelty or anomaly detection–based approach.

Figure 15. The proposed personalized infection detection algorithms for detecting microevents (incidences of infections) in people with type 1 diabetes. These approaches are alternative means of achieving the same objective, which is detecting infection incidences.



Data Sources and Input

The patient unit is a mobile health app, as shown in Figure 14, which integrates data from different sensors and wearables that record key diabetes parameters, such as blood glucose levels, insulin dosage, diet, physical activity, and other optional physiological parameters including body temperature, heart rate, blood pressure, and others [26,61]. The app is also expected to record the geographical location of the individual along with the time of data registration. For example, one way of estimating user location can be carried out based on global positioning system (GPS) information from the mobile phone during data registration [61]. The geographical location of the user can be the geographical coordinates of longitude and latitude [62], postal code address [63], or any local reference coordinates.

Personalized Infection Detection Algorithm

Detection of microevents can be carried out using individual self-recorded historical data based on a personalized health model, that is, either a prediction model [64,65] or novelty detection algorithms [66-68], as shown in Figure 15. The prediction model–based algorithm requires learning the individual blood glucose dynamics for accurate prediction, and for the purpose of detecting the microevents, it can be implemented as either a residual-based [69-72] or a conformal prediction–based approach [73-80]. In a similar fashion, novelty detection–based algorithms can be other alternatives for detecting novel microevents relying on either supervised, semisupervised, and unsupervised approaches [4,66,67,81]. Different categories of novelty detection approaches could be exploited for detecting infection-induced deviations in blood glucose dynamics, including approaches based on statistical techniques [68], prediction, density [82-85], distance [67,86], classification or domain [4,62,63,87-92], clustering [62,93], and ensemble [67,68,80-82,85,86,92-95].

Alarm Management (Decision Making)

The alarm management module accepts the score computed by the personalized infection detection algorithm as input and evaluates the degree of severity of the infection incidence. The severity is evaluated based on the degree of abnormalities of the anomalies score, and a label could be assigned to the individual patient status as normal (0), suspicious (−1), and infected (1). For example, a rule-based fuzzy logic with membership functions of infected, normal, and suspicious can be used to assign the label indicating the severity of the infection incidence using the anomaly score. The output from the alarm management will be directly fed to the cluster detection analysis, which is used to detect a group of patients based on geography (space) and time so as to revel if there is any ongoing infectious disease outbreak.

Macroevents: Population-Level Infectious Disease Outbreak Detection

Cluster Detection Mechanism

Cluster detection is defined as the process of identifying a group of infected individuals with similar spatial, temporal, or spatio-temporal attributes [96]. A spatial cluster analysis only considers a patient's geographical location, and a temporal cluster analysis considers only the time aspect of the events. However, a spatio-temporal cluster analysis is conducted to look for aberrant patterns and detect a cluster of infected people within a specified geographical region and predefined timeframe [96,97]. The analysis of space time clusters is carried out based on a couple of steps: geocoding and identification, which transforms the patient address into meaningful coordinates and

detecting the clusters based on the transformed location and time. A space time cluster analysis is the most favored approach when it comes to early detection of an infectious disease outbreak. A space time cluster analysis can be designed by performing a spatial analysis first and then superimposing the temporal aspect [97]. Regarding the proposed system, the input to the space time cluster detection analysis consists of the individual patient status from alarm management, user location, and time of data registration [26]. The status of the individual patient at any time can be normal (0), infected (1), or suspicious (−1), which comes from alarm management. The user's geographical location can be geographical coordinates of longitude and latitude [98], postal code address [99], or any local reference coordinates. Estimation of user location can be carried out using GPS information from the user's mobile phone, which can be accessed during each data registration. The time aspects depend on the requirement of detection frequency and can be set to either an hourly or daily window. One optimal approach could be tracking the individual during each hour of the day for any statistically significant deviations and performing a concluding analysis at the end of each day based on the daily analysis. Various algorithms have been implemented in the literature, including the density-based clustering algorithm, Bayesian spatial scan statistics, K-NN with Haversian distance (K-nearness), cumulative summation, space time scan statistics, space time permutation scan statistic, and space scan statistics [96,97,100], which can be further tested and adopted. The most important challenge is the sparsity of the data set considering the small proportion of people with type 1 diabetes that can be under surveillance over a large region. Therefore, it is necessary to adopt these cluster detection techniques to overcome data sparsity and produce acceptable detection accuracy. In the proposed system, the detected clusters, if there is any, can be displayed and viewed based on real-time and interactive data visualization tools.

Data Visualization

Data visualization is a mechanism by which detected clusters of disease outbreaks, if there is any, are presented to the responsible bodies for quicker public health actions and responses. Generally, such a visualization tool could report outbreaks of epidemic cases for investigation and follow-up, and it could also report the duration of the epidemic (timing), degree of severity of the epidemic, and the region under threat. In the literature, there are various implemented visualization tools and visual displays with regard to disease outbreak detection systems, including ArcGIS, Google map API, TwiInfo, OpenStreetMap, and JFreeChart, and display mechanisms such as maps, time series, graphs, and color indicators [96]. These visualization tools and display mechanisms can be further tested and adopted in the proposed system. The real-time health status of an individual from the ongoing tracking could be accessible to the end user and can be displayed in a stand-alone software app based on smartphones, tablets, and computers or a dedicated website [26]. Generally, both the data providers (participants) and the general population could benefit from the system in the sense that they can take actions needed to avoid being infected. Moreover, the individual patient could also receive analysis and feedback from the system to learn the situation, such as the degree and severity of deviation of different parameters, including blood glucose, insulin, diet, and insulin-to-carbohydrate ratio, along with their trend as compared with the noninfection period.

Ethical and Motivational Challenges

The implementation of a digital infectious disease detection system based on self-recorded data poses serious challenges that require special attention, such as user privacy and security, data confidentiality, user acceptance, and motivations [26,101], especially during data collection, transmission, and data storage [102,103]. Personal health–related data are sensitive, and the data collection, transmission, and data storage procedure need to follow the standards and regulations provided by the major governing bodies, such as General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA) [104,105]. This includes privacy-preserving mechanisms such as pseudonymization and anonymization to meet the necessary data compliance requirement along with user informed consent [102,103]. According to GDPR, the deidentification procedure is one of the recommended anonymization standards to preserve data confidentiality [104,105]. Moreover, from the technology perspective, it is necessary to look for a robust mechanism to ensure that user privacy and security are respected during data collection, transmission, and storage, as this is highly critical for successful acceptance of the proposed system [26,106]. One such alternative is to look for the possibility of deploying the infection detection algorithm (app logic) on the user (client) mobile device terminal to avoid transmission of patient data to a central server, where only the timely computed infection status of the patient will be sent to the central server for further cluster detection processing. However, this choice requires further feasibility studies to determine the cost, especially in terms of power constraints related to the mobile device terminal, since the detection algorithms need to continuously run in the background to compute the individual's infection status, at the most each hour of the day [26]. In addition, users might also lack willingness to adopt a new technology or system for various reasons ranging from lack of trust, lack of motivation, lack of perceived usefulness, and ease of use [26,101]. However, these challenges can be mitigated by properly buying user trust by developing state-of-the-art technology for preserving privacy, security, and confidentiality of the user and addressing factors that enhance user motivation, including usability knowledge, simplicity and ease of use, reduced time and frequency of interaction with the system, incentives, and others [101].

Conclusions

The relationship between infection incidents and elevated blood glucose levels has been known for a long time. People with type 1 diabetes often experience prolonged episodes of elevated blood glucose levels as a result of infection incidence. Despite the fact that patients increasingly gather data about themselves, there are no solid findings on how to use such self-recorded data as a secondary source of information for other purposes, such as self-management–related decision support during infection incidence and digital infectious disease detection systems. We presented the effect of infection incidence on key

parameters of the blood glucose dynamics along with the necessary framework to exploit the information for realizing a digital infectious disease detection system and further shed light on the possibility of assisting individuals during infection-related blood glucose management crises. The results demonstrated that despite tight blood glucose control, blood glucose level is still elevated during infection incidence. The analysis shows that infection incidences have a significant impact on blood glucose dynamics as compared with the other patient-uncontrollable factors. All of these findings indicate that blood glucose levels were elevated despite a higher amount of insulin injection and reduced carbohydrate consumption, which are quite significant changes that could possibly be detected through personalized modeling that spans from prediction models to anomaly detection algorithms. However, further large-scale studies are required to strengthen the findings. Moreover, future research should investigate the possibility of improving detection time and disease characterization. Early detection, that is, during the incubation period, is a critical component of any outbreak detection system and therefore needs to be improved by analyzing how various features of CGM can be used in context with other parameters, such as diet, insulin, and physical activity data. For instance, different individuals with type 1 diabetes often reported the experiences of an elevated episode of blood glucose levels before the onset of the first symptoms. Disease characterization involves determining the type and nature of pathogens that cause the infection, which is an important component of outbreak reporting. The extent and degree of the impact of infection incidence on blood glucose dynamics are highly correlated with the disease pathogens involved. In this regard, carefully analyzing a large-scale self-recorded data set containing several infection incidences (different pathogens) could characterize them based on their effect on blood glucose dynamics. Generally, we foresee that these findings can benefit the efforts toward building next-generation digital infectious disease surveillance systems and provoke further thoughts in this challenging field.

Acknowledgments

The work presented in this paper is part of the project “Electronic Disease Surveillance Monitoring Network (EDMON) system,” which is funded by the University of Tromsø—The Arctic University of Norway and National Library of Medicine “Mechanistic machine learning” (grant number: LM012734) and is the PhD program of the first author, AW. The authors would like to extend their sincere gratitude to all the participants of the study.

Authors' Contributions

The first author, AW, conceived the study, designed and performed the experiments, and wrote the manuscript. IK, EÅ, AH, DA, and GH provided successive inputs and revised the manuscript. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparative analysis of parameters of blood glucose dynamics with and without infection incidences.
[DOCX File , 9296 KB - jmir_v22i8e18911_app1.docx]

Multimedia Appendix 2
Analytical plot of the normal/regular patient years.
[DOCX File , 1346 KB - jmir_v22i8e18911_app2.docx]

Multimedia Appendix 3
Analytical plot of the patient years with acute infection.
[DOCX File , 1662 KB - jmir_v22i8e18911_app3.docx]

References

1. Choi J, Cho Y, Shim E, Woo H. Web-based infectious disease surveillance systems and public health perspectives: a systematic review. BMC Public Health 2016 Dec 8;16(1):1238 [FREE Full text] [doi: 10.1186/s12889-016-3893-0] [Medline: 27931204]
2. Hope K, Durrheim DN, d'Espaignet ET, Dalton C. Syndromic Surveillance: is it a useful tool for local outbreak detection? J Epidemiol Community Health 2006 May;60(5):374-375 [FREE Full text] [doi: 10.1136/jech.2005.035337] [Medline: 16680907]
3. Banaee H, Ahmed MU, Loutfi A. Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. Sensors (Basel) 2013 Dec 17;13(12):17472-17500 [FREE Full text] [doi: 10.3390/s131217472] [Medline: 24351646]

4. Woldaregay AZ, Årsand E, Botsis T, Albers D, Mamykina L, Hartvigsen G. Data-driven blood glucose pattern classification and anomalies detection: machine-learning applications in type 1 diabetes. J Med Internet Res 2019 May 1;21(5):e11030 [FREE Full text] [doi: 10.2196/11030] [Medline: 31042157]
5. Fawcett T. Mining the quantified self: personal knowledge discovery as a challenge for data science. Big Data 2015 Dec;3(4):249-266. [doi: 10.1089/big.2015.0049] [Medline: 27441406]
6. Gurrin C, Smeaton AF, Doherty AR. LifeLogging: personal big data. FNT Inf Ret 2014;8(1):1-125 [FREE Full text] [doi: 10.1561/1500000033]
7. Prince JD. The quantified self: operationalizing the quotidien. J Electron Resour Med Libr 2014 May 30;11(2):91-99 [FREE Full text] [doi: 10.1080/15424065.2014.909145]
8. Rawassizadeh R, Momeni E, Dobbins C, Mirza-Babaei P, Rahnamoun R. Lesson learned from collecting quantified self information via mobile and wearable devices. J Sens Actuator Netw 2015 Nov 5;4(4):315-335 [FREE Full text] [doi: 10.3390/jsan4040315]
9. Bellazzi R, Abu-Hanna A. Data mining technologies for blood glucose and diabetes management. J Diabetes Sci Technol 2009 May 1;3(3):603-612 [FREE Full text] [doi: 10.1177/193229680900300326] [Medline: 20144300]
10. Rodríguez-Rodríguez I, Rodríguez JV, Zamora-Izquierdo M. Variables to be monitored via biomedical sensors for complete type 1 diabetes mellitus management: an extension of the 'on-board' concept. J Diabetes Res 2018;2018:4826984 [FREE Full text] [doi: 10.1155/2018/4826984] [Medline: 30363935]
11. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb;11(2):e1003904 [FREE Full text] [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]
12. Denecke K. An ethical assessment model for digital disease detection technologies. Life Sci Soc Policy 2017 Sep 20;13(1):16 [FREE Full text] [doi: 10.1186/s40504-017-0062-x] [Medline: 28929347]
13. Tsui F, Espino JU, Dato VM, Gesteland PH, Hutman J, Wagner MM. Technical description of RODS: a real-time public health surveillance system. J Am Med Inform Assoc 2003;10(5):399-408 [FREE Full text] [doi: 10.1197/jamia.M1345] [Medline: 12807803]
14. Ogurtsova K, da Rocha Fernandes J, Huang Y, Linnenkamp U, Guariguata L, Cho N, et al. IDF diabetes atlas: global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract 2017 Jun;128:40-50. [doi: 10.1016/j.diabres.2017.03.024] [Medline: 28437734]
15. Clark M. What is diabetes? In: Ogden J, editor. Understanding Diabetes. New Jersey, United States: John Wiley & Sons Ltd; 2004.
16. Bazaev NA, Pletenev AN, Pozhar KV. Classification of factors affecting blood glucose concentration dynamics. Biomed Eng 2013 Jul 18;47(2):100-103. [doi: 10.1007/s10527-013-9344-7] [Medline: 23789157]
17. Finan DA, Zisser H, Jovanovič L, Bevier WC, Seborg DE. Automatic detection of stress states in type 1 diabetes subjects in ambulatory conditions. Ind Eng Chem Res 2010 Sep 1;49(17):7843-7848 [FREE Full text] [doi: 10.1021/ie901891c] [Medline: 20953334]
18. Botsis T, Hejlesen O, Bellika JG, Hartvigsen G. Electronic disease surveillance for sensitive population groups - the diabetics case study. Stud Health Technol Inform 2008;136:365-370. [Medline: 18487758]
19. Rayfield EJ, Ault MJ, Keusch GT, Brothers MJ, Nechemias C, Smith H. Infection and diabetes: the case for glucose control. Am J Med 1982 Mar;72(3):439-450. [doi: 10.1016/0002-9343(82)90511-3] [Medline: 7036735]
20. Simonsen JR, Harjutsalo V, Järvinen A, Kirveskari J, Forsblom C, Groop P, FinnDiane Study Group. Bacterial infections in patients with type 1 diabetes: a 14-year follow-up study. BMJ Open Diabetes Res Care 2015;3(1):e000067 [FREE Full text] [doi: 10.1136/bmjdrc-2014-000067] [Medline: 25767718]
21. Yki-Järvinen H, Sammalkorpi K, Koivisto VA, Nikkilä EA. Severity, duration, and mechanisms of insulin resistance during acute infections. J Clin Endocrinol Metab 1989 Aug;69(2):317-323. [doi: 10.1210/jcem-69-2-317] [Medline: 2666428]
22. Botsis T, Lai AM, Hripcsak G, Palmas W, Starren JB, Hartvigsen G. Proof of concept for the role of glycemic control in the early detection of infections in diabetics. Health Informatics J 2012 Mar;18(1):26-35. [doi: 10.1177/1460458211428427] [Medline: 22447875]
23. Botsis T, Hartvigsen G. Exploring new directions in disease surveillance for people with diabetes: lessons learned and future plans. Stud Health Technol Inform 2010;160(Pt 1):466-470. [Medline: 20841730]
24. Arsand E, Walseth OA, Andersson N, Fernando R, Granberg O, Bellika JG, et al. Using blood glucose data as an indicator for epidemic disease outbreaks. Stud Health Technol Inform 2005;116:217-222. [Medline: 16160262]
25. Granberg O, Bellika JG, Arsand E, Hartvigsen G. Automatic infection detection system. Stud Health Technol Inform 2007;129(Pt 1):566-570. [Medline: 17911780]
26. Woldaregay AZ, Årsand E, Giordanengo A, Albers D, Mamykina L, Botsis T, et al. EDMON-A Wireless Communication Platform for a Real-Time Infectious Disease Outbreak Detection System Using Self-Recorded Data from People with Type 1 Diabetes. In: The 15th Scandinavian Conference on Health Informatics. 2017 Presented at: CHI'17; August 29, 2017; Kristiansand, Norway URL: https://ep.liu.se/konferensartikel.aspx?series=ecp&issue=145&Article_No=3
27. Lauritzen JN, Årsand E, Vuurden KV, Bellika JG, Hejlesen OK, Hartvigsen G. Towards a Mobile Solution for Predicting Illness in Type 1 Diabetes Mellitus: Development of a Prediction Model for Detecting Risk of Illness in Type 1 Diabetes Prior to Symptom Onset. In: 2nd International Conference on Wireless Communication, Vehicular Technology, Information

Theory and Aerospace & Electronic Systems Technology. 2011 Presented at: Wireless VITAE; February 28-March 3, 2011; Chennai, India. [doi: 10.1109/wirelessvitae.2011.5940877]
28. Skrøvseth SO, Bellika JG, Godtliebsen F. Causality in scale space as an approach to change detection. PLoS One 2012;7(12):e52253 [FREE Full text] [doi: 10.1371/journal.pone.0052253] [Medline: 23300626]
29. Botsis T, Bellika JG, Hartvigsen G. Disease Surveillance Systems for Sensitive Population Groups. Faculty & Staff Insider - University of Washington. 2007. URL: https://faculty.washington.edu/lober/www.isdsjournal.org/htdocs/articles/2026.pdf [accessed 2020-07-23]
30. Botsis T, Hejlesen O, Bellika JG, Hartvigsen G. Blood Glucose Levels as an Indicator for the Early Detection of Infections In Type-1 Diabetics. Faculty & Staff Insider - University of Washington. 2007. URL: http://faculty.washington.edu/lober/www.isdsjournal.org/htdocs/articles/2025.pdf [accessed 2020-07-23]
31. Lauritzen J, Årsand E, Van Vuurden K, Hejlesen O, Hartvigsen G. Exploring illness prediction in Type 1 Diabetes Mellitus Pre-Symptom Onset. E-health Research. 2010. URL: https://ehealthresearch.no/en/publications/exploring-illness-prediction-in-type-1-diabetes-mellitus-pre-symptom-onset-1 [accessed 2020-07-23]
32. Hartvigsen G, Årsand E, Botsis T, van Vuurden K, Johansen M, Bellika JG. Reusing patient data to enhance patient empowerment and electronic disease surveillance. J Inf Technol 2009;7(1):4-12.
33. Botsis T, Bellika JG, Hartvigsen G. New Directions in Electronic Disease Surveillance: Detection of Infectious Diseases during the Incubation Period. In: International Conference on eHealth. 2009 Presented at: EH'09; February 1-7, 2009; Cancun, Mexico p. 176-183. [doi: 10.1109/etelemed.2009.9]
34. Botsis T, Hejlesen O, Bellika JG, Hartvigsen G. Blood Glucose Levels as a Censor for Early Detection of Infection in Type-1 Diabetics. Faculty & Staff Insider - University of Washington. 2007. URL: http://faculty.washington.edu/lober/www.isdsjournal.org/htdocs/articles/2025.pdf [accessed 2020-07-23]
35. Botsis T, Hejlesen O, Bellika JG, Hartvigsen G. Disease surveillance systems for diabetics. In: Telemedicine and e-Health. 2008 Mar Presented at: The American Telemedicine Association Annual International Meeting and Exposition; 2008; Seattle p. 108-108. [doi: 10.1089/tmj.2008.9983.supp]
36. Botsis T, Hejlesen O, Bellika JG, Hartvigsen G. Electronic infectious disease surveillance systems for diabetics. 2008 Presented at: International Conference on Advanced Technologies & Treatments for Diabetes; Prague, Czech Republic.
37. Brealey D, Singer M. Hyperglycemia in critical illness: a review. J Diabetes Sci Technol 2009 Nov 1;3(6):1250-1260 [FREE Full text] [doi: 10.1177/193229680900300604] [Medline: 20144378]
38. Marik PE, Bellomo R. Stress hyperglycemia: an essential survival response!. Crit Care 2013 Mar 6;17(2):305 [FREE Full text] [doi: 10.1186/cc12514] [Medline: 23470218]
39. Årsand E, Muzny M, Bradway M, Muzik J, Hartvigsen G. Performance of the first combined smartwatch and smartphone diabetes diary application study. J Diabetes Sci Technol 2015 May;9(3):556-563 [FREE Full text] [doi: 10.1177/1932296814567708] [Medline: 25591859]
40. Features. Spike App. URL: https://spike-app.com/#features1 [accessed 2019-09-20]
41. Gillespie SJ, Kulkarni KD, Daly AE. Using carbohydrate counting in diabetes clinical practice. J Am Diet Assoc 1998 Aug;98(8):897-905. [doi: 10.1016/S0002-8223(98)00206-5] [Medline: 9710660]
42. Rasoulzadeh V, Erkus EC, Yogurt TA, Ulusoy I, Zergeroğlu SA. A comparative stationarity analysis of EEG signals. Ann Oper Res 2016 Apr 26;258(1):133-157 [FREE Full text] [doi: 10.1007/s10479-016-2187-3]
43. Azami H, Mohammadi K, Bozorgtabar B. An improved signal segmentation using moving average and Savitzky-Golay filter. J Signal Inf Process 2012;03(01):39-44. [doi: 10.4236/jsip.2012.31006]
44. Botev ZI, Grotowski JF, Kroese DP. Kernel density estimation via diffusion. Ann Statist 2010 Oct;38(5):2916-2957. [doi: 10.1214/10-aos799]
45. Gramacki A. Nonparametric density estimation. In: Nonparametric Kernel Density Estimation and Its Computational Aspects. Cham, Switzerland: Springer International Publishing; 2018:7-24.
46. Gramacki A. Bandwidth selectors for kernel density estimation. In: Nonparametric Kernel Density Estimation and Its Computational Aspects. Cham, Switzerland: Springer International Publishing; 2018:63-83.
47. Heidenreich N, Schindler A, Sperlich S. Bandwidth selection for kernel density estimation: a review of fully automatic selectors. AStA Adv Stat Anal 2013 Jun 2;97(4):403-433 [FREE Full text] [doi: 10.1007/s10182-013-0216-y]
48. Gramacki A. Kernel density estimation. In: Nonparametric Kernel Density Estimation and Its Computational Aspects. Cham, Switzerland: Springer International Publishing; 2018:25-62.
49. Zdravko B. Kernel Density Estimator. MathWorks. 2015. URL: https://www.mathworks.com/matlabcentral/fileexchange/14034-kernel-density-estimator [accessed 2019-06-22]
50. Rossini AJ. 'Applied smoothing techniques for data analysis: the kernel approach with s-plus illustrations' by Adrian W Bowman and Adelchi Azzalini. Comput Stat 2000 Sep 11;15(2):301-302 [FREE Full text] [doi: 10.1007/s001800000033]
51. Bowman AW, Azzalini A. Applied Smoothing Techniques for Data Analysis: The Kernel Approach With S-Plus Illustrations. Oxford, UK: OUP Oxford; 1997.
52. Yi C. Bivariant Kernel Density Estimation (V2.1). MathWorks. 2013. URL: https://la.mathworks.com/matlabcentral/fileexchange/19280-bivariant-kernel-density-estimation-v2-1?s_tid=prof_contriblnk [accessed 2019-06-22]

53. Yamaguchi M, Kaseda C, Yamazaki K, Kobayashi M. Prediction of blood glucose level of type 1 diabetics using response surface methodology and data mining. Med Biol Eng Comput 2006 Jun;44(6):451-457. [doi: 10.1007/s11517-006-0049-x] [Medline: 16937196]
54. van Herpe T, Pluymers B, Espinoza M, van den Berghe G, de Moor BA. A minimal model for glycemia control in critically ill patients. Conf Proc IEEE Eng Med Biol Soc 2006;2006:5432-5435 [FREE Full text] [doi: 10.1109/IEMBS.2006.260613] [Medline: 17946700]
55. Waldhäusl WK, Bratusch-Marrain P, Komjati M, Breitenecker F, Troch I. Blood glucose response to stress hormone exposure in healthy man and insulin dependent diabetic patients: prediction by computer modeling. IEEE Trans Biomed Eng 1992 Aug;39(8):779-790 [FREE Full text] [doi: 10.1109/10.148386] [Medline: 1354649]
56. O'Shea J. Digital disease detection: a systematic review of event-based internet biosurveillance systems. Int J Med Inform 2017 May;101:15-22 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.01.019] [Medline: 28347443]
57. Mizock BA. Alterations in carbohydrate metabolism during stress: a review of the literature. Am J Med 1995 Jan;98(1):75-84 [FREE Full text] [doi: 10.1016/S0002-9343(99)80083-7] [Medline: 7825623]
58. Wagner MM, Moore AW, Aryel RM, editors. Handbook of Biosurveillance. New York, USA: Elsevier; 2006.
59. Event-Based Surveillance. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/globalhealth/healthprotection/gddopscenter/how.html [accessed 2019-06-26]
60. World Health Organization. A Guide to Establishing Event-Based Surveillance. Geneva, Switzerland: World Health Organization; 2008.
61. Coucheron S, Woldaregay AZ, Årsand E, Botsis T, Hartvigsen G. EDMON - A System Architecture for Real-Time Infection Monitoring and Outbreak Detection Based on Self-Recorded Data from People with Type 1 Diabetes: System Design and Prototype Implementation. In: The 17th Scandinavian Conference on Health Informatics. 2019 Presented at: CHI'19; November 12, 2019; Oslo, Norway p. 37-44 URL: https://ep.liu.se/ecp/161/007/ecp19161007.pdf
62. Tax DM. One-Class Classification: Concept Learning in the Absence of Counter-Examples. South Holland, Netherlands: Technische Universiteit Delft; 2002.
63. Mazhelis O. One-class classifiers: a review and analysis of suitability in the context of mobile-masquerader detection. S Afr Comput J 2006;36:29-48 [FREE Full text]
64. Oviedo S, Vehí J, Calm R, Armengol J. A review of personalized blood glucose prediction strategies for T1DM patients. Int J Numer Method Biomed Eng 2017 Jun;33(6):---. [doi: 10.1002/cnm.2833] [Medline: 27644067]
65. Woldaregay AZ, Årsand E, Walderhaug S, Albers D, Mamykina L, Botsis T, et al. Data-driven modeling and prediction of blood glucose dynamics: machine learning applications in type 1 diabetes. Artif Intell Med 2019 Jul;98:109-134 [FREE Full text] [doi: 10.1016/j.artmed.2019.07.007] [Medline: 31383477]
66. Dunning T, Friedman E. In: Loukides M, editor. Practical Machine Learning: A New Look at Anomaly Detection. New York, USA: O'Reilly Media Inc; 2018.
67. Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey. ACM Comput Surv 2009 Jul;41(3):1-58 [FREE Full text] [doi: 10.1145/1541880.1541882]
68. Pimentel MA, Clifton DA, Clifton L, Tarassenko L. A review of novelty detection. Sig Process 2014 Jun;99:215-249 [FREE Full text] [doi: 10.1016/j.sigpro.2013.12.026]
69. Haslett J, Haslett SJ. The three basic types of residuals for a linear model. Int Statistical Rev 2007 Apr;75(1):1-24 [FREE Full text] [doi: 10.1111/j.1751-5823.2006.00001.x]
70. Harvey AC, Koopman SJ. Diagnostic checking of unobserved-components time series models. J Bus Econ Stat 1992 Oct;10(4):377-389. [doi: 10.1080/07350015.1992.10509913]
71. Yu Y, Zhu Y, Li S, Wan D. Time series outlier detection based on sliding window prediction. Math Probl Eng 2014;2014:1-14 [FREE Full text] [doi: 10.1155/2014/879736]
72. Woldaregay AZ, Årsand E, Botsis T, Hartvigsen G. An early infectious disease outbreak detection mechanism based on self-recorded data from people with diabetes. Stud Health Technol Inform 2017;245:619-623. [Medline: 29295170]
73. Vovk V, Gammerman A, Shafer G. Conformal prediction. In: Algorithmic Learning in a Random World. Boston, MA: Springer US; 2005:17-51.
74. Vovk V. The basic conformal prediction framework. In: Conformal Prediction for Reliable Machine Learning. New York, USA: Morgan Kaufmann; 2014:3-19.
75. Vovk V. Beyond the basic conformal prediction framework. In: Conformal Prediction for Reliable Machine Learning. New York, USA: Morgan Kaufmann; 2014:21-46.
76. James S. The Efficiency of Conformal Predictors for Anomaly Detection. Royal Holloway, University of London: Research. 2016. URL: https://pure.royalholloway.ac.uk/portal/en/publications/the-efficiency-of-conformal-predictors-for-anomaly-detection(d68cdec7-7d84-414b-9498-0c8539bb57b8)/export.html [accessed 2020-07-23]
77. Laxhammar R, Falkman G. Inductive conformal anomaly detection for sequential detection of anomalous sub-trajectories. Ann Math Artif Intell 2013 Sep 20;74(1-2):67-94 [FREE Full text] [doi: 10.1007/s10472-013-9381-7]
78. Shafer G, Vovk V. A tutorial on conformal prediction. J Mach Learn Res 2008 Jun 1;9(-):371-421 [FREE Full text] [doi: 10.5555/1390681.1390693]

79. Smith J, Nouretdinov I, Craddock R, Offer C, Gammerman A. Anomaly Detection of Trajectories with Kernel Density Estimation by Conformal Prediction. In: IFIP International Conference on Artificial Intelligence Applications and Innovations. 2014 Presented at: AIAI'14; September 19-21, 2014; Rhodos, Greece URL: https://doi.org/10.1007/978-3-662-44722-2_29 [doi: 10.1007/978-3-662-44722-2_29]
80. Laxhammar R. Conformal Anomaly Detection: Detecting Abnormal Trajectories in Surveillance Applications. Skövde: School of Informatics, University of Skövde; 2014.
81. Mehrotra K, Mohan CK, Huang H. Anomaly Detection Principles and Algorithms. New York, USA: Springer; 2017.
82. Zhao Z, Mehrotra KG, Mohan CK. Ensemble Algorithms for Unsupervised Anomaly Detection. In: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. 2015 Presented at: IEA/AIE'15; June 10-12, 2015; Seoul, Korea (Republic of) URL: https://link.springer.com/content/pdf/10.1007%2F978-3-319-19066-2.pdf [doi: 10.1007/978-3-319-19066-2_50]
83. Breunig MM, Kriegel HP, Ng RT, Sanders J. LOF: Identifying Density-Based Local Outliers. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. 2000 Presented at: SIGMOD'00; May 16-18, 2000; Dallas, Texas, USA. [doi: 10.1145/342009.335388]
84. Tang J, Chen Z, Fu AW, Cheung DW. Enhancing Effectiveness of Outlier Detections for Low Density Patterns. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. 2002 Presented at: PAKDD'02; May 6-8, 2002; Taipei, Taiwan. [doi: 10.1007/3-540-47887-6_53]
85. Nguyen HV, Nguyen TT, Nguyen QU. Sequential Ensemble Method for Unsupervised Anomaly Detection. In: 9th International Conference on Knowledge and Systems Engineering (KSE). 2017 Presented at: KSE'17; October 19-21, 2017; Hue, Vietnam. [doi: 10.1109/kse.2017.8119437]
86. Goldstein M, Uchida S. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLoS One 2016;11(4):e0152173 [FREE Full text] [doi: 10.1371/journal.pone.0152173] [Medline: 27093601]
87. Irigoien I, Sierra B, Arenas C. Towards application of one-class classification methods to medical data. ScientificWorldJournal 2014;2014:730712 [FREE Full text] [doi: 10.1155/2014/730712] [Medline: 24778600]
88. Tax DM, Duin RP. Support vector data description. Mach Learn 2004 Jan;54(1):45-66 [FREE Full text] [doi: 10.1023/b:mach.0000008084.60811.49]
89. Schölkopf B, Williamson RC, Smola AJ, Shawe-Taylor J, Platt JC. Support Vector Method for Novelty Detection. In: The 12th International Conference on Neural Information Processing. 1999 Presented at: NIPS'99; November 29-December 4, 1999; Denver, Colorado, USA URL: https://dl.acm.org/doi/10.5555/3009657.3009740 [doi: 10.5555/3009657.3009740]
90. Juszczak P, Tax DM, Pe kalska E, Duin RP. Minimum spanning tree based one-class classifier. Neurocomputing 2009 Mar;72(7-9):1859-1869 [FREE Full text] [doi: 10.1016/j.neucom.2008.05.003]
91. Khan SS, Madden MG. One-class classification: taxonomy of study and review of techniques. Knowl Eng Rev 2014 Jan 24;29(3):345-374 [FREE Full text] [doi: 10.1017/s026988891300043x]
92. Wang B, Mao Z. One-class classifiers ensemble based anomaly detection scheme for process control systems. T I Meas Control 2017 Sep 21;40(12):3466-3476 [FREE Full text] [doi: 10.1177/0142331217724508]
93. Ahmad S, Lavin A, Purdy S, Agha Z. Unsupervised real-time anomaly detection for streaming data. Neurocomputing 2017 Nov;262:134-147 [FREE Full text] [doi: 10.1016/j.neucom.2017.04.070]
94. Yu E, Parekh P. A bayesian ensemble for unsupervised anomaly detection. arXiv 2016:- preprint [FREE Full text]
95. Shoemaker L, Hall LO. Anomaly Detection Using Ensembles. In: International Workshop on Multiple Classifier Systems. 2011 Presented at: MCS'11; June 15-17, 2011; Naples, Italy URL: https://doi.org/10.1007/978-3-642-21557-5_3 [doi: 10.1007/978-3-642-21557-5_3]
96. Yeng PK, Woldaregay AZ, Solvoll T, Hartvigsen G. Cluster detection mechanisms for syndromic surveillance systems: systematic review and framework development. JMIR Public Health Surveill 2020 May 26;6(2):e11512 [FREE Full text] [doi: 10.2196/11512] [Medline: 32357126]
97. Shi Z, Pun-Cheng L. Spatiotemporal data clustering: a survey of methods. Int J Geogr Inf Sci 2019 Feb 28;8(3):112 [FREE Full text] [doi: 10.3390/ijgi8030112]
98. Duangchaemkarn K, Chaovatut V, Wiwatanadate P, Boonchieng E. Symptom-based data preprocessing for the detection of disease outbreak. Conf Proc IEEE Eng Med Biol Soc 2017 Jul;2017:2614-2617 [FREE Full text] [doi: 10.1109/EMBC.2017.8037393] [Medline: 29060435]
99. Grubesic TH, Matisziw TC. On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. Int J Health Geogr 2006 Dec 13;5:58 [FREE Full text] [doi: 10.1186/1476-072X-5-58] [Medline: 17166283]
100. Glatman-Freedman A, Kaufman Z, Kopel E, Bassal R, Taran D, Valinsky L, et al. Near real-time space-time cluster analysis for detection of enteric disease outbreaks in a community setting. J Infect 2016 Aug;73(2):99-106. [doi: 10.1016/j.jinf.2016.04.038] [Medline: 27311747]
101. Woldaregay AZ, Issom D, Henriksen A, Marttila H, Mikalsen M, Pfuhl G, et al. Motivational factors for user engagement with mhealth apps. Stud Health Technol Inform 2018;249:151-157. [Medline: 29866972]

102. Klingler C, Silva DS, Schuermann C, Reis AA, Saxena A, Strech D. Ethical issues in public health surveillance: a systematic qualitative review. BMC Public Health 2017 Apr 4;17(1):295 [FREE Full text] [doi: 10.1186/s12889-017-4200-4] [Medline: 28376752]
103. Geneviève LD, Martani A, Wangmo T, Paolotti D, Koppeschaar C, Kjelsø C, et al. Participatory disease surveillance systems: ethical framework. J Med Internet Res 2019 May 23;21(5):e12273 [FREE Full text] [doi: 10.2196/12273] [Medline: 31124466]
104. Rumbold JM, Pierscionek B. The effect of the general data protection regulation on medical research. J Med Internet Res 2017 Feb 24;19(2):e47 [FREE Full text] [doi: 10.2196/jmir.7108] [Medline: 28235748]
105. Hintze M. Viewing the GDPR through a de-identification lens: a tool for compliance, clarification, and consistency. Int Data Priv Law 2017;8(1):86-101 [FREE Full text] [doi: 10.1093/idpl/ipx020]
106. Sun W, Cai Z, Li Y, Liu F, Fang S, Wang G. Security and privacy in the medical internet of things: a review. Secur Commun Netw 2018;2018:1-9 [FREE Full text] [doi: 10.1155/2018/5978636]

Abbreviations

CGM: continuous glucose monitoring
CRH: counterregulatory hormone
GDPR: General Data Protection Regulation
GPS: global positioning system
IoT: Internet of Things
SMBG: self-monitoring of blood glucose

Edited by G Eysenbach; submitted 26.03.20; peer-reviewed by M Nomali, S Sarbadhikari; comments to author 15.04.20; revised version received 06.06.20; accepted 11.06.20; published 12.08.20.

Please cite as:
Woldaregay AZ, Launonen IK, Årsand E, Albers D, Holubová A, Hartvigsen G
Toward Detecting Infection Incidence in People With Type 1 Diabetes Using Self-Recorded Data (Part 1): A Novel Framework for a Personalized Digital Infectious Disease Detection System
J Med Internet Res 2020;22(8):e18911
URL: https://www.jmir.org/2020/8/e18911
doi:10.2196/18911
PMID:32784178

Original Paper

A Novel Approach for Continuous Health Status Monitoring and Automatic Detection of Infection Incidences in People With Type 1 Diabetes Using Machine Learning Algorithms (Part 2): A Personalized Digital Infectious Disease Detection Mechanism

Ashenafi Zebene Woldaregay[1], MSc; Ilkka Kalervo Launonen[2], PhD; David Albers[3,4], PhD; Jorge Igual[5], PhD; Eirik Årsand[1], PhD; Gunnar Hartvigsen[1], PhD

[1]Department of Computer Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway
[2]Department of Clinical Research, University Hospital of North Norway, Tromsø, Norway
[3]Department of Pediatrics, Informatics and Data Science, University of Colorado, Aurora, CO, United States
[4]Department of Biomedical Informatics, Columbia University, New York, NY, United States
[5]Universitat Politècnica de València, Valencia, Spain

Corresponding Author:
Ashenafi Zebene Woldaregay, MSc
Department of Computer Science
University of Tromsø – The Arctic University of Norway
Hansine Hansens veg 54, Science building Realfagbygget, office A124
Tromsø
Norway
Phone: 47 46359333
Email: ashenafi.z.woldaregay@uit.no

Abstract

Background: Semisupervised and unsupervised anomaly detection methods have been widely used in various applications to detect anomalous objects from a given data set. Specifically, these methods are popular in the medical domain because of their suitability for applications where there is a lack of a sufficient data set for the other classes. Infection incidence often brings prolonged hyperglycemia and frequent insulin injections in people with type 1 diabetes, which are significant anomalies. Despite these potentials, there have been very few studies that focused on detecting infection incidences in individuals with type 1 diabetes using a dedicated personalized health model.

Objective: This study aims to develop a personalized health model that can automatically detect the incidence of infection in people with type 1 diabetes using blood glucose levels and insulin-to-carbohydrate ratio as input variables. The model is expected to detect deviations from the norm because of infection incidences considering elevated blood glucose levels coupled with unusual changes in the insulin-to-carbohydrate ratio.

Methods: Three groups of one-class classifiers were trained on target data sets (regular days) and tested on a data set containing both the target and the nontarget (infection days). For comparison, two unsupervised models were also tested. The data set consists of high-precision self-recorded data collected from three real subjects with type 1 diabetes incorporating blood glucose, insulin, diet, and events of infection. The models were evaluated on two groups of data: raw and filtered data and compared based on their performance, computational time, and number of samples required.

Results: The one-class classifiers achieved excellent performance. In comparison, the unsupervised models suffered from performance degradation mainly because of the atypical nature of the data. Among the one-class classifiers, the boundary and domain-based method produced a better description of the data. Regarding the computational time, nearest neighbor, support vector data description, and self-organizing map took considerable training time, which typically increased as the sample size increased, and only local outlier factor and connectivity-based outlier factor took considerable testing time.

Conclusions: We demonstrated the applicability of one-class classifiers and unsupervised models for the detection of infection incidence in people with type 1 diabetes. In this patient group, detecting infection can provide an opportunity to devise tailored services and also to detect potential public health threats. The proposed approaches achieved excellent performance; in particular, the boundary and domain-based method performed better. Among the respective groups, particular models such as one-class

support vector machine, K-nearest neighbor, and K-means achieved excellent performance in all the sample sizes and infection cases. Overall, we foresee that the results could encourage researchers to examine beyond the presented features into other additional features of the self-recorded data, for example, continuous glucose monitoring features and physical activity data, on a large scale.

(J Med Internet Res 2020;22(8):e18912) doi:10.2196/18912

KEYWORDS

type 1 diabetes; self-recorded health data; infection detection; decision support techniques; outbreak detection system; syndromic surveillance

Introduction

Anomaly or novelty detection problem involves identifying the anomalous or novel instances, which exhibit different characteristics, from the rest of the data set and has been widely used in various applications including machine fault and sensor failure detection, prevention of credit card or identity fraud, health and medical diagnostics and monitoring, cyber-intrusion detection, and others [1-7]. The term anomaly was precisely coined by Hawkins [8] as "observations that deviate much from the other observations so as to arouse suspicions that it could be generated by a different process." Anomalousness is usually described as point, contextual, and collective, depending on how the degree of anomaly is computed [1,7,9]. On the basis of the necessity of having labeled data instances for the respective class, the anomaly detection problem can be approached as supervised, semisupervised, and unsupervised [3,7,9-11]. Supervised anomaly detection, for example, multiclass classification, requires labeled data instances for both the target and the nontarget (anomaly) classes. This characteristic makes it impractical for tasks where there is difficulty in either finding enough samples for the anomaly class, that is, poorly sampled and unbalanced data, or demarcating boundaries of the anomaly class [7,10,12]. Moreover, anomalies could also evolve over time, and what is known today might not be valid through time, making the characterization of anomalies class more challenging. In this case, semisupervised anomaly detection, that is, one-class classification, is preferred given that it only requires characterizing what is believed to be normal (target data instances) to detect the abnormal (nontarget data instances) [7]. Under certain circumstances, for example, medical domain, obtaining and demarcating the anomalous (nontarget) data instances can become very difficult, expensive, and time consuming, if not impossible [7,13]. For instance, assume a health diagnostic and monitoring system that detects health changes in an individual by tracking the individual's physiological parameters, where the current health status is examined based on a set of parameters, and raises a notification alarm when the individual health deteriorates [12]. In such a system, it becomes feasible to rely on a method that can be trained using only the regular or normal day measurements (target days) so as to detect deviation from normality [12,14]. This is because demarcating the exact boundaries between normal and abnormal health conditions is very challenging given that each pathogen has a different effect on the individual physiology. The one-class classifiers–based anomaly detection methods can be roughly grouped into 3 main groups: boundary and domain-based, density-based, and reconstruction-based methods based on how their internal function is defined and the approach used for minimization [3,10,12,13,15,16]. These models take into account different characteristics of the data set, and depending on the data set under consideration, these models could achieve different generalization performance, overfitting, and bias [12]. Unlike supervised and semisupervised anomaly detection methods, unsupervised methods do not require labeled instances to detect the anomaly (nontarget) instances because they rely on the entire data set to determine the anomalies and can be another possible alternative to semisupervised anomaly detection methods [7,10,12]. One of the drawbacks of unsupervised methods is that they require significant amount of data to achieve comparable performance. Both semisupervised and unsupervised methods have been used in various applications to detect anomalous instances [1,7,10,16]. In particular, these methods have been popular in the medical domain owing to their suitability for such applications, where there is lack of a sufficient data set for the other classes [13]. Accordingly, considering the difficulty and expense of obtaining enough sample data sets for the infection days from people with type 1 diabetes, a one-class classifier and unsupervised models are proposed for detecting infection incidence in people with type 1 diabetes.

Type 1 diabetes, also known as insulin-dependent diabetes, is a chronic disease of blood glucose regulation (hemostasis), and is caused by the lack of insulin secretion from pancreatic cells [17,18]. In people with type 1 diabetes, the incidence of infection often results in hyperglycemia and frequent insulin injection [19-26]. Infection-induced anomalies are characterized by violation of the norm of blood glucose dynamics, where blood glucose remains elevated despite taking a higher amount of insulin injection with less carbohydrate consumption [19]. Despite these potentials, there have been very few studies that focused on detecting infection incidence in individuals with type 1 diabetes using a dedicated personalized health model. Therefore, the objective of this study was to develop an algorithm, that is, a personalized health model that can automatically detect the incidence of infection in people with type 1 diabetes using blood glucose levels and insulin-to-carbohydrate ratio as input variables. For this, a one-class classifier and unsupervised models are proposed. The model is expected to detect any deviations from the norm because of infection incidences considering elevated blood glucose level (hyperglycemia incidences) coupled with unusual changes in the insulin-to-carbohydrate ratio, that is, frequent insulin injections and unusual reduction in the amount of carbohydrate intake [19]. Three groups of one-class classifiers

and two unsupervised density-based models were explored. A detailed theoretical description of the proposed models is given in Multimedia Appendix 1 [1,7-16,27-37]. The anomaly detection problem studied in this paper can be regarded as a contextual anomaly, where the ratio of insulin-to-carbohydrate is the context and the average blood glucose level is the behavioral attribute. This is mainly because of the fact that elevated blood glucose levels do not always signify being anomalies without looking at the context of the ratio of insulin-to-carbohydrate in this case. Throughout the paper, the term object is used to describe a feature vector incorporating the number of parameters under consideration. For example, an object X can define a specific event of an individual blood glucose dynamics at a specified time index *k* and is represented by a feature vector $X_k=(x_{k,1}, x_{k,2})$, where $x_{k,1}$ represents the ratio of total insulin-to-total carbohydrate and $x_{k,2}$ represents the average blood glucose level in a specific time-bin (interval) around *k*.

Methods

A group of one-class classifiers and unsupervised models were tested and compared. The one-class classifier incorporates 3 groups: boundary and domain-based, density-based, and reconstruction-based methods. The boundary and domain-based method contains support vector data description (SVDD) [27], one-class support vector machine (V-SVM) [28], incremental support vector machine [29], nearest neighbor (NN) [12], and minimum spanning tree (MST) [15]. Density-based method includes normal Gaussian [32], minimum covariance Gaussian [38], mixture of Gaussian (MOG) [32], Parzen [39], naïve Parzen [32], K-nearest neighbor (KNN) [12,30], and local outlier factor (LOF) [31]. The reconstruction-based method includes principal component analysis (PCA) [12,32], K-means [32], self-organizing maps (SOM) [12,32], and auto-encoder networks [12]. In addition, the unsupervised models were also tested, including the LOF [31,33] and the connectivity-based outlier factor (COF) [33,34]. The input variables, average blood glucose levels and ratio of total insulin (bolus) to total carbohydrate, used in training and testing of the models were selected in accordance with the description provided by Woldaregay et al [19], and the ratio was calculated by dividing the total insulin with the total carbohydrate within a specified time-bin. The data set consists of high-precision self-recorded data collected from 3 real subjects (2 males and 1 female; average age 34 [SD 13.2] years) with type 1 diabetes. It incorporates blood glucose levels, insulin, carbohydrate information, and self-reported infections cases of influenza (flu) and, mild and light common cold without fever, as shown in Table 1. Exemplar data depicting the model's input features for 2 specific patient years with and without infection are shown in Figures 1-4, and a more detailed description of the input features for 10-patient years with and without infection incidences can be found in Multimedia Appendix 2 [12,19]. The data were resampled and imputed in accordance with the description provided by Woldaregay et al [19], and the preprocessed data were smoothed using a moving average filter of 2 days' (48 hours) window size to remove short-term and small-scale features [19,40,41]. Feature scaling was carried out using min-max scaling [42] to normalize the data between 0 and 1, which is important to ensure that larger parameters do not dominate the smaller ones. The data sets are labeled as target and nontarget data sets, where the target data sets include all the self-recorded normal period of the year and the nontarget data set includes only the self-reported infection periods when the individual was sick. Accordingly, the one-class classifiers were trained using only the target data sets containing the regular or normal period of the year and tested using both the target and the nontarget (infection period) data sets. For the unsupervised models, all the data sets containing both the target and the nontarget data sets were presented during testing. The hyperparameters of most of the one-class classifiers were optimized using a consistency approach [43]. Models such as naïve Parzen and Parzen were optimized using the leave-one-out method. For MST, the entire MST was used. For PCA, the fraction of variance retained from the training data set was set to be 0.67. The models were evaluated based on different characteristics including data nature (with and without filter), data granularity (hourly and daily), data sample size, and required computational time. All the experiments were conducted using MATLAB 2018b (Mathworks, Inc). Most of the models were implemented using *ddtools*, *prtools*, and *anomaly detection toolbox*, which are MATLAB toolboxes [32,33,35].

Table 1. Equipments used in the self-management of diabetes.

Patients	Self-management				
	BG[a]	Insulin administration	Diet	Body weight (kg)	HbA_{1c}[b] (%)
Subject 1	Finger pricks recorded in the Diabetes Diary mobile app and Dexcom CGM[c]	Insulin Pen (multiple bolus and 1-time basal in the morning) recorded in the Diabetes Diary mobile app	Carbohydrate in grams recorded in the Diabetes Diary mobile app; level 3 (advanced carb counting)	83	6.0
Subject 2	Finger pricks recorded in the Spike mobile app and Dexcom G4 CGM[c]	Insulin Pen (multiple bolus [Humalog] and 1-time basal [Toujeo] before bed) recorded in the Spike mobile app	Carbohydrate in grams recorded in the Spike mobile app; level 3 (advanced carb counting)	77	7.3
Subject 3	Enlite (Medtronic) CGM[c] and Dexcom G4	Medtronic MinMed G640 insulin pump (basal rates profile [Fiasp] and multiple bolus [Fiasp])	Carbohydrate in grams recorded in pump information; level 3 (advanced carb counting)	70	6.2

[a]BG: blood glucose.

[b]HbA_{1c}: hemoglobin A_{1c}.

[c]CGM: continuous glucose monitoring.

Figure 1. Daily scatter plot of average blood glucose levels versus total insulin (bolus) to total carbohydrate ratio for a specific regular or normal patient year without any infection incidences.

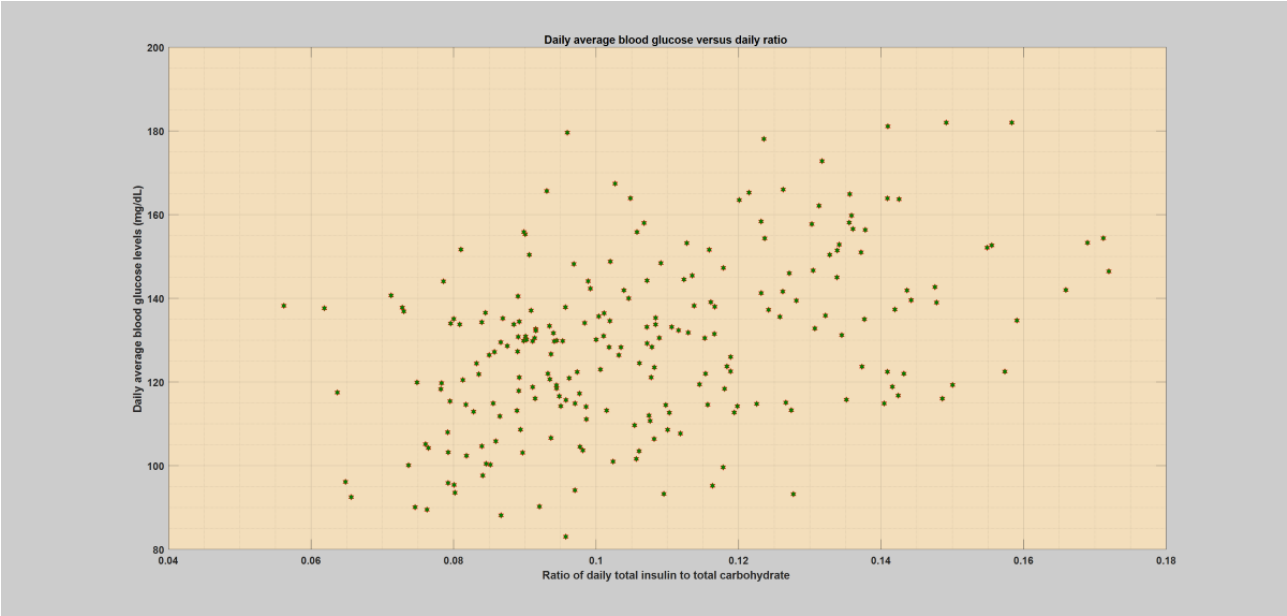


Figure 2. Hourly scatter plot of average blood glucose levels versus total insulin (bolus) to total carbohydrate ratio for a specific regular or normal patient year without any infection incidences.

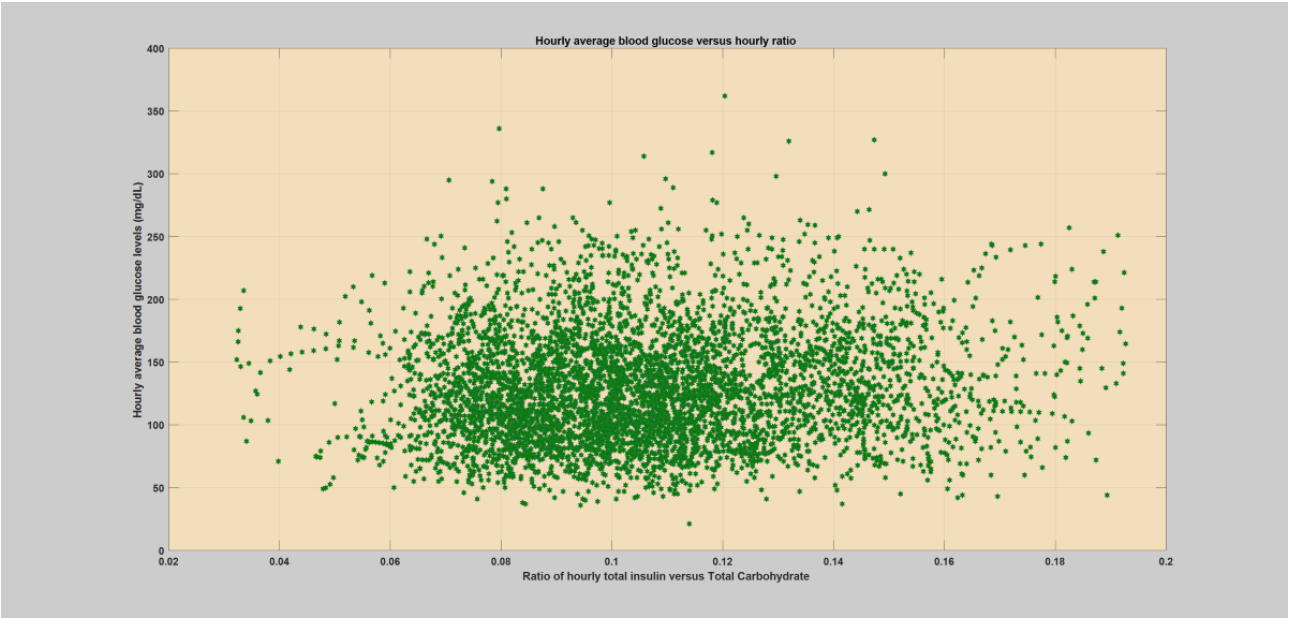


Figure 3. Daily scatter plot of average blood glucose levels versus total insulin (bolus) to total carbohydrate ratio for a specific patient year with an infection incidence (flu).

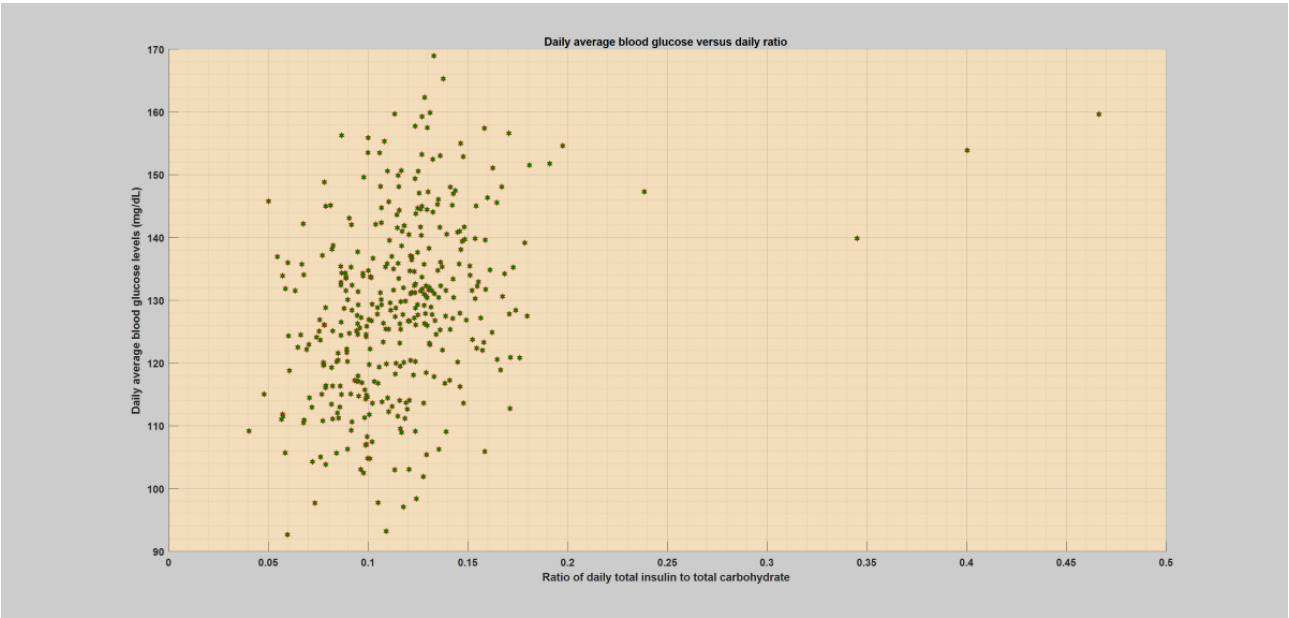
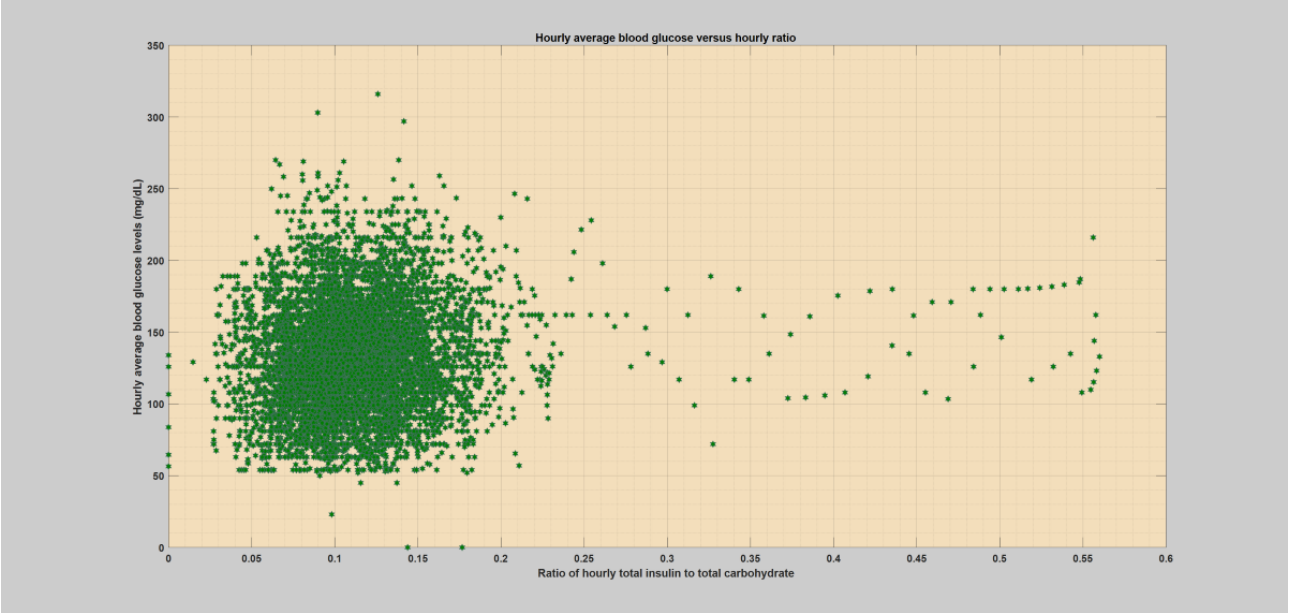


Figure 4. Hourly scatter plot of average blood glucose levels versus total insulin (bolus) to total carbohydrate ratio for a specific patient year with an infection incidence (flu).



Model Evaluation

The performance of the one-class classifiers was evaluated using 20 times 5-fold stratified cross-validation. For both daily and hourly cases, the user-specified outlier fraction threshold ß was set to 0.01 such that 1% of the training target data are allowed to be classified as outlier or get rejected [12]. Class imbalance was mitigated by oversampling of the nontarget data sets through random sampling [44]. Performance was measured using the area under the receiver operating characteristic (ROC) curve (AUC), specificity, and F1-score [45-48]. The AUC, specificity, and F1-score were reported as the average (SD) of twenty times five-fold stratified cross-validation rounds. AUC is the result of integration (summation) of the ROC curve over a range of possible classification thresholds [49]. It is regarded as robust (insensitive) when it comes to the presence of data imbalance; however, it is impractical for real-world implementation because it is independent of a single threshold [48]. Specificity measures the ratio of correctly classified negative samples from the total number of available negative samples [50]. Thus, it depicts the proportion of infection days (nontarget samples) that are correctly classified as such to the total number of infection days (period). It is only used to examine how the model performs in regard to the nontarget class (infection days). F1-score is the harmonic mean of precision and recall, where the value ranges from 0 to 1, and high F1 scores depict high classification performance [45]. F1-score is considered appropriate when evaluating model performance with regard to one target class and in the presence of unbalanced data sets [10,46-48]. The models were further compared based on various criteria, which can contribute to the implementation of the models in real-world settings, including computation time, sample size, number of user-defined parameters, and sensitivity to outliers in the training data sets:

- Computation time: this characteristic defines the amount of time taken to train and test the model. Regarding personal use, response time is crucial for acceptance of the services by a wide range of users. Furthermore, with regard to the outbreak detection settings, this is an important parameter given that a system that uses data from many participants needs to have an acceptable response time. However, in real-world applications, the training phase can be performed in an offline mode, which makes the testing response time very crucial.
- Sample size: this characteristic specifies the minimum amount of training data required to generate an acceptable performance. This is an important factor given that the system relies on self-recorded data; it is difficult to accumulate a large set of data for an individual initially.
- Number of user-defined parameters: this characteristic defines the complexity of the model. It is simpler and less data are required to estimate a model with fewer parameters. This is an important factor because it is easier for an individual to implement the simple model compared with the complex model.
- Sensitivity to outliers in the training data sets: this characteristic defines how the model estimation is affected by outliers in the training set. This is a crucial characteristic because the model training depends on self-reported data, which are highly dependent on the accuracy of the user data registration. It is possible that the user might forget to report some infection incidence and hence might be considered as target data sets and be used as a training data set. Furthermore, errors incurred during manual registration of data can also affect model generalization.

Data Collection and Ethical Declaration

The study protocol has been submitted to the Norwegian Regional Committees for Medical Health Research Ethics Northern Norway for evaluation and was found exempted from regional ethics review because it is outside the scope of medical research (reference number: 108435). Written consent was obtained, and the participants donated the data sets. All data from the participants were anonymized.

Results

The models were evaluated based on two different versions of the same data set: raw and filtered. The input variables to the models were the average blood glucose levels and the ratio of total insulin (bolus)-to-total carbohydrate. The necessary computational time for both training and testing of the models was also estimated. A comparison of the classifiers was carried out taking into account their performance, necessary sample size for producing acceptable performance, and computational time. These models were further compared based on their theoretical guarantee provided for robustness to outliers in the target data set and based on their complexity. In addition, these classifiers were compared with the unsupervised version of some selected models.

Model Evaluation

Model training and evaluations were carried out on an individual basis taking into account different characteristics of the data, specified time window or resolution (hourly and daily), and nature of the data (raw data and its smoothed version). For daily evaluation, we compared the performance of the models on raw data and its smoothed version with a 2-day moving average filter. For hourly evaluation, we compared the performance of the model on a smoothed version of the data set. The purpose of the comparison was to study the performance gain achieved by removing short-time noises from the data set through smoothing. The average and SD of AUC, specificity, and F1-score are computed and reported for each model. The top performing models from each category are highlighted in italics within each tables.

Semisupervised Models

The regular or normal days were labeled as the target class data set and the infection period as the nontarget class data set. Three groups of one-class classifiers were trained on the target class and tested on a data set containing both the target and the nontarget classes. In addition to the data characteristics stated above, resolution and data nature, the one-class classifier performance was also assessed taking into account the required sample object size to produce acceptable data description. In this direction, we consider four groups of sample size: 1 month, 2 months, 3 months, and 4 months data sets. In the model evaluation, the data set containing the infection period was presented during testing. The evaluation was carried out based on 20 times 5-fold stratified cross-validation. The performance of the model was reported as the average and SD of AUC, specificity, and F1-score of the rounds. A score plot of each model for both the hourly and the daily scenarios using the smoothed version of the data can be found in Multimedia Appendix 3, where the models were trained on random 120 regular or normal days of the patient year and tested over the whole year.

Daily

As can be seen in Tables 2 and 3 below (see also Multimedia Appendix 4), the performance of the models generally improves as the size of the sample increases. The models performed well with respect to the raw data sets; however, the performance significantly improved with the smoothed version of the data. The results indicate that the sample size greatly affects the model performance and that there is a larger variation in performance when the training data set is small. Generally, it can be seen that the models generalize well with the 3-month data set (90 sample objects) and further improve after 3 months. In general, on average, with both the raw and smoothed data sets, the boundary and domain-based method performed better with a small sample size. As the sample size increased, all the three groups produced comparable descriptions of the data. From each respective category, models such as V-SVM, K-NN, and K-means performed well across all the sample sizes.

First Case of Infection (Flu)

The boundary and domain-based method achieved a better description of the data with a small sample size when compared with the other two groups. However, as the sample size increased, all the three groups achieved relatively comparable descriptions of the data. Specific models such as V-SVM, K-NN, and K-means performed better from their respective group. Regarding the raw data, as seen in Table 2, all the models failed to generalize from the 1-month data set as compared with the large sample objects, that is, 3 months, which was expected:

1. From the boundary and domain-based method, V-SVM performed better in all the sample sizes and achieved comparable performance even with 60 objects and improved significantly afterward. SVDD produced a comparable description with higher sample sizes, that is, 3 months and later.
2. From the density-based method, K-NN performed better in all the sample sizes and achieved better performance even with 60 objects. Naïve Parzen produced comparable performance with higher sample sizes, that is, 3 months and later.
3. From the reconstruction-based method, K-means achieved better performance for all sample sizes.

Smoothing the data, as shown in Table 3, improved the model performance even with 30 sample objects:

1. From the boundary and domain-based method, V-SVM achieved better performance in all sample sizes.
2. From the density-based method, K-NN achieved better performance for all sample sizes, minimum covariance determinant (MCD) Gaussian produced a comparable description with 30 and 60 sample objects, and naïve Parzen achieved comparable description of the data with 4-month sample objects.
3. Regarding the reconstruction-based method, PCA achieved good performance with 30 and 60 sample objects, whereas K-means performed better with larger sample objects.

Table 2. Average (SD) of area under the receiver operating characteristic curve, specificity, F1-score for the raw data set (without smoothing), and different sample size. Fraction=0.01.

Models	1 month			2 months			3 months			4 months		
	AUC[a], mean (SD)	Specificity, mean (SD)	F1, mean (SD)	AUC, mean (SD)	Specificity, mean (SD)	F1, mean (SD)	AUC, mean (SD)	Specificity, mean (SD)	F1, mean (SD)	AUC, mean (SD)	Specificity, mean (SD)	F1, mean (SD)
Boundary and domain–based method												
SVDD[b]	90.7 (8.8)	71.7 (7.7)	73.6 (5.5)	93.4 (6.2)	81.7 (5.0)	87.4 (8.1)	96.4 (2.9)	87.8 (3.3)	91.3 (6.0)	94.6 (3.7)	81.7 (5.0)	90.0 (4.6)
IncSVDD[c]	90.4 (8.9)	66.7 (7.5)	72.7 (4.9)	91.8 (5.9)	66.7 (7.5)	84.4 (3.2)	95.8 (2.9)	70.0 (7.1)	85.4 (1.2)	93.7 (3.6)	55 (10.7)	81.0 (2.7)
V-SVM[d]	93.1 (6.0)	63 (10.6)	*78.9 (6.2)* [e]	96.5 (2.3)	81.9 (4.7)	*90.7 (3.4)*	97.9 (1.5)	88.9 (0.0)	*94.1 (2.0)*	96.2 (2.3)	83.3 (0.0)	*91.7 (1.4)*
NN[f]	74.2 (9.3)	38.3 (7.7)	61.0 (4.7)	89.5 (9.3)	20.0 (6.7)	70.0 (4.6)	90.1 (6.6)	11.1 (18)	69.2 (3.8)	92.8 (3.3)	33.3 (0.0)	75.1 (0.4)
MST[g]	89.4 (8.1)	50.0 (0.0)	62.7 (6.6)	95.4 (5.6)	61.7 (7.7)	82.3 (5.9)	96.6 (2.7)	68.9 (4.5)	83.6 (4.7)	94.1 (2.8)	55.0 (7.7)	80.6 (2.3)
Density-based method												
Gaussian	90.6 (7.1)	60.0 (8.2)	68.8 (8.4)	95.4 (4.6)	70.0 (6.7)	85.3 (4.6)	97.3 (2.5)	80.0 (4.5)	89.2 (3.3)	95.5 (3.2)	66.7 (0.0)	84.5 (2.0)
MOG[h]	88.1 (9.9)	80.1 (17.3)	67.8 (16.4)	93.1 (7.1)	75.8 (14.8)	82.5 (10.1)	95.6 (3.4)	80.2 (7.5)	86.0 (6.7)	93.7 (3.9)	68.7 (11.6)	84.2 (5.7)
MCD[i] Gaussian	89.0 (8.5)	55.0 (7.7)	66.4 (9.0)	94.0 (4.6)	68.3 (5.0)	84.6 (6.3)	97.0 (2.7)	80.0 (4.5)	89.9 (2.4)	94.5 (3.2)	65.0 (5.0)	84.0 (3.2)
Parzen	89.0 (9.2)	70.0 (6.7)	70.7 (5.9)	94.6 (4.9)	83.3 (0.0)	87.9 (6.3)	97.2 (2.4)	88.9 (0.0)	90.5 (5.9)	95.2 (2.9)	83.3 (0.0)	88.9 (3.3)
Naïve Parzen	90.1 (7.6)	55 (10.7)	65.0 (5.0)	95.7 (3.9)	76.7 (8.2)	87.2 (3.5)	98.3 (1.4)	88.9 (0.0)	*93.6 (2.4)*	96.8 (2.1)	83.3 (0.0)	90.7 (2.0)
K-NN[j]	91.8 (6.9)	50.0 (0.0)	66.0 (2.0)	95.6 (3.1)	81.7 (5.0)	*90.9 (3.2)*	97.9 (1.6)	88.9 (0.0)	93.5 (3.7)	97.0 (2.2)	83.3 (0.0)	*92.0 (1.0)*
LOF[k]	88.5 (6.1)	66.7 (7.5)	*72.7 (4.9)*	97.0 (1.9)	71.7 (7.7)	86.1 (2.4)	96.8 (2.8)	78.9 (3.3)	88.7 (2.8)	92.6 (4.8)	50.0 (0.0)	79.3 (2.6)
Reconstruction-based method												
PCA[l]	87.8 (11.9)	50.0 (7.5)	62.4 (8.5)	93.5 (6.2)	51.7 (5.0)	78.2 (4.1)	93.6 (4.7)	60 (10.2)	81.8 (4.4)	91.3 (5.2)	46.7 (6.7)	78.7 (2.3)
Auto-encoder	82.2 (12.0)	57.9 (15.3)	64.7 (12.0)	88.2 (9.5)	61.6 (14.0)	81.4 (7.1)	93.4 (5.7)	74.4 (11)	86.4 (5.9)	88.4 (8.8)	61.3 (14.3)	82.7 (5.7)
SOM[m]	86.9 (9.4)	78.3 (13.3)	66.7 (16.9)	92.8 (7.3)	64.2 (12.4)	80.9 (7.0)	95.8 (3.7)	80.1 (6.3)	86.9 (5.5)	92.2 (4.1)	76.5 (9.0)	87.5 (4.5)
K-means	91.8 (6.9)	65.0 (9.0)	*71.8 (5.1)*	96.0 (2.4)	83.3 (0.0)	*91.5 (2.8)*	97.6 (1.6)	88.9 (0.0)	*93.5 (3.7)*	96.2 (2.2)	83.3 (0.0)	*91.5 (1.6)*

[a]AUC: area under the receiver operating characteristic curve.

[b]SVDD: support vector data description.

[c]IncSVDD: incremental support vector data description.

[d]V-SVM: one-class support vector machine.

[e]Italicized values indicates the top performing models.

[f]NN: nearest neighbor.

[g]MST: minimum spanning tree.

[h]MOG: mixture of Gaussian.

[i]MCD: minimum covariance determinant.

[j]K-NN: K-nearest neighbor.

[k]LOF: local outlier factor.

[l]PCA: principal component analysis.
[m]SOM: self-organizing maps.

Table 3. Average of area under the receiver operating characteristic curve, specificity, and F1-score for smoothed version of the data with a 2-day moving average filter and different sample size. Fraction=0.01.

Models	1 month			2 months			3 months			4 months		
	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1
Boundary and domain–based method												
SVDD[b]	99.6 (1.3)	100 (0.0)	93.6 (15.2)	100 (0.0)	100 (0.0)	94.8 (10.1)	100 (0.0)	100 (0.0)	97.0 (4.1)	100 (0.0)	100 (0.0)	96.9 (4.0)
IncSVDD[c]	99.6 (1.3)	100 (0.0)	93.6 (15.2)	100 (0.0)	100 (0.0)	97.1 (6.3)	100 (0.0)	100 (0.0)	97.6 (4.1)	100 (0.0)	100 (0.0)	98.3 (2.8)
V-SVM[d]	100 (0.0)	99.5 (2.9)	*98.9 (3.2)* [e]	100 (0.0)	100 (0.0)	*99.1 (2.6)*	100 (0.0)	100 (0.0)	*99.4 (1.7)*	100 (0.0)	100 (0.0)	*99.6 (1.2)*
NN[f]	98.1 (3.9)	58.3 (15.4)	72.3 (9.9)	86.9 (12.5)	16.7 (22.4)	70.5 (5.3)	88.1 (6.5)	54.4 (22.5)	80.0 (8.6)	92.4 (5.3)	8.3 (17.1)	69.0 (4.8)
MST[g]	98.5 (2.4)	85.0 (5.0)	85.5 (2.1)	99.7 (0.8)	100 (0.0)	97.1 (6.3)	99.9 (0.4)	97.8 (4.5)	97.2 (4.0)	99.7 (0.8)	100 (0.0)	97.0 (7.9)
Density-based method												
Gaussian	100 (0.0)	98.3 (5.0)	92.1 (15.2)	100 (0.0)	100 (0.0)	97.1 (6.3)	99.8 (0.7)	100 (0.0)	97.6 (4.1)	99.4 (1.7)	100 (0.0)	97.0 (7.9)
MOG[h]	98.6 (3.2)	99.8 (1.7)	88.5 (16.8)	99.6 (1.2)	100 (0.0)	92.2 (11.1)	99.7 (0.7)	99.8 (1.4)	94 (10.3)	99.3 (2.0)	99.9 (1.2)	94.4 (11.8)
MCD[i] Gaussian	98.9 (2.2)	91.7 (8.4)	*90.9 (7.7)*	100 (0.0)	100 (0.0)	*98.0 (6.0)*	99.5 (1.1)	96.7 (5.1)	96.6 (5.9)	99.4 (1.7)	88.3 (7.7)	92.0 (6.8)
Parzen	99.6 (1.3)	100 (0.0)	87.7 (17.0)	100 (0.0)	100 (0.0)	95.1 (8.0)	100 (0.0)	100 (0.0)	94.6 (9.8)	99.9 (0.4)	100 (0.0)	94.6 (12.3)
Naïve Parzen	99.2 (2.5)	100 (0.0)	94.7 (11.1)	100 (0.0)	100 (0.0)	93.8 (11.0)	99.6 (1.1)	100 (0.0)	97.5 (5.0)	100 (0.0)	100 (0.0)	*98.7 (2.7)*
K-NN[j]	98.1 (3.9)	68.3 (5.0)	75.2 (4.3)	100 (0.0)	100 (0.0)	*98.0 (6.0)*	100 (0.0)	100 (0.0)	*98.8 (3.8)*	100 (0.0)	100 (0.0)	*97.7 (4.7)*
LOF[k]	98.6 (2.9)	75.0 (13.5)	80.2 (10.8)	100 (0.0)	100 (0.0)	*98.0 (6.0)*	100 (0.0)	100 (0.0)	96.9 (5.0)	99.7 (0.8)	100 (0.0)	97.4 (7.9)
Reconstruction-based method												
PCA[l]	98.9 (2.2)	85.0 (5.0)	*85.5 (2.1)*	99.2 (1.3)	85.0 (5.0)	*91.4 (2.7)*	98.6 (1.9)	88.9 (0.0)	92.2 (6.0)	97.8 (2.2)	83.3 (0.0)	89.1 (9.7)
Auto-encoder	97.4 (6.0)	89.1 (13.0)	86.0 (14.2)	98.5 (3.2)	94.5 (9.6)	91.8 (9.4)	99.2 (2.4)	93.7 (10.2)	93.7 (8.3)	98.6 (3.8)	94.4 (9.5)	93.7 (9.7)
SOM[m]	99.3 (1.9)	99.9 (1.2)	84.7 (19.8)	99.8 (0.7)	100 (0.0)	91.4 (9.6)	99.9 (0.3)	100 (0.0)	95.2 (7.9)	99.6 (1.3)	100 (0.0)	93.4 (12.1)
K-means	99.2 (2.5)	85.0 (11.7)	87.0 (10.4)	100 (0.0)	100 (0.0)	97.1 (6.3)	100 (0.0)	100 (0.0)	*98.8 (3.8)*	100 (0.0)	100 (0.0)	*99.2 (2.5)*

[a]AUC: area under the receiver operating characteristic curve.

[b]SVDD: support vector data description.

[c]IncSVDD: incremental support vector data description.

[d]V-SVM: one-class support vector machine.

[e]Italicized values indicates the top performing models.

[f]NN: nearest neighbor.

[g]MST: minimum spanning tree.

[h]MOG: mixture of Gaussian.

[i]MCD: minimum covariance determinant.

[j]K-NN: K-nearest neighbor.

[k]LOF: local outlier factor.

[l]PCA: principal component analysis.
[m]SOM: self-organizing maps.

Second Case of Infection (Flu)

The boundary and domain-based method achieved better performance with a small sample size compared with the density and reconstruction-based methods. However, as the sample size increased, all the three groups achieved comparable performance. The detailed numerical values of comparison are given in Multimedia Appendix 4. Specific models such as V-SVM**,** K-NN, and K-means performed better from their respective group. Regarding the raw data, all the models failed to generalize from the 1-month data set as compared with the higher sample objects, that is, 3 months (Multimedia Appendix 4):

1. From the boundary and domain-based method, SVDD, MST, and incremental support vector data description (incSVDD) performed better with a larger sample object, and V-SVM achieved better description with 30 sample objects.
2. From the density-based method, all the models exhibited similar performance. Naïve Parzen and K-NN, with only 60 sample objects, achieved comparable performance with the higher sample objects.
3. From the reconstruction-based method, K-means achieved better performance for all sample sizes.

Smoothing the data significantly improved the performance of the model even with 30 objects, compared with the raw data (Multimedia Appendix 4):

1. From the boundary and domain-based method, the V-SVM achieved higher performance in all the sample sizes.
2. From the density-based method, LOF achieved better description with small sample objects, and K-NN produced better description with all the sample sizes. Gaussian families achieved improved and comparable performance with increased sample objects. Among them, K-NN with only 60 objects achieved comparable performance with larger sample objects.
3. Regarding the reconstruction-based method, K-means and SOM achieved better performance, whereas K-means performed better in all the sample sizes.

Third Case of Infection (Flu)

The boundary and domain-based method achieved better performance with a small sample size compared with the density and reconstruction-based methods. However, as the sample size increased, all the three groups produced comparable descriptions. The detailed numerical values of comparison are given in Multimedia Appendix 4. Specific models such as V-SVM, MST, LOF, and PCA performed better from their respective group. Regarding the raw data, surprisingly, in contrast to the previous two infection cases, all the models achieved higher generalization from the 1-month data set (Multimedia Appendix 4):

1. From the boundary and domain-based method, SVDD, V-SVM, MST, and incSVDD performed better in all the cases, with MST achieving better performance.
2. From the density-based method, normal and MCD Gaussian achieved better description of the data with 1-month sample objects. K-NN and LOF performed better with sample sizes larger than 1-month sample objects, and LOF outperformed all sample sizes. The LOF with only 60 objects achieved comparable performance with the higher sample objects.
3. From the reconstruction-based method, PCA produced better description for all sample sizes, whereas K-means and SOM achieved comparable performance with sample size larger than 1-month sample objects.

Smoothing the data allowed the models to generalize well and significantly improved the performance of the model even with 30 objects, compared with the raw data (Multimedia Appendix 4):

1. From the boundary and domain-based method, the V-SVM and MST achieved higher performance in all the sample sizes, whereas V-SVM outperformed all the models.
2. From the density-based method, the Gaussian families, LOF, and K-NN achieved better performance, whereas LOF achieved better performance in all sample sizes.
3. Regarding the reconstruction-based method, K-means and PCA achieved better performance, whereas PCA performed better in all the sample sizes.

Fourth Case of Infection (Flu)

The boundary and domain-based method achieved better performance with small sample sizes compared with the density and reconstruction-based methods. All the three groups improved with increasing sample size. The detailed numerical values of comparison are given in Multimedia Appendix 4. Specific models such as V-SVM, LOF, and K-means performed better from their respective group. Regarding the raw data, surprisingly, in contrast to all the previous three infection cases, all the models achieved higher generalization from the 1-month data set (Multimedia Appendix 4):

1. From the boundary and domain-based method, SVDD, V-SVM, and incSVDD performed better for all the sample sizes.
2. From the density-based method, MCD Gaussian performed better with a 1-month sample size, and all the models produced comparable descriptions as the sample size increased, whereas the LOF performed better for all the sample sizes.
3. From the reconstruction-based method, PCA performed relatively better for all the sample sizes, and K-means and SOM achieved comparable performance with a larger sample size.

Smoothing the data significantly improved the model performance even with 30 objects compared with the raw data (Multimedia Appendix 4):

1. From the boundary and domain-based method, the V-SVM achieved higher performance in all the sample sizes. As the sample size increased, the incSVDD and MST achieved comparable performance.

2. From the density-based method, K-NN and LOF produced better descriptions with a 1-month sample size. K-NN performed better in almost all sample sizes.
3. From the reconstruction-based method, K-means achieved better performance for all sample sizes.

Hourly

As can be seen in Table 4 (see also Multimedia Appendix 4), the performance of the model generally improved as more training sample data were presented. The models produced comparable performance even with the 1-month data set compared with the daily scenario. This is mainly because of the presence of more samples per day (24 samples per day), which enables the models to reach a better generalization. Generally, the results indicate that the models generalize well after 2 months. Both the boundary and domain-based method and reconstruction-based method achieved better performance even with a 1-month sample size. However, the density-based method suffers from large variation with 1-month training samples. In general, the boundary and domain-based method performed better in all the infection cases compared with the other two methods. In addition, specific models such as V-SVM, K-NN, and K-means performed well from their respective groups.

First Case of Infection (Flu)

The boundary and domain-based method achieved better performance compared with the density and reconstruction-based methods. As can be seen in Table 4, the boundary and domain-based method achieved better generalization from the 1-month data set. Specific models such as V-SVM, K-NN, and K-means performed better from their respective group:

1. From the boundary and domain-based method, V-SVM achieved better description in all sample sizes, whereas SVDD, incSVDD, and V-SVM achieved comparable performance with a larger sample size.
2. From the density-based method, Gaussian families and naïve Parzen performed better at large sample sizes, whereas K-NN and LOF achieved better performance in all the sample sizes. K-NN outperformed all the models.
3. From the reconstruction-based method, K-means performed better in all the sample sizes, and all the other models performed better with larger sample sizes.

Table 4. Average (SD) of area under the receiver operating characteristic curve, specificity, F1-score for the smoothed version of the data with a 48-hour moving average filter and different sample size. Fraction=0.01.

Models	1 month			2 months			3 months			4 months		
	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1
Boundary and domain–based method												
SVDD[b]	97.6 (1.9)	83.2 (3.4)	85.8 (1.7)	97.8 (1.2)	85.7 (5.0)	90.5 (9.6)	97.7 (1.2)	90.4 (5.1)	94.2 (2.9)	98.1 (0.9)	91.0 (3.7)	96.8 (0.9)
IncSVDD[c]	97.4 (1.9)	84.5 (2.8)	86.8 (1.9)	97.7 (1.2)	86.7 (2.0)	93.9 (1.0)	97.5 (1.2)	88.5 (1.5)	96.0 (1.1)	97.9 (0.9)	88.9 (1.2)	*97.1 (0.7)*
V-SVM[d]	98.1 (2.1)	84.5 (1.1)	*90.5 (1.1)* [e]	99.0 (1.1)	92.6 (0.0)	*96.1 (1.3)*	99.5 (0.6)	93.8 (0.5)	*96.9 (1.4)*	99.4 (0.4)	94.2 (0.0)	97.1 (1.3)
NN[f]	84.8 (6.0)	75.9 (4.5)	74.8 (6.0)	89.3 (2.2)	76.5 (4.1)	87.1 (3.3)	89.0 (4.0)	77.5 (3.9)	89.3 (4.4)	90.2 (4.7)	77.5 (3.8)	91.4 (6.4)
MST[g]	90.5 (3.1)	85.4 (3.9)	67.6 (14.5)	94.4 (2.0)	85.7 (4.0)	85.1 (7.0)	94.7 (2.4)	88.8 (3.5)	87.8 (8.5)	95.8 (2.2)	88.8 (3.0)	90.9 (5.9)
Density–based method												
Gaussian	98.1 (2.2)	79.8 (4.9)	83.9 (2.7)	99.5 (0.9)	90.1 (1.7)	95.2 (1.8)	99.6 (0.7)	92.9 (1.3)	97.1 (2.5)	99.5 (0.5)	92.2 (1.0)	97.7 (1.1)
MOG[h]	95.8 (3.6)	82.7 (4.3)	83.7 (5.0)	98.3 (1.5)	86.2 (2.7)	92.3 (2.7)	98.7 (1.4)	88.7 (4.6)	94.7 (3.5)	98.6 (1.6)	88.2 (3.1)	95.3 (3.2)
MCD[i] Gaussian	98.6 (2.1)	75.3 (6.9)	81.3 (2.5)	99.6 (0.9)	89.6 (1.9)	95.0 (1.8)	99.6 (0.7)	92.5 (1.8)	97.0 (2.3)	99.6 (0.4)	92.0 (1.2)	97.7 (1.1)
Parzen	91.9 (2.9)	93.6 (2.0)	63.4 (16.5)	96.2 (2.3)	94.4 (2.0)	81.6 (10.2)	96.6 (2.6)	94.8 (1.7)	84.2 (9.5)	97.4 (2.2)	95.6 (1.2)	87.9 (7.1)
Naïve Parzen	94.8 (3.7)	76.4 (5.6)	77.6 (7.9)	98.7 (1.2)	85.2 (3.3)	91.8 (2.9)	99.1 (1.1)	89.1 (3.8)	94.8 (2.5)	98.9 (0.9)	89.7 (2.4)	96.2 (1.6)
K-NN[j]	97.1 (3.4)	78.8 (2.0)	*84.2 (2.1)*	99.1 (1.0)	92.9 (0.7)	*96.0 (1.8)*	99.6 (0.4)	93.8 (0.7)	*97.3 (1.9)*	99.5 (0.3)	94.0 (0.6)	*98.2 (0.9)*
LOF[k]	96.9 (3.5)	78.3 (3.0)	84.2 (2.4)	99.2 (1.1)	91.9 (0.9)	*96.0 (1.8)*	99.6 (0.5)	93.7 (0.8)	97.3 (2.1)	99.5 (0.4)	93.1 (0.4)	97.8 (1.2)
Reconstruction–based method												
PCA[l]	97.1 (3.4)	63.9 (8.8)	75.4 (0.3)	99.4 (1.2)	76.4 (6.6)	90.2 (1.1)	99.1 (1.3)	75.1 (6.8)	92.4 (1.1)	98.9 (1.2)	69.1 (4.1)	93.1 (0.8)
Auto-encoder	92.0 (4.8)	79.5 (7.6)	78.9 (8.3)	96.2 (2.6)	83.1 (7.2)	91.1 (3.9)	96.3 (3.2)	84.3 (7.7)	92.7 (5.0)	96.7 (3.0)	84.0 (8.0)	94.6 (4.4)
SOM[m]	94.1 (2.3)	82.2 (3.3)	82.6 (4.9)	95.6 (1.1)	82.9 (3.1)	91.6 (1.9)	94.8 (2.3)	83.4 (5.8)	92.3 (4.1)	95.5 (1.9)	84.1 (3.8)	94.3 (3.8)
K-means	97.3 (3.2)	80.9 (2.5)	*85.5 (2.5)*	98.9 (1.1)	92.6 (0.7)	*95.8 (1.8)*	99.3 (0.6)	92.9 (0.7)	*97.3 (1.4)*	99.4 (0.4)	94.1 (0.2)	*98.1 (1.1)*

[a]AUC: area under the receiver operating characteristic curve.

[b]SVDD: support vector data description.

[c]IncSVDD: incremental support vector data description.

[d]V-SVM: one-class support vector machine.

[e]Italicized values indicates the top performing models.

[f]NN: nearest neighbor.

[g]MST: minimum spanning tree.

[h]MOG: mixture of Gaussian.

[i]MCD: minimum covariance determinant.

[j]K-NN: K-nearest neighbor.

[k]LOF: local outlier factor.

[l]PCA: principal component analysis.

[m]SOM: self-organizing maps.

Second Case of Infection (Flu)

The boundary and domain-based method and reconstruction-based method achieved better performance for all sample sizes compared with the density-based method. Specifically, the boundary and domain-based method achieved better generalization from the 1-month data set. The detailed numerical values of comparison are given in Multimedia Appendix 4. Specific models such as V-SVM, K-NN, and K-means performed better from their respective group:

1. From the boundary and domain-based method, V-SVM achieved better description for all the sample sizes, and SVDD, NN, and incSVDD improved with larger training sample size; however, V-SVM outperformed all the models for all the sample sizes.
2. From the density-based method, normal and MCD Gaussian performed better with the 1- and 2-month sample sizes, and models such as K-NN performed better on all the sample sizes, whereas naïve Parzen outperformed all the models with the 3- and 4-month data sets.
3. From the reconstruction-based method, K-means produced better description for all the sample sizes and the auto-encoder and SOM performed better with larger sample sizes.

Third Case of Infection (Flu)

Generally, in comparison, all the groups performed better at large training sample sizes; however, the boundary and domain-based method achieved better performance with small training sample sizes. It achieved comparable generalization from the 1-month data set. The detailed numerical values of comparison are given in Multimedia Appendix 4. Specific models such as V-SVM, families that utilize nearest neighbor distance (K-NN and LOF), and PCA performed better from their respective group:

1. From the boundary and domain-based method, SVDD, NN, MST, incSVDD, and V-SVM achieved better performance at larger training sample sizes, whereas V-SVM outperformed all the models for all the sample sizes.
2. From the density-based method, the Gaussian families, K-NN, LOF, and naïve Parzen achieved better performance at larger training sample sizes, whereas K-NN and LOF outperformed all the models for all the sample sizes.
3. From the reconstruction-based method, K-means, PCA, auto-encoder, and SOM achieved better performance at larger training sample sizes, whereas PCA performed better for all sample sizes.

Fourth Case of Infection (Flu)

Generally, in comparison, all the group performed better at large training sample size; however, the boundary and domain-based method achieved better performance with small training sample sizes, for example, 1-month data set. It achieved comparable generalization from the 1-month data set. The detailed numerical values of comparison are given in Multimedia Appendix 4. Specific models such as V-SVM, Gaussian families (Gaussian, MOG, and MCD Gaussian), and PCA performed better from their respective groups:

1. From the boundary and domain-based method, NN, incSVDD, and V-SVM achieved better performance at larger training sample sizes, whereas V-SVM outperformed all the models for all the sample sizes.
2. From the density-based method, Gaussian families, K-NN, LOF, and naïve Parzen achieved better performance at larger training sample sizes, whereas Gaussian families outperformed all the models for all the sample sizes.
3. From the reconstruction-based method, K-means, SOM, auto-encoder, and PCA achieved better performance at larger training sample sizes, whereas PCA performed better for all sample sizes.

Average Performance Across all the Infection Cases

The average performances of the models across all the infection cases for different sample sizes, levels of data granularity (hourly and daily), and nature of data (raw and smoothed) are shown in Tables 5-7. In general, the boundary and domain-based method performed better than the other two groups in both daily and hourly smoothed data sets; however, all the groups achieved comparable performance with respect to the daily raw data set. Specific models such as V-SVM, K-NN, and K-means performed better in all these circumstances.

Daily Raw Data Set

Regarding the daily raw data set, as shown in Table 5, specific models such as V-SVM, MCD Gaussian, K-NN, and K-means produced relatively better descriptions of the 1-month data. For the 2-month sample size, models such as incSVDD, K-NN, LOF, and K-means achieved better performance. For the 3-month sample size, SVDD, incSVDD, V-SVM, Gaussian, MCD Gaussian, K-NN, LOF, and K-means produced comparable descriptions. As expected, SVDD and most of the density-based method improved with larger training sizes. For the 4-month sample size, almost all the models produced much improved performance. In the group comparison, all three groups produced comparable descriptions in all the sample sizes.

Table 5. Average performance of each model across all the infection cases for the daily raw data set (without smoothing) and different sample sizes. Fraction=0.01.

Models	1 month			2 months			3 months			4 months		
	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1
Boundary and domain-based method												
SVDD[b]	87.1 (11)	66.0 (13.5)	74.8 (9.5)	91.7 (7.3)	61.7 (10.6)	84.1 (5.5)	93.3 (4.6)	67.3 (10.5)	86.2 (4.4)	91.4 (4.3)	61.7 (10.6)	*85.7 (4.1)*[c]
IncSVDD[d]	85.2 (11)	63.0 (4.6)	74.7 (10.4)	90.5 (8.5)	57.9 (11)	*83.8 (3.6)*	92.8 (5.1)	62.8 (10.9)	84.9 (3.2)	90.8 (4.4)	55.0 (11.7)	83.5 (3.7)
V-SVM[e]	91.5 (8.0)	55.7 (7.0)	*77.4 (6.4)*	92.2 (5.1)	60.6 (5.0)	82.8 (4.5)	94.2 (3.8)	66.9 (6.1)	*86.6 (3.5)*	93.8 (4.1)	63.1 (11.9)	84.5 (5.1)
NN[f]	73.4 (12)	31.3 (6.5)	65.0 (5.4)	72.1 (11.9)	25.0 (9.6)	75.7 (3.7)	70.8 (11.2)	8.6 (17.6)	72.0 (4.7)	70.0 (9.0)	16.0 (14.4)	75.7 (3.4)
MST[g]	82.4 (8.7)	52.1 (0.0)	71.2 (6.1)	82.6 (9.1)	50.4 (9.0)	82.0 (5.1)	84.0 (6.3)	56.2 (9.3)	82.9 (3.5)	84.2 (6.6)	50.0 (11.4)	82.6 (2.7)
Density-based method												
Gaussian	91.5 (9.9)	56.9 (7.7)	72.9 (7.8)	93.6 (6.1)	58.8 (10.9)	84.0 (4.0)	95.1 (4.3)	65.3 (10.6)	86.3 (3.2)	95.0 (3.5)	57.9 (10.3)	84.6 (3.2)
MOG[h]	89.9 (12)	69.2 (11.9)	71.3 (14.3)	91.7 (6.1)	64.1 (14.0)	83.8 (6.8)	94.0 (4.4)	67.0 (11.4)	85.0 (5.6)	94.5 (3.7)	61.6 (12.6)	84.9 (5.1)
MCD[i] Gaussian	90.8 (9.1)	54.0 (5.5)	*72.0 (6.8)*	93.1 (6.0)	58.0 (8.1)	84.1 (4.3)	95.3 (4.2)	65.3 (10.6)	86.4 (3.0)	94.8 (3.5)	57.9 (10.6)	84.9 (3.0)
Parzen	89.7 (10)	59.6 (8.3)	70.6 (9.4)	91.7 (6.5)	62.1 (10.3)	83.9 (5.3)	93.9 (5.0)	68.7 (11.2)	85.6 (5.4)	94.3 (3.8)	66.1 (12.7)	86.1 (3.8)
Naïve Parzen	88.1 (8.7)	54.2 (6.5)	69.1 (9.6)	90.2 (7.1)	60.4 (11.2)	83.7 (4.9)	91.9 (5.5)	66.5 (12.8)	86.6 (4.4)	92.8 (4.7)	64.6 (10.0)	*86.9 (3.4)*
K-NN[j]	91.1 (7.8)	52.9 (5.1)	71.6 (7.9)	91.6 (5.0)	61.1 (11.3)	*85.9 (3.1)*	94.8 (4.8)	66.9 (11.2)	*87.1 (3.2)*	95.0 (3.8)	62.1 (10.3)	*86.5 (3.3* **)**
LOF[k]	89.2 (8.9)	56.3 (3.9)	73.0 (8.6)	92.4 (6.0)	59.2 (11.1)	84.9 (2.8)	94.0 (4.8)	64.4 (11.4)	86.2 (2.8)	93.7 (4.3)	53.8 (10.3)	83.8 (2.5)
Reconstruction-based method												
PCA[l]	87.6 (8.8)	58.8 (4.6)	73.7 (8.3)	90.2 (6.4)	55.0 (6.8)	82.7 (4.5)	91.4 (4.9)	59.7 (6.2)	84.1 (3.2)	90.5 (4.5)	53.8 (7.2)	83.6 (2.9)
Auto-encoder	83.6 (14)	58.3 (17.7)	71.0 (12.5)	84.6 (12.5)	53.1 (20.0)	82.1 (7.0)	88.4 (10.0)	57.7 (21.5)	83.3 (6.8)	88.5 (10.6)	52.3 (21.0)	83.2 (5.8)
SOM[m]	85.6 (12)	63.4 (10.3)	72.7 (11.7)	87.6 (7.2)	57.1 (10.2)	81.6 (5.8)	93.5 (5.4)	64.4 (8.5)	84.8 (4.0)	94.7 (4.0)	59.0 (5.8)	85.0 (3.1)
K-means	94.2 (7.6)	57.2 (7.6)	*73.1 (7.1)*	93.7 (6.2)	62.2 (10.5)	*85.4 (4.2)*	96.0 (4.4)	67.6 (10.3)	*87.4 (3.1)*	95.8 (3.9)	62.1 (10.3)	*86.5 (2.9)*

[a]AUC: area under the receiver operating characteristic curve.

[b]SVDD: support vector data description.

[c]Italicized values indicates the top performing models.

[d]IncSVDD: incremental support vector data description.

[e]V-SVM: one-class support vector machine.

[f]NN: nearest neighbor.

[g]MST: minimum spanning tree.

[h]MOG: mixture of Gaussian.

[i]MCD: minimum covariance determinant.

[j]K-NN: K-nearest neighbor.

[k]LOF: local outlier factor.

[l]PCA: principal component analysis.

[m]SOM: self-organizing maps.

Daily Smoothed Data Set

Regarding the daily smoothed data set, as shown in Table 6, almost all models achieved excellent performance and much improved data description compared with the daily raw data set. As shown in Table 6, specific models such as V-SVM, K-NN, and K-means produced excellent descriptions of the data for all the sample sizes; however, V-SVM achieved superior performance compared with these models. In the group comparison, the boundary and domain-based method produced excellent description of the data for all sample sizes.

Table 6. Average performance of each model across all the infection cases for the daily smoothed data set (with filter) and different sample size. Fraction=0.01.

Models	1 month			2 months			3 months			4 months		
	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1
Boundary and domain-based method												
SVDD[b]	99.9 (0.7)	100 (0.0)	94.1 (14.2)	100 (0.0)	100 (0.0)	96.1 (7.6)	100 (0.0)	100 (0.0)	96.5 (6.5)	100 (0.0)	100 (0.0)	97.9 (3.9)
IncSVDD[c]	99.9 (0.7)	100 (0.0)	94.1 (14.2)	100 (0.0)	100 (0.0)	96.9 (6.5)	100 (0.0)	100 (0.0)	97.3 (5.9)	100 (0.0)	100 (0.0)	98.6 (2.9)
V-SVM[d]	100 (0.0)	100 (0.0)	*99.1 (3.2)* [e]	100 (0.0)	100 (0.0)	*99.1 (2.9)*	100 (0.0)	100 (0.0)	*99.4 (1.9)*	100 (0.0)	100 (0.0)	*99.5 (1.5)*
NN[f]	90.1 (14.5)	40.0 (30.5)	69.5 (13.2)	88.9 (9.9)	33.1 (22.6)	78.4 (6.8)	89.2 (7.9)	33.6 (14.6)	77.7 (5.3)	90.5 (6.8)	23.5 (18.6)	77.1 (5.7)
MST[g]	98.9 (3.6)	85 (6.1)	86.7 (9.4)	99.8 (0.7)	96.7 (3.4)	95.1 (6.2)	99.9 (0.2)	98.9 (4.1)	98.0 (3.5)	99.9 (0.5)	100 (0.0)	98.0 (5.4)
Density-based method												
Gaussian	99.2 (5.1)	92.6 (9.0)	87.2 (15.2)	99.5 (2.5)	96.7 (7.5)	94.8 (10.4)	99.9 (0.4)	100 (0.0)	98.1 (4.9)	99.8 (0.8)	100 (0.0)	98.3 (5.9)
MOG[h]	98.8 (5.4)	92.9 (8.6)	85.2 (17.1)	99.4 (2.6)	97.0 (5.4)	92.1 (11.6)	99.9 (0.4)	99.9 (0.7)	95.4 (7.8)	99.8 (1.0)	99.9 (0.6)	96.4 (7.7)
MCD[i] Gaussian	98.4 (5.6)	86.6 (8.8)	86.6 (11.9)	99.3 (2.7)	90.0 (8.7)	93.4 (8.1)	99.8 (0.5)	99.2 (2.6)	98.0 (5.3)	99.8 (0.9)	97.1 (3.9)	97.0 (5.5)
Parzen	99.2 (3.5)	100 (0.0)	90.8 (16.4)	99.9 (0.4)	100 (0.0)	93.7 (9.8)	100 (0.0)	100 (0.0)	93.6 (8.9)	99.9 (0.3)	100 (0.0)	95.8 (8.2)
Naïve Parzen	99.8 (1.2)	100 (0.0)	94.4 (14.6)	100 (0.0)	100 (0.0)	96.1 (7.9)	99.9 (0.5)	100 (0.0)	97.4 (5.6)	100 (0.0)	100 (0.0)	98.2 (4.2)
K-NN[j]	99.5 (2.0)	91.6 (3.6)	*90.7 (9.6)*	99.9 (0.4)	100 (0.0)	*98.3 (4.9)*	100 (0.0)	100 (0.0)	*98.4 (5.1)*	100 (0.0)	100 (0.0)	*98.8 (3.6)*
LOF[k]	99.6 (1.5)	93.3 (7.3)	92.4 (10.6)	99.9 (0.5)	99.2 (3.4)	97.1 (7.3)	99.9 (0.2)	98.6 (2.8)	97.4 (4.5)	99.9 (0.4)	100 (0.0)	98.2 (5.9)
Reconstruction-based method												
PCA[l]	93.8 (6.7)	82.0 (7.3)	83.8 (10.4)	91.3 (4.3)	77.9 (7.3)	89.3 (8.7)	88.7 (5.9)	76.3 (8.6)	89.5 (5.3)	90.7 (3.6)	76.2 (8.6)	89.0 (6.9)
Auto-encoder	97.0 (8.1)	91.6 (14.6)	87.7 (16.0)	98.1 (5.4)	92.6 (15.3)	92.0 (10.7)	98.6 (4.6)	92.8 (14.8)	94.0 (8.3)	98.7 (4.0)	92.7 (15.8)	94.9 (7.7)
SOM[m]	99.1 (3.2)	99.9 (0.6)	85.2 (20.5)	99.8 (0.7)	100 (0.0)	88.9 (16.1)	99.9 (0.2)	100 (0.0)	94.6 (8.0)	99.8 (0.6)	100 (0.0)	95.9 (8.1)
K-means	99.8 (1.2)	96.2 (6.0)	*93.2 (12.7)*	100 (0.0)	100 (0.0)	*97.8 (5.6)*	100 (0.0)	100 (0.0)	*98.0 (5.6)*	100 (0.0)	100 (0.0)	*99.0 (2.9)*

[a]AUC: area under the receiver operating characteristic curve.

[b]SVDD: support vector data description.

[c]IncSVDD: incremental support vector data description.

[d]V-SVM: one-class support vector machine.

[e]Italicized values indicates the top performing models.

[f]NN: nearest neighbor.

[g]MST: minimum spanning tree.

[h]MOG: mixture of Gaussian.

[i]MCD: minimum covariance determinant.

[j]K-NN: K-nearest neighbor.

[k]LOF: local outlier factor.

[l]PCA: principal component analysis.

[m]SOM: self-organizing maps.

Hourly Smoothed Data Set

Regarding the hourly smoothed data set, as shown in Table 7, almost all the models failed to produce acceptable data description from the 1-month sample size except V-SVM, which achieved the best description. The high variability between the performance of the models with the 1-month hourly data set could be associated with the high data granularity, and, in fact, the models require more data sets to capture the high variability among the data objects. Models such as V-SVM, MCD Gaussian, and K-means achieved superior performance from their respective groups. In general, V-SVM outperformed in all the sample sizes. The density and reconstruction-based models improved with larger sample size. In the group comparison, the boundary and domain-based method produced better description in all the sample sizes, and the density and reconstruction-based method achieved equivalent performance with larger sample sizes.

Table 7. Average performance of each model across all the infection cases for the hourly data set with smoothing and different sample size. Fraction=0.01.

Models	1 month			2 months			3 months			4 months		
	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1	AUC[a], mean (SD)	Specificity	F1
Boundary and domain-based method												
SVDD[b]	97.4 (2.9)	89.0 (3.4)	89.4 (7.1)	97.4 (1.8)	86.7 (4.4)	91.5 (10.9)	97.2 (2.6)	80.1 (5.5)	93.5 (3.4)	97.6 (1.7)	81.8 (5.3)	94.6 (6.0)
IncSVDD[c]	97.1 (2.9)	87.7 (2.7)	89.5 (5.9)	97.2 (1.8)	86.4 (2.8)	93.6 (4.8)	97.0 (2.7)	76.2 (6.3)	93.2 (2.6)	97.4 (1.7)	79.0 (4.8)	*95.4 (1.9)* [d]
V-SVM[e]	98.1 (2.0)	85.5 (0.6)	*92.3 (1.3)*	98.9 (1.4)	89.8 (0.2)	*95.4 (1.6)*	98.7 (1.4)	86.4 (0.4)	*94.4 (2.0)*	99.0 (0.9)	89.2 (0.3)	*95.4 (2.1)*
NN[f]	93.2 (7.8)	92.0 (2.4)	83.9 (12.0)	94.4 (2.5)	88.4 (3.4)	90.9 (5.3)	93.3 (2.8)	83.0 (3.7)	92.0 (4.2)	94.0 (2.8)	82.9 (3.6)	94.0 (4.0)
MST[g]	96.1 (2.6)	94.4 (2.2)	72.9 (18.5)	97.3 (1.4)	94.2 (2.1)	86.1 (11.0)	96.1 (2.1)	93.5 (1.9)	90.2 (7.3)	97.0 (1.4)	93.6 (1.7)	92.6 (5.0)
Density-based method												
Gaussian	98.4 (1.6)	91.2 (2.6)	89.6 (12.5)	99.3 (0.9)	92.3 (1.7)	95.7 (4.9)	98.8 (1.3)	88.1 (4.0)	95.9 (2.7)	99.2 (0.7)	89.8 (3.1)	*97.2 (1.8)*
MOG[h]	97.5 (3.0)	91.7 (3.2)	87.8 (13.3)	98.9 (1.2)	90.9 (2.7)	94.0 (6.3)	98.2 (2.0)	85.4 (6.6)	94.2 (4.1)	98.5 (1.5)	88.0 (4.9)	96.0 (3.1)
MCD[i] Gaussian	98.5 (1.5)	89.9 (3.7)	*89.1 (11.8)*	99.5 (0.9)	92.2 (92.2)	*95.8 (4.5)*	98.9 (1.1)	87.9 (3.3)	*96.0 (2.5)*	99.2 (0.7)	90.4 (3.4)	*97.4 (1.7)*
Parzen	96.4 (2.6)	97.8 (1.1)	59.9 (18.9)	98.0 (1.6)	97.7 (1.1)	79.5 (14.5)	97.2 (2.3)	96.4 (1.2)	85.1 (10)	98.1 (1.6)	96.7 (1.1)	88.6 (7.1)
Naïve Parzen	96.4 (3.0)	87.5 (3.5)	85.1 (10.9)	98.7 (1.5)	89.2 (2.8)	92.8 (7.5)	96.0 (2.3)	90.8 (2.6)	95.0 (4.1)	98.2 (1.6)	90.0 (1.8)	96.2 (2.8)
K-NN[j]	97.6 (2.9)	91.1 (1.6)	87.6 (13.6)	99.0 (1.4)	92.4 (2.4)	94.5 (6.6)	98.4 (1.4)	92.6 (1.4)	95.7 (4.8)	98.7 (1.1)	93.3 (1.3)	*97.3 (2.8)*
LOF[k]	96.9 (2.9)	91.2 (1.6)	86.2 (13.0)	97.4 (1.8)	89.8 (4.8)	93.1 (4.9)	95.0 (3.0)	85.2 (4.6)	92.9 (4.8)	95.8 (1.7)	85.3 (4.7)	94.7 (3.2)
Reconstruction-based method												
PCA[l]	97.4 (3.2)	78.2 (6.1)	82.5 (10.9)	94.8 (3.8)	77.6 (4.5)	90.9 (3.6)	92.6 (4.2)	72.4 (3.8)	92.5 (1.9)	93.4 (3.2)	71.1 (2.5)	93.9 (1.1)
Auto-encoder	95.4 (5.3)	88.7 (9.5)	86.1 (13.1)	96.9 (3.2)	87.1 (9.9)	92.8 (6.4)	95.0 (5.3)	79.3 (14.5)	93.1 (4.8)	95.9 (4.3)	80.3 (14.4)	95.0 (3.6)
SOM[m]	95.9 (2.9)	91.6 (2.6)	86.1 (14.4)	95.7 (1.7)	87.6 (4.1)	92.7 (5.7)	93.9 (3.5)	79.1 (10.9)	92.3 (4.5)	96.0 (2.5)	87.5 (7.0)	96.1 (3.2)
K-means	97.1 (3.9)	89.7 (6.7)	*88.7 (12.1)*	98.6 (1.7)	91.1 (4.2)	*95.2 (4.4)*	98.5 (1.5)	92.3 (2.9)	*96.9 (3.3)*	98.9 (1.0)	93.9 (1.3)	*97.9 (2.2)*

[a]AUC: area under the receiver operating characteristic curve.

[b]SVDD: support vector data description.

[c]IncSVDD: incremental support vector data description.

[d]Italicized values indicates the top performing models.

[e]V-SVM: one-class support vector machine.

[f]NN: nearest neighbor.

[g]MST: minimum spanning tree.

[h]MOG: mixture of Gaussian.

[i]MCD: minimum covariance determinant.

[j]K-NN: K-nearest neighbor.

[k]LOF: local outlier factor.

[l]PCA: principal component analysis.

[m]SOM: self-organizing maps.

Unsupervised Methods

Two density-based unsupervised models were tested and evaluated on the same set of data as used in the one-class classifiers: LOF and COF. The average AUC, specificity, and F1-score were computed after 20 runs. The best performing thresholds for all the infection cases along with the optimal value of *k* (number of neighbors) are given in Table 8. As can be seen from the table, both the LOF and the COF achieved better performance on the smoothed data set as compared with its raw version. In all the infection cases, LOF performed better than COF. This is mainly because of the characteristics of the data sets, which fulfill the LOF spherical assumption of neighbor distribution. Considering the average F1-score across all the infection cases, LOF achieved 74.7% on the raw daily data, 91.1% on the smoothed daily data, and 72.7% on the hourly data, whereas COF achieved 71.9% on the raw daily data, 85.8% on the smoothed daily data, and 68.9% on the hourly data. However, compared with the one-class classifier, it suffers from performance degradation mainly because the data are not distributed uniformly, where some regions may contain high density and others might be sparse. However, the region of sparse density does not always signify anomalies (infection incidence). For example, an individual patient on certain days might prefer to take little insulin compared with most of the days and perform heavy physical activity to replace their insulin needs. This scenario could generate an outlier, a small ratio of insulin-to-carbohydrate, which will be considered and detected as outliers by unsupervised models. A detailed score plot of each model for the different infection cases can be found in Multimedia Appendix 3.

Table 8. Average area under the receiver operating characteristic curve, specificity, and F1-score for both with and without smoothed versions of the data. The parameters kd and kh represent the optimal number of nearest neighbors for the daily and hourly cases, respectively.

Frequencies, density-based methods												
Pre-pro	Models (threshold)	1st case of infection (k_d=30, k_h=240)			2nd case of infection (k_d=30, k_h=240)			3rd case of infection (k_d=30, k_h=240)			4th case of infection (k_d=30, k_h=240)		
		AUC[a]	Specific	F1	AUC[a]	Specific	F1	AUC[a]	Specific	F1	AUC[a]	Specific	F1
Daily													
Without filter	LOF[b] (T_1=2.4, T_2=1.2, T_3=1.45, T_4=1.8)[c]	75.0	50.0	*85.6*	90.0	100	67.4	92.1	66.7	*70.1*	98.2	100	*75.8*
	COF[d] (T_1=1.4, T_2=1.3, T_3=1.4, T_4=1.4)	82.1	66.7	72.6	97.4	100	*75.8*	75.2	66.7	67.6	96.7	100	71.8
With filter	LOF[b] (T_1=1.7, T_2=1.6, T_3=1.95, T_4=2.2)	99.0	100	*84.1*	99.2	100	*85.4*	100	100	*100*	99.9	100	94.7
	COF[d] (T_1=1.3, T_2=1.3, T_3=1.8, T_4=1.8)	97.6	100	76.6	97.9	100	77.6	99.5	100	88.8	100	100	*100*
Hourly													
	LOF[b] (T_1=1.4, T_2=1.3, T_3=1.35, T_4=1.5)	98.0	86.0	*74.6*	95.5	100	*70.2*	94.3	91.4	*75.0*	85.2	72.6	*71.1*
	COF[d] (T_1=1.2, T_2=1.1, T_3=, T_4=1.1)	92.4	88.4	*74.6*	77.0	66.0	62.5	90.3	82.7	74.6	82.6	82.2	63.7

[a]AUC: area under the receiver operating characteristic curve.

[b]LOF: local outlier factor.

[c]T_k: threshold for the kth month.

[d]COF: connectivity-based outlier factor.

Computational Time

Computational time is the amount of time a particular model needs to learn and execute a given task [12]. It can be regarded as one of the best performance indicators for real-time systems. For a real-time application, an optimal model is the one that achieves superior detection performance with small training and testing time**.** Depending on the application, sometimes models can be trained offline, which makes the training time less important [12]. In this regard, the computational times of all the models were estimated and compared with each other. The computational time was measured for different sample sizes of the training and testing data sets. The sample size of the training and testing data includes 240, 480, 720, 960, 1200, 1440, 1680, 1920, 2160, 2400, 2640, and 2880 sample objects (data points) each. The required computational time for both training and testing each model is depicted in Figures 5 and 6. The figures demonstrate a rough estimation of the computational time, where each model learns the data set and classifies the sample objects. During the training phase, NN, SVDD, and SOM took considerable time. For a training sample size of 2880 objects, NN requires 296 times, SVDD requires 206 times, and SOM requires 42 times the time taken by K-NN on the same sample size. Generally, as the number of sample objects

increases, these models require much more time. However, K-means, Gaussian families, LOF, MST, K-NN, V-SVM, PCA, auto-encoder, and incSVDD took less time. These models took almost constant time even when the number of samples increased. During the testing phase, only the LOF took considerable time compared with the other models, as can be seen in Figure 6.

Figure 5. Plot of models' average computational time for the training phase. The x-axis depicts the sample size, and each label stands for total sample size divided by 24. The y-axis depicts the computational time required by each model. Gauss: Gaussian; IncSVDD: incremental support vector data description; K-NN: K-nearest neighbor; LOF: local outlier factor; MCD: minimum covariance determinant; MOG: mixture of Gaussian; MST: minimum spanning tree; NN: nearest neighbor; NParzen: naïve Parzen; PCA: principal component analysis; SOM: self-organizing maps; SVDD: support vector data description; V-SVM: one-class support vector machine.

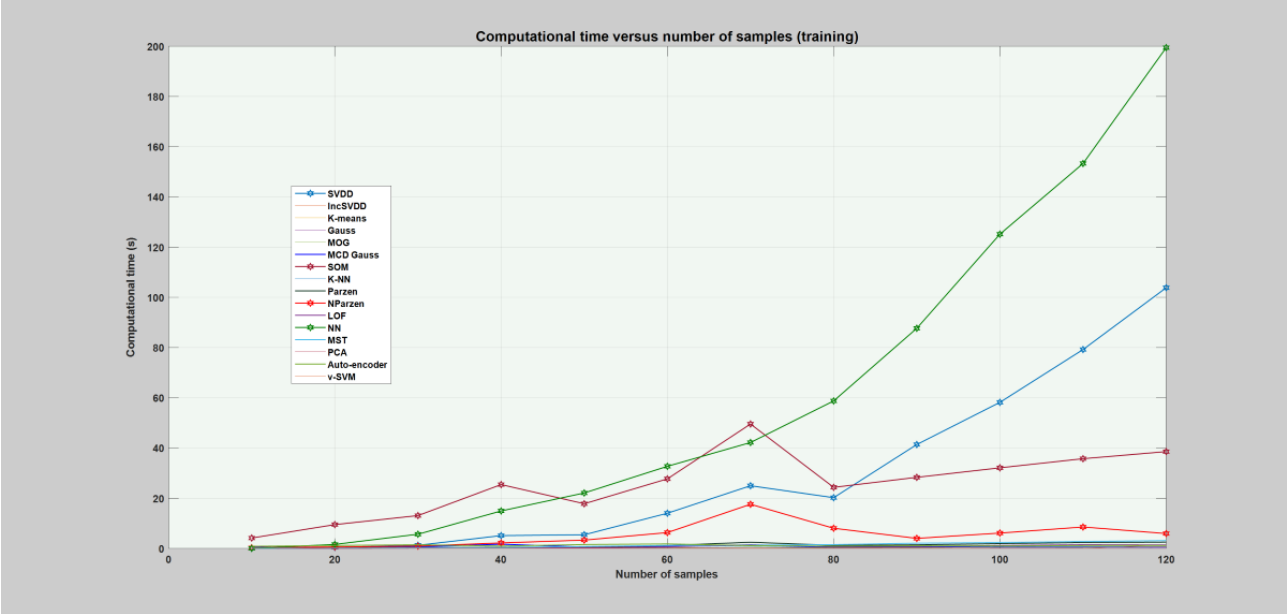
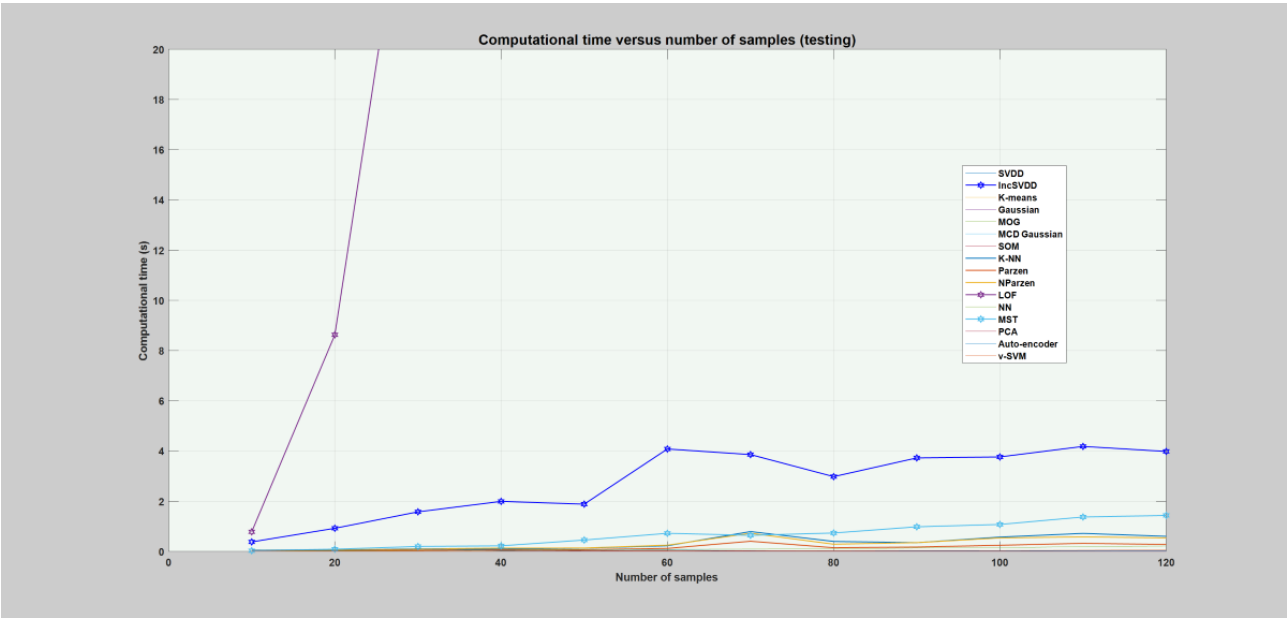


Figure 6. Plot of models' average computational time for the testing phase. The x-axis depicts the sample size, and each label stands for total sample size divided by 24. The y-axis depicts the computational time required by each model. Gauss: Gaussian; IncSVDD: incremental support vector data description; K-NN: K-nearest neighbor; LOF: local outlier factor; MCD Gauss: Gaussian: SOM: self-organizing maps; MOG: mixture of Gaussian; MST: minimum spanning tree; NN: nearest neighbor; NParzen: naïve Parzen; PCA: principal component analysis; SVDD: support vector data description; V-SVM: one-class support vector machine.



Discussion

Principal Findings

Anomaly or novelty detection problem has been widely used in various applications including machine fault and sensor failure detection, prevention of credit card or identity fraud, health and medical diagnostics and monitoring, cyber-intrusion detection, and others [1-3]. In applications related to health and medical diagnostics and monitoring, the anomaly detection problem has been used to detect and identify the abnormal health

state of an individual, for example, detecting abnormal patterns of heartbeat recorded using an electrocardiogram [1,51-54]. The omnipresence of various physiological sensors has facilitated circumstances for individuals to easily self-record health-related events and data for the purpose of self-informatics and management [55]. Currently, people are generating huge amounts of data on a daily basis that can contribute to both individual and public health purposes [54]. To this end, people with diabetes are not an exception, generating rich data in both quality and quantity, which is expected to further improve with advances in diabetes technologies. These data can provide valuable information if processed with the right tools and methodology, and in this regard, particular instance includes detecting novel or anomalous data points for various purposes. The availability of labeled data constrains the choice of methods in the anomaly detection problem [3,9-11]. Supervised anomaly detection methods are impractical for applications such as detecting infection incidences in people with type 1 diabetes for a number of reasons [10,12]. Blood glucose dynamics are affected by various other factors apart from infection incidences [19,56,57], and characterization of infection-induced anomalies (abnormal class) from the normal class [13] is a challenging task because of the following reasons:

1. There are no well-defined boundaries regarding how different pathogens affect various key parameters of blood glucose dynamics, including blood glucose levels, insulin injections, carbohydrate ingestions, physical activity or exercise load, and others. This results in poor boundary demarcation between the normal and abnormal classes.
2. Class boundaries defined for a single pathogen might not work for the other pathogens because the effect of different pathogens on the blood glucose dynamics could be different.
3. It is expensive and time consuming to collect infection-related data to explore and characterize pathogen-specific class boundaries. This results in ill-defined class boundaries even for an infection related to a single pathogen.
4. The degree of effect of the same pathogens on the blood glucose dynamics could differ between different individuals because of the difference in individual immunity, which further complicates the characterization task.
5. Lack of sufficient sample size for both the abnormal and the normal classes results in poor training and testing data sample size or imbalanced class problems.

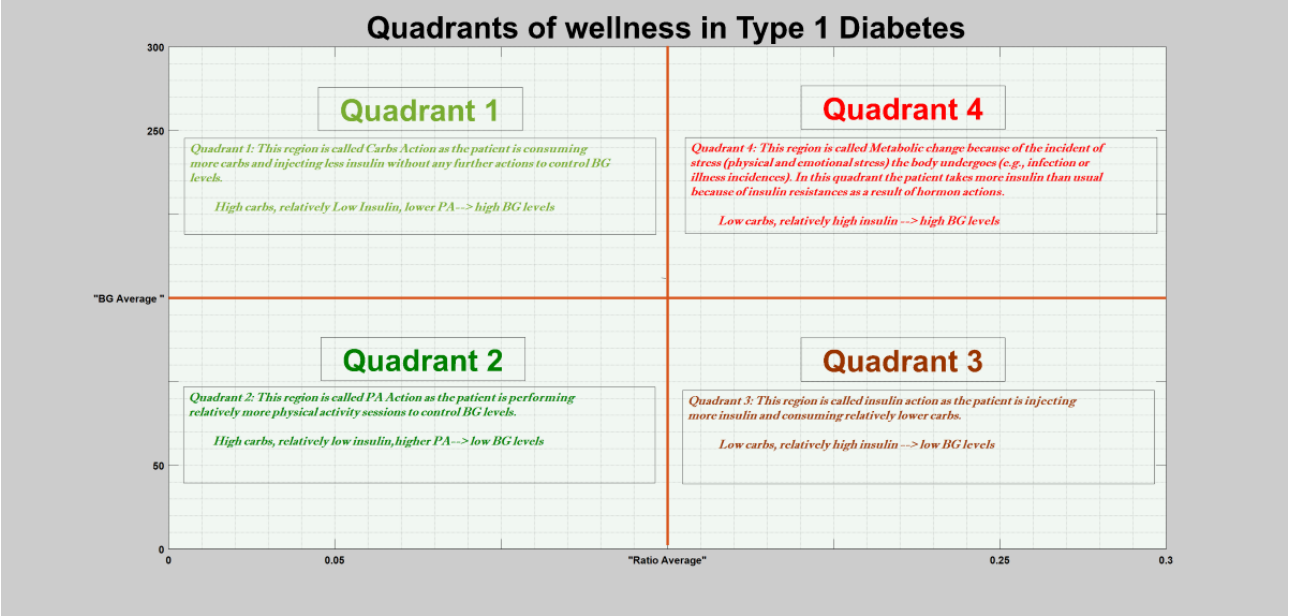
Given these challenges, the best possible approach is to identify methods that can learn from the normal health state of an individual and classify abnormalities relying on the boundaries learnt from the normal health state, which is a one-class classifier approach. This definitely reduces the challenge because it only requires the characterization of what is believed to be a normal health state. For instance, assume a health diagnostic and monitoring system that detects health changes in an individual by tracking the individual's physiological parameters, where the current health status is examined based on set of parameters, and raises a notification alarm when the individual health deteriorates [12]. In such a system, it becomes feasible to rely on a method that can be trained using only the regular or normal day measurements (target days) so as to detect deviation from normality [12,14]. Another possible alternative approach is to identify a method that does not require any characterization and labeling of classes, which is unsupervised methods [7]. Accordingly, considering the previously mentioned challenges, one-class classifiers and unsupervised models were proposed for detecting infection incidence in people with type 1 diabetes. The objective was to develop a personalized health model that can automatically detect the incidence of infection in people with type 1 diabetes using blood glucose levels and insulin-to-carbohydrate ratio as input variables. The model is expected to detect any deviations from the norm as a result of infection incidences considering blood glucose level (hyperglycemia incidences) coupled with unusual changes in the insulin-to-carbohydrate ratio, that is, frequent insulin injections and unusual reduction in the amount of carbohydrate intake [19]. A personalized health model based on one-class classifiers and unsupervised methods was tested using blood glucose levels and the insulin-to-carbohydrate ratio as a bivariate input. The result demonstrated the potential of the proposed approach, which achieved excellent performance in describing the data set, that is, detecting infection days from the regular or normal days, and, in particular, the boundary and domain-based method performed better. Among the respective group, particular models such as V-SVM, K-NN, and K-means achieved excellent performance in all the sample sizes and infection cases. However, the unsupervised approaches suffer performance degradation compared with the one-class classifier mainly because of the atypical nature of the data, which are not distributed uniformly, where some regions may contain high density and others might be sparse (Multimedia Appendix 2). There are rare events (sparse region) of blood glucose dynamics that are a normal response; however, the unsupervised methods can still detect and flag false alarms including the following:

1. Carbohydrate action: a situation in which the ratio of insulin-to-carbohydrate is small and the blood glucose levels are high (hyperglycemia), *Carb Action-Quadrant 1* in Figure 7. This is a normal response to blood glucose dynamics as consumption of more carbohydrates and less insulin intake can derive blood glucose dynamics into the hyperglycemia region (high blood glucose levels) if there is no physical activity session. A typical example of this particular situation is holiday seasons, where people consume too many carbohydrates.
2. Physical activity action: despite a small ratio of insulin-to-carbohydrate, the blood glucose levels still drop to low levels (hypoglycemia), *PA Action-Quadrant 2* in Figure 7. Normally, a small ratio of insulin-to-carbohydrate signifies that the patient consumed more carbohydrates and injected less insulin, which normally derives the blood glucose dynamics into the hyperglycemia region. However, despite taking more carbohydrates and less insulin, a rigorous physical exercise can still derive the blood glucose dynamics into the hypoglycemia region. Therefore, this is a normal response of blood glucose dynamics as the action of physical activity or exercise can derive the patient into hypoglycemic regions even if the patient consumes more carbohydrates. For example, an individual patient on certain days might prefer to take little insulin as compared with most of the days and perform heavy physical activity to

replace their insulin needs. This scenario could generate an outlier, a small ratio of insulin-to-carbohydrate, which will be considered and detected as anomalies by the unsupervised models. However, this could be mitigated by incorporating physical activity data as an input variable.

3. Insulin action: the ratio of insulin-to-carbohydrate is large, that is, high insulin intake and low carbohydrate consumption, and blood glucose levels are low (hypoglycemia), *Insulin Action-Quadrant 3* in Figure 7. This is a normal response to blood glucose dynamics as administration of high insulin with little carbohydrate consumption can derive the blood glucose dynamics into the hypoglycemic region.

Figure 7. Quadrants of wellness in people with type 1 diabetes. The figure depicts the 4 possible scenarios of different parameters: carbohydrate action, insulin action, physical activity action, and abnormality because of metabolic change such as infection and stress. BG: blood glucose; PA: physical activity.



The drawback of unsupervised methods is that they do not have any mechanism to handle rare events even if the events are normal. This is mainly because unsupervised methods define an anomaly on the basis of the entire data set. However, the one-class classifier can learn and handle such scenarios appropriately if presented during the training phase. This is mainly because one-class classifiers produce a reference description based on the available normal (target) data set, including the rare events. With regard to the one-class classifiers, the boundary and domain-based method achieved a better description of the data set compared with the density and reconstruction-based methods, mainly because of the ability of such models to handle the atypical nature of the data [12]. Detectability of the infection incidence is directly related to the extent and degree of the effect it induces on the blood glucose dynamics. The type of pathogen, individual's immunity, and hormones involved could play a role in determining the degree of severity in this regard [19,24,58-62]. To this end, the results demonstrated that the models were capable of detecting all the infection incidences that can significantly alter the blood glucose dynamics, such as influenza. Moreover, infection incidence that had a moderate effect on the blood glucose dynamics, such as mild common cold without fever, was also detected. However, as expected, infection incidences that had almost little effect on the blood glucose dynamics, such as light common cold without fever, as reported by the individual patient, were not detected. Regarding the computational time, NN, SVDD, and SOM took considerable training time, which typically increased as the number of sample objects increased. Moreover, compared with the other models, only LOF and COF took considerable testing time.

Comparative Analysis of the Methods

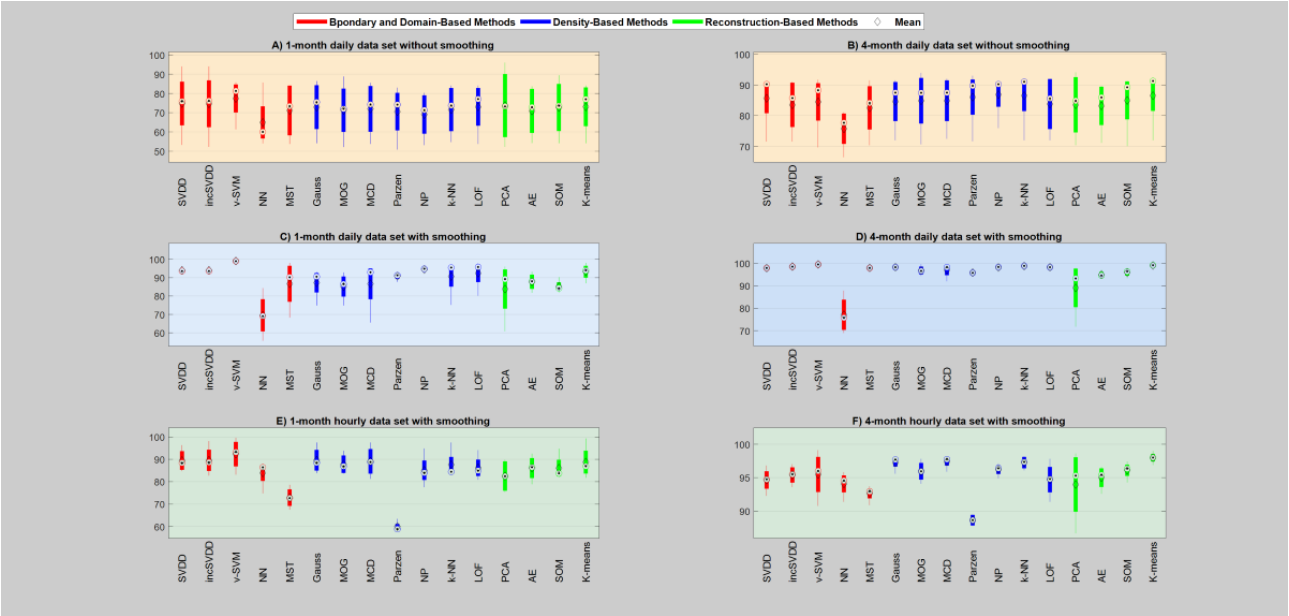
Selecting the proper model for implementation in a real-world setting requires considering different characteristics of the model. This includes typical model characteristics such as performance in limited training sample size, robustness to outliers in the training data, required training and testing time, and complexity of the model (in terms of the number of model parameters).

Performance and Sample Size

The sample size, N, is the number of sample objects used during the training phase and highly affects the generalization power of the model [12,13]. Models trained with small sample sizes often fail to produce satisfactory descriptions mainly associated with the presence of large variance in the sample objects [3,12,13,63]. To this end, the results indicate that most of the models fail to make good descriptions with a 1-month (30 objects) data set, mainly with the daily raw data set, as shown in Figure 8. The figure depicts the average performance of each model across all the infection cases over the 1- and 4-month sample sizes. Specifically, MST, Gaussian families, SOM, and auto-encoders require a considerable amount of training sample objects to better describe the data. There is some exception, for instance V-SVM, which produces a satisfactory description of the 1-month data sets in all the infection cases and data granularity. Models such as NN and PCA produced the worst

description in most cases. As the number of training sample objects increased, all the models improved and produced a comparable description of the data. As a rule of thumb, for the daily scenario, a 3-month training sample (90 sample objects) produces a good description of the data, which can be considered for real-world applications. Moreover, if smoothing is considered, a 1-month sample size produces better description than the 4-month sample size without smoothing, as shown in Figure 8. However, for the hourly scenario, a 1-month training sample object produces a comparable description and anything more than this size will be enough.

Figure 8. Average performance (F1-score) of each model across all the infection cases. AE: auto-encoder; Gauss: Gaussian; IncSVDD: incremental support vector data description; K-NN: K-nearest neighbor; LOF: local outlier factor; MCD: minimum covariance determinant; MOG: mixture of Gaussian; MST: minimum spanning tree; NN: nearest neighbor; NP: naïve Parzen; PCA: principal component analysis; SOM: self-organizing maps; SVDD: support vector data description; V-SVM: one-class support vector machine.



Computational Time

For real-time applications, the time a model takes to learn and classify the sample object is essential in model selection. Table 9 depicts the rough estimation of average training and testing time required by different classifiers, both the one-class classifiers and the unsupervised models, based on 2880 training and testing sample objects each. Most of the models, as shown in Figures 5 and 6 and Table 9, require reasonable training and testing time, except NN, SVDD, and SOM, which took a considerably longer time. However, it is possible that in some cases models can be trained offline, which makes the training time less important. With regard to the testing time, most of the models executed the classification task in a reasonable time except COF and one class classifier version of LOF, which consume considerable time to classify the 2880 objects. The computational time in these particular models grows exponentially as the sample size increases, which makes them resource demanding in a big data setting.

Table 9. Rough estimation of average training and testing time required by the different classifiers.

Methods	Training time, mean (SD)	Testing time, mean (SD)
One-class classifiers		
SVDD[a]	105.2 (2.03)	0.008 (0.002)
IncSVDD[b]	0.05 (0.16)	2.41 (0.83)
K-means	0.0047 (0.0014)	0.0032 (0.0010)
Gaussian	0.0055 (0.0032)	0.0032 (0.0012)
MOG[c]	0.076 (0.018)	0.0036 (0.0011)
MCD[d] Gaussian	0.27 (0.075)	0.0034 (0.0015)
SOM[e]	21.62 (5.91)	0.0033 (0.00087)
K-NN[f]	0.51 (0.11)	0.52 (0.12)
Parzen	2.02 (0.41)	0.21 (0.052)
Naïve Parzen	4.02 (0.82)	0.40 (0.10)
LOF[g]	1.15 (0.28)	1198.05 (323.07)
NN[h]	151.34 (22.52)	0.18 (0.024)
MST[i]	2.39 (0.31)	1.24 (0.19)
PCA[j]	0.046 (0.20)	0.0031 (0.00086)
Auto-encoder	0.65 (0.094)	0.017 (0.0034)
V-SVM[k]	0.32 (0.024)	0.035 (0.0066)
Unsupervised		
LOF[l]	N/A[m]	0.2 (0.0)
COF[n]	N/A	82.8 (1.5)

[a]SVDD: support vector data description.
[b]IncSVDD: incremental support vector data description.
[c]MOG: mixture of Gaussian.
[d]MCD: minimum covariance determinant.
[e]SOM: self-organizing maps.
[f]K-NN: K-nearest neighbor.
[g]LOF: local outlier factor.
[h]NN: nearest neighbor.
[i]MST: minimum spanning tree.
[j]PCA: principal component analysis.
[k]V-SVM: one-class support vector machine.
[l]LOF: local outlier factor.
[m]N/A: not applicable.
[n]COF: connectivity-based outlier factor.

Robustness to Outliers in the Training Data Set

The presence of outliers in the training data set could significantly affect the model's generalization ability. Outlier objects are samples that exhibit different characteristics compared with the rest of the objects in the data set [8,63]. For instance, an individual might forget a previous infection incident and could label these days as a regular or normal period during self-reporting, which could end up being used as target data sets for training. Another important example could be error recorded during data registration, that is, carbohydrate, blood glucose levels, and insulin registration. Such errors could occur during the manual registration of carbohydrates, associated with infusion set failures and other similar situations. In this scenario, an individual could record lower or higher values incorrectly affecting the input features, for example, ratio of insulin-to-carbohydrate and blood glucose levels, resulting in an outlier that could greatly affect the model's generalization ability. In this type of situation, a model's sensitivity to outliers in the training data is crucial to curb the influence of outliers on the accuracy of the description generated. To some extent,

a user-specified empirical rejection rate is incorporated in the models to reduce the effect of outliers in the training data by rejecting the most dissimilar objects from the description generated. For example, a rejection rate of 1% on training data sets implies that 1% of outliers in the training data set are rejected. Nevertheless, the sensitivity of models to outliers in the training data sets differs greatly between models. Among the models, NN is regarded as the most sensitive model to outliers in the training data set [12]. The presence of outliers in the training data changes the shape of the description generated by the model, forcing a larger portion of the feature space to be accepted as the target class [10,12]. Furthermore, models that rely on an estimation of the covariance matrix, for example, Gaussian families, also suffer from the presence of outliers in the training data sets [12,36]. However, when equipped with regularization, Gaussian models can withstand such outliers. Local density estimators such as Parzen can withstand outliers, considering the fact that only the local density is affected [12]. Models that rely on prototype estimation, such as SOM and K-means, are highly affected by the presence of outliers in the training data set, which could force the estimated prototype to be placed near or at the nontarget data set [2,12,13]. Nevertheless, boundary and domain-based method such as SVDD and V-SVM and reconstruction-based method such as auto-encoders are more or less insensitive to outliers and can generate acceptable solutions [3,12,64].

Model Parameters and Associated Complexity

The parameters of a model can be either free or user defined. These two parameters, free and user defined, provide insight into how flexible the model is, how sensitive the model is to overtraining, and how easy the model is to configure (simplicity) [12,16]. Considering the number of these parameters, there exist large variations among the models. For instance, NN does not possess any free parameters; therefore, its performance completely relies on the training data set [12]. This constraint has limitations, mainly because training data that contain outliers could ruin the model's performance [12,15,16]. A model that possess large number of free and user defined parameters is too flexible and complex [12]**.** Regarding the user-defined parameters, also known as hyper-parameters, a model equipped with small number of parameters and preferably with intuitive meaning are easy to configure. Setting up the user defined parameters incorrectly can degrade the model's performance and selecting the proper values (optimization) becomes complex and vague as the number of model parameters become too large. One of the simplest models is Parzen density and NN, which do not require the user to specify any parameters [3,12,13]. Some models, such as support vector families, require the user to specify parameters that have intuitive meaning, for example, the ratio of training objects to be rejected by the description [12,65]. There are also models that are complex enough given that the user is expected to specify many parameters, which are not intuitive and require careful choice. Examples of such models include SOM and auto-encoders, where the user is expected to supply the number of neuron, hidden units, and learning rate [10,12,37,66].

Practical Illustration and Area of Applications

For a real-world application, apart from the performance of the model, it is important to consider two important aspects of the data set, the time window of detection (data granularity) and the required sample size. The time window or data granularity, that is, hourly and daily, defines the frequency (continuity) of computation one needs to conduct throughout the day to screen the health status of the individual with type 1 diabetes. In an hourly time window, one is expected to carry out the computation at the end of each hour throughout the day. However, in the daily time window, one needs to carry out one aggregate computation at the end of the day. Decreasing the time window (increasing the granularity of the data) enhances early detections; however, at the coast of accuracy, for example, more unwanted features (noise) in the data. The results demonstrated that almost all the models produced fairly comparable detection performances in both time windows. Moreover, the required sample size determines the necessary amount of data an individual with type 1 diabetes needs to collect in advance before joining such an infection detection system. Models that could generalize well with small sample sizes could be preferred in a real-world application to enable more people to join the system with ease. Generally, the results demonstrated that the models require at least a sample size of 3-month data for the daily case and 2-month data for hourly case to perform better. Automating the detection of infection incidences among people with type 1 diabetes can deliver a means to provide personalized decision support and learning platforms for the individuals and, at the same time, can be used to detect infectious disease outbreaks on a large scale through spatio-temporal cluster detection [19,67,68]. Detailed descriptions of these instances are given below:

1. A personalized decision support system and learning platform relies on an individual's self-recorded data to provide relevant information in relation to decision making to assist the individuals during crises [19,67,68]. Moreover, it can also provide a learning platform concerning the extent to which infection incidence affects the key parameters of the blood glucose dynamics. Information regarding what to expect at each stage of the course of infection could be very important to the individuals [19]. During infection incidences, various kinds of information could be vital for an individual to properly manage blood glucose levels, including time in range (blood glucose), to what extent is the evolution of blood glucose affected during the course of infection, to what extent does insulin sensitivity change, and how much does the insulin-to-carbohydrate ratio shift, that is, changes in insulin requirements for each gram of carbohydrate intake.
2. A population-based early outbreak detection system relies on self-recorded information from an individual with type 1 diabetes to detect individuals' infection cases and, thereby, detect a group of infected individuals on a spatio-temporal basis. Such a system should collect individuals' self-recoded data to a central server, analyze individuals' data on a timely basis, identify and locate a cluster of people based on space and time, and notify the responsible bodies if there is an ongoing outbreak [19,67-71].

Conclusions

Anomaly or novelty detection problem has been widely used in various applications including machine fault and sensor failure detection, prevention of credit card or identity fraud, health and medical diagnostics and monitoring, cyber-intrusion detection, and others. In this study, we demonstrated the applicability of one-class classifiers and unsupervised anomaly detection methods for the purpose of detecting infection incidences in people with type 1 diabetes. In general, the proposed methods produced excellent performance in describing the data set, and particularly the boundary and domain-based method performed better. In contrast to the specific models, V-SVM, K-NN, and K-means achieved better generalization in describing the data set in all infection cases. Detecting the incidence of infection in people with type 1 diabetes can provide an opportunity to devise tailored services, that is, personalized decision support and a learning platform for the individuals, and can simultaneously be used for detecting potential public health threats, that is, infectious disease outbreaks, on a large-scale basis through a spatio-temporal cluster detection. Generally, we foresee that the results presented could encourage researchers to further examine the presented features along with other additional features of self-recorded data, for example, various CGM features and physical activity data, on a large-scale basis.

Acknowledgments

The work presented in this paper is part of the project *Electronic Disease Surveillance Monitoring Network (EDMON) system*, which is funded by the University of Tromsø–The Arctic University of Norway and National Library of Medicine *Mechanistic machine learning* (grant number: LM012734) and is the PhD program of the first author, AW. The authors would like to extend their sincere gratitude to all the participants of the study.

Authors' Contributions

The first author, AW, conceived the study, designed and performed the experiments, and wrote the manuscript. IK, EÅ, JI, DA, and GH provided successive inputs and revised the manuscript. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Theoretical background of the methods.
[DOCX File , 73 KB - jmir_v22i8e18912_app1.docx]

Multimedia Appendix 2
Detailed description of the models input features.
[DOCX File , 12076 KB - jmir_v22i8e18912_app2.docx]

Multimedia Appendix 3
Score plot of the models for each patient year.
[DOCX File , 12308 KB - jmir_v22i8e18912_app3.docx]

Multimedia Appendix 4
Model evaluations – performance of the models for each patient year.
[DOCX File , 62 KB - jmir_v22i8e18912_app4.docx]

References

1. Dunning T, Friedman E. In: Loukides M, editor. Practical Machine Learning: A New Look at Anomaly Detection. New York, USA: O'Reilly Media Inc; 2014.
2. Agrawal S, Agrawal J. Survey on anomaly detection using data mining techniques. Procedia Comput Sci 2015;60:708-713 [FREE Full text] [doi: 10.1016/j.procs.2015.08.220]
3. Pimentel MA, Clifton DA, Clifton L, Tarassenko L. A review of novelty detection. Sig Process 2014 Jun;99:215-249 [FREE Full text] [doi: 10.1016/j.sigpro.2013.12.026]
4. Cohen G, Hilario M, Sax H, Hugonnet S, Pellegrini C, Geissbuhler A. An application of one-class support vector machine to nosocomial infection detection. Stud Health Technol Inform 2004;107(Pt 1):716-720. [Medline: 15360906]
5. Cohen G, Sax H, Geissbuhler A. Novelty detection using one-class Parzen density estimator. An application to surveillance of nosocomial infections. Stud Health Technol Inform 2008;136:21-26. [Medline: 18487702]
6. Cohen G, Hilario M, Sax H, Hugonnet S, Geissbuhler A. Learning from imbalanced data in surveillance of nosocomial infection. Artif Intell Med 2006 May;37(1):7-18. [doi: 10.1016/j.artmed.2005.03.002] [Medline: 16233974]

7. Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Comput Surv 2009 Jul;41(3):1-58 [FREE Full text] [doi: 10.1145/1541880.1541882]
8. Hawkins DM. Identification of Outliers. Netherlands: Springer; 1980.
9. Mehrotra KG, Mohan CK, Huang H. In: Subrahmanian VS, editor. Anomaly Detection Principles and Algorithms. Cham, Switzerland: Springer International Publishing; 2017.
10. Khan SS, Madden MG. One-class classification: taxonomy of study and review of techniques. Knowl Eng Rev 2014 Jan 24;29(3):345-374 [FREE Full text] [doi: 10.1017/s026988891300043x]
11. Ding X, Li Y, Belatreche A, Maguire LP. An experimental evaluation of novelty detection methods. Neurocomputing 2014 Jul;135:313-327 [FREE Full text] [doi: 10.1016/j.neucom.2013.12.002]
12. Tax DM. One-Class Classification: Concept Learning in the Absence of Counter-Examples. South Holand: Technische Universiteit Delft; 2002.
13. Irigoien I, Sierra B, Arenas C. Towards application of one-class classification methods to medical data. ScientificWorldJournal 2014;2014:730712 [FREE Full text] [doi: 10.1155/2014/730712] [Medline: 24778600]
14. Japkowicz N. Concept Learning in the Absence of Counterexamples: an Autoassociation-Based Approach to Classification. Semantic Scholar. 1999. URL: https://www.semanticscholar.org/paper/Concept-learning-in-the-absence-of-counterexamples%3A-Japkowicz-Hanson/03ed0a73d1f7a7b16505d6cb9c8bfbeeef7b19bb [accessed 2020-07-23]
15. Juszczak P, Tax DM, Pe¸kalska E, Duin RP. Minimum spanning tree based one-class classifier. Neurocomputing 2009 Mar;72(7-9):1859-1869 [FREE Full text] [doi: 10.1016/j.neucom.2008.05.003]
16. Mazhelis O. One-class classifiers: a review and analysis of suitability in the context of mobile-masquerader detection. S Afr Comput J 2006;2006(36):29-48 [FREE Full text]
17. Clark M. What is diabetes? In: Ogden J, editor. Understanding Diabetes. New Jersey, United States: John Wiley & Sons, Ltd; Jan 20, 2004.
18. Ogurtsova K, da Rocha FJ, Huang Y, Linnenkamp U, Guariguata L, Cho N, et al. IDF diabetes atlas: global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract 2017 Jun;128:40-50. [doi: 10.1016/j.diabres.2017.03.024] [Medline: 28437734]
19. Woldaregay A, Årsand E, Albers D, Launonen I, Hartvigsen G. Towards detecting infection incidences in people with type 1 diabetes using self-recorded data: a novel framework for a digital infectious disease detection mechanism. JMIR preprints 2020:- preprint [FREE Full text] [doi: 10.2196/preprints.18911]
20. Marcovecchio ML, Chiarelli F. The effects of acute and chronic stress on diabetes control. Sci Signal 2012 Oct 23;5(247):pt10. [doi: 10.1126/scisignal.2003508] [Medline: 23092890]
21. Rayfield EJ, Ault MJ, Keusch GT, Brothers MJ, Nechemias C, Smith H. Infection and diabetes: the case for glucose control. Am J Med 1982 Mar;72(3):439-450. [doi: 10.1016/0002-9343(82)90511-3] [Medline: 7036735]
22. Botsis T, Lai AM, Hripcsak G, Palmas W, Starren JB, Hartvigsen G. Proof of concept for the role of glycemic control in the early detection of infections in diabetics. Health Informatics J 2012 Mar;18(1):26-35. [doi: 10.1177/1460458211428427] [Medline: 22447875]
23. Botsis T, Hejlesen O, Bellika JG, Hartvigsen G. Blood Glucose Levels as an Indicator for the Early Detection of Infections In Type-1 Diabetics. Faculty & Staff Insider - University of Washington. 2007. URL: http://faculty.washington.edu/lober/www.isdsjournal.org/htdocs/articles/2025.pdf [accessed 2020-07-23]
24. Mizock BA. Alterations in carbohydrate metabolism during stress: a review of the literature. Am J Med 1995 Jan;98(1):75-84. [doi: 10.1016/S0002-9343(99)80083-7] [Medline: 7825623]
25. Bosarge PL, Kerby JD. Stress-induced hyperglycemia: is it harmful following trauma? Adv Surg 2013;47:287-297. [doi: 10.1016/j.yasu.2013.03.002] [Medline: 24298857]
26. Kajbaf F, Mojtahedzadeh M, Abdollahi M. Mechanisms underlying stress-induced hyperglycemia in critically ill patients. Therapy 2007 Jan;4(1):97-106 [FREE Full text] [doi: 10.2217/14750708.4.1.97]
27. Tax DM, Duin RP. Support vector data description. Mach Learn 2004 Jan;54(1):45-66 [FREE Full text] [doi: 10.1023/b:mach.0000008084.60811.49]
28. Schölkopf B, Williamson RC, Smola AJ, Shawe-Taylor J, Platt JC. Support Vector Method for Novelty Detection. NIPS Proceedings. 1999. URL: https://papers.nips.cc/paper/1723-support-vector-method-for-novelty-detection.pdf [accessed 2020-07-23]
29. Tax DM, Duin RP. Support vector domain description. Pattern Recognit Lett 1999 Nov;20(11-13):1191-1199 [FREE Full text] [doi: 10.1016/s0167-8655(99)00087-2]
30. Ridder DD, Tax DM, Duin, RP. An Experimental Comparison of One-Class Classification Methods. In: Proceedings of the 4th Annual Conference of the Advanced School for Computing and Imaging. 1998 Presented at: Annual Conference of the Advacned School for Computing and Imaging; September 29-October 4, 1998; Delft, Netherlands URL: https://www.researchgate.net/publication/282673032_An_experimental_comparison_of_classification_algorithm_performances_for_highly_imbalanced_datasets
31. Breunig MM, Kriegel H, Ng RT, Sander J. LOF: identifying density-based local outliers. SIGMOD Rec 2000 Jun 1;29(2):93-104 [FREE Full text] [doi: 10.1145/335191.335388]

32. Tax D. Software. DDTools. Delft: Delft University of Technology; 2015. URL: https://www.tudelft.nl/ewi/over-de-faculteit/afdelingen/intelligent-systems/pattern-recognition-bioinformatics/pattern-recognition-bioinformatics/data-and-software/dd-tools [accessed 2019-02-10]
33. A Collection of Algorithms for Anomaly/Outlier Detection. Anomaly Detection Toolbox. 2016. URL: http://dsmi-lab-ntust.github.io/AnomalyDetectionToolbox/ [accessed 2019-03-25]
34. Tang J, Chen Z, Fu AW, Cheung DW. Enhancing effectiveness of outlier detections for low density patterns. In: Advances in Knowledge Discovery and Data Mining. New York, USA: Springer; 2002.
35. Duin R, Juszczak P, Paclik P, Pekalska E, De Ridder D, Tax DM. Software. Delft University of Technology. 2007. URL: http://prtools.tudelft.nl/Guide/37Pages/software.html [accessed 2019-02-25]
36. Goldstein M, Uchida S. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLoS One 2016;11(4):e0152173 [FREE Full text] [doi: 10.1371/journal.pone.0152173] [Medline: 27093601]
37. Swersky L, Marques H, Sander J, Campello RJ, Zimek A. On the Evaluation of Outlier Detection and One-Class Classification Methods. In: IEEE International Conference on Data Science and Advanced Analytics (DSAA). 2016 Presented at: DSAA'16; October 17-19, 2016; Montreal, QC, Canada URL: https://doi.org/10.1109/DSAA.2016.8 [doi: 10.1109/dsaa.2016.8]
38. Rousseeuw PJ, Driessen KV. A fast algorithm for the minimum covariance determinant estimator. Technometrics 1999 Aug;41(3):212-223 [FREE Full text] [doi: 10.1080/00401706.1999.10485670]
39. Parzen E. On estimation of a probability density function and mode. Ann Math Statist 1962 Sep;33(3):1065-1076 [FREE Full text] [doi: 10.1214/aoms/1177704472]
40. Rasoulzadeh V, Erkus EC, Yogurt TA, Ulusoy I, Zergeroğlu SA. A comparative stationarity analysis of EEG signals. Ann Oper Res 2016 Apr 26;258(1):133-157 [FREE Full text] [doi: 10.1007/s10479-016-2187-3]
41. Azami H, Mohammadi K, Bozorgtabar B. An improved signal segmentation using moving average and Savitzky-Golay filter. J Signal Inf Process 2012;03(01):39-44. [doi: 10.4236/jsip.2012.31006]
42. Kandanaarachchi S, Muñoz MA, Hyndman RJ, Smith-Miles K. On normalization and algorithm selection for unsupervised outlier detection. Data Min Knowl Disc 2019 Nov 21;34(2):309-354 [FREE Full text] [doi: 10.1007/s10618-019-00661-z]
43. Tax DM, Muller KA. A Consistency-based Model Selection for One-Class Classification. In: Proceedings of the 17th International Conference on Pattern Recognition. 2004 Presented at: ICPR'04; August 26, 2004; Cambridge, UK. [doi: 10.1109/icpr.2004.1334542]
44. Maldonado S, Montecinos C. Robust classification of imbalanced data using one-class and two-class SVM-based multiclassifiers. Intell Data Anal 2014 Jan 1;18(1):95-112. [doi: 10.3233/ida-130630]
45. Tharwat A. Classification assessment methods. Appl Comput Inf 2018 Aug:- epub ahead of print [FREE Full text] [doi: 10.1016/j.aci.2018.08.003]
46. Nguyen GH, Bouzerdoum A, Phung SL. Learning pattern classification tasks with imbalanced data sets. Pattern Recog 2009:- [FREE Full text] [doi: 10.5772/7544]
47. Hajizadeh S, Li Z, Dollevoet RP, Tax DM. Evaluating classification performance with only positive and unlabeled samples. In: Fränti P, Brown G, Loog M, Escolano F, Pelillo M, editors. Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshop. Berlin, Heidelberg: Springer; 2014.
48. Hajizadeh S, Núñez A, Tax DM. Semi-supervised rail defect detection from imbalanced image data. IFAC-PapersOnLine 2016;49(3):78-83 [FREE Full text] [doi: 10.1016/j.ifacol.2016.07.014]
49. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recog 1997 Jul;30(7):1145-1159 [FREE Full text] [doi: 10.1016/s0031-3203(96)00142-2]
50. McNamara LA, Martin S. Principles of epidemiology and public health. In: Long SS, Prober CG, Fischer M, editors. Principles and Practice of Pediatric Infectious Diseases. Fifth Edition. New York, USA: Elsevier; 2018.
51. Zhu Y. Automatic detection of anomalies in blood glucose using a machine learning approach. J Commun Netw 2011 Apr;13(2):125-131 [FREE Full text] [doi: 10.1109/jcn.2011.6157411]
52. Zhu Y. Automatic Detection of Anomalies in Blood Glucose Using a Machine Learning Approach. In: International Conference on Information Reuse & Integration. 2010 Presented at: IER'10; August 4-6, 2010; Las Vegas, NV, USA URL: https://ieeexplore.ieee.org/document/5558959 [doi: 10.1109/jcn.2011.6157411]
53. Spinosa EJ, Carvalho AC. Support vector machines for novel class detection in Bioinformatics. Genet Mol Res 2005 Sep 30;4(3):608-615. [Medline: 16342046]
54. Lotze TH. Anomaly Detection in Time Series: Theoretical and Practical Improvements for Disease Outbreak Detection. Digital Repository at the University of Maryland. 2009. URL: https://drum.lib.umd.edu/handle/1903/9857 [accessed 2020-07-23]
55. Tsui F, Espino JU, Dato VM, Gesteland PH, Hutman J, Wagner MM. Technical description of RODS: a real-time public health surveillance system. J Am Med Inform Assoc 2003;10(5):399-408 [FREE Full text] [doi: 10.1197/jamia.M1345] [Medline: 12807803]
56. Woldaregay AZ, Årsand E, Botsis T, Albers D, Mamykina L, Hartvigsen G. Data-driven blood glucose pattern classification and anomalies detection: machine-learning applications in type 1 diabetes. J Med Internet Res 2019 May 1;21(5):e11030 [FREE Full text] [doi: 10.2196/11030] [Medline: 31042157]

57. Oviedo S, Vehí J, Calm R, Armengol J. A review of personalized blood glucose prediction strategies for T1DM patients. Int J Numer Method Biomed Eng 2017 Jun;33(6):---. [doi: 10.1002/cnm.2833] [Medline: 27644067]
58. Yki-Järvinen H, Sammalkorpi K, Koivisto VA, Nikkilä EA. Severity, duration, and mechanisms of insulin resistance during acute infections. J Clin Endocrinol Metab 1989 Aug;69(2):317-323. [doi: 10.1210/jcem-69-2-317] [Medline: 2666428]
59. Rayfield EJ, Curnow RT, George DT, Beisel WR. Impaired carbohydrate metabolism during a mild viral illness. N Engl J Med 1973 Sep 20;289(12):618-621. [doi: 10.1056/NEJM197309202891207] [Medline: 4198822]
60. McGuinness OP. Defective glucose homeostasis during infection. Annu Rev Nutr 2005;25:9-35. [doi: 10.1146/annurev.nutr.24.012003.132159] [Medline: 16011457]
61. Brealey D, Singer M. Hyperglycemia in critical illness: a review. J Diabetes Sci Technol 2009 Nov 1;3(6):1250-1260 [FREE Full text] [doi: 10.1177/193229680900300604] [Medline: 20144378]
62. Mizock BA. Alterations in fuel metabolism in critical illness: hyperglycaemia. Best Pract Res Clin Endocrinol Metab 2001 Dec;15(4):533-551. [doi: 10.1053/beem.2001.0168] [Medline: 11800522]
63. Tax DM, Duin RP. Characterizing One-Class Datasets. CiteSeerX. 2005. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.460.8322&rep=rep1&type=pdf [accessed 2019-07-10]
64. Wang D, Yeung DS, Tsang EC. Structured one-class classification. IEEE Trans Syst Man Cybern B Cybern 2006 Dec;36(6):1283-1295. [doi: 10.1109/tsmcb.2006.876189] [Medline: 17186805]
65. Janssens JH. Outlier Selection and One-Class Classification. Netherlands: Embedded Systems Institute, Tilburg University; 2013.
66. Wang B, Mao Z. One-class classifiers ensemble based anomaly detection scheme for process control systems. T I Meas Control 2017 Sep 21;40(12):3466-3476 [FREE Full text] [doi: 10.1177/0142331217724508]
67. Samerski S. Individuals on alert: digital epidemiology and the individualization of surveillance. Life Sci Soc Policy 2018 Jun 14;14(1):13 [FREE Full text] [doi: 10.1186/s40504-018-0076-z] [Medline: 29900518]
68. Radin JM, Wineinger NE, Topol EJ, Steinhubl SR. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study. Lancet Digit Health 2020 Feb;2(2):e85-e93 [FREE Full text] [doi: 10.1016/s2589-7500(19)30222-5]
69. Woldaregay A, Årsand E, Giordanengo A, Albers D, Mamykina L, Botsis T, et al. EDMON-A Wireless Communication Platform for a Real-time Infectious Disease Outbreak Detection System Using Self-recorded Data From People With Type 1 Diabetes. In: The 15th Scandinavian Conference on Health Informatics. 2017 Presented at: SHI'17; August 29, 2017; Kristiansand, Norway URL: https://ep.liu.se/konferensartikel.aspx?series=ecp&issue=145&Article_No=3
70. Coucheron S, Woldaregay AZ, Årsand E, Botsis T, Hartvigsen G. EDMON - A System Architecture for Real-Time Infection Monitoring and Outbreak Detection Based on Self-Recorded Data from People with Type 1 Diabetes: System Design and Prototype Implementation. In: The 17th Scandinavian Conference on Health Informatics. 2019 Presented at: CHI'19; November 12-13, 2019; Oslo, Norway URL: https://ep.liu.se/ecp/161/007/ecp19161007.pdf
71. Yeng PK, Woldaregay AZ, Solvoll T, Hartvigsen G. Cluster detection mechanisms for syndromic surveillance systems: systematic review and framework development. JMIR Public Health Surveill 2020 May 26;6(2):e11512 [FREE Full text] [doi: 10.2196/11512] [Medline: 32357126]

Abbreviations

AUC: area under the receiver operating characteristic curve
COF: connectivity-based outlier factor
IncSVDD: incremental support vector data description
K-NN: K-nearest neighbor
LOF: local outlier factor
MCD: minimum covariance determinant
MOG: mixture of Gaussian
MST: minimum spanning tree
NN: nearest neighbor
PCA: principal component analysis
SOM: self-organizing maps
SVDD: support vector data description
ROC: receiver operating characteristic curve
V-SVM: one-class support vector machine

Edited by G Eysenbach; submitted 26.03.20; peer-reviewed by S Sarbadhikari, M Nomali; comments to author 13.04.20; revised version received 06.06.20; accepted 11.06.20; published 12.08.20.

Please cite as:
Woldaregay AZ, Launonen IK, Albers D, Igual J, Årsand E, Hartvigsen G
A Novel Approach for Continuous Health Status Monitoring and Automatic Detection of Infection Incidences in People With Type 1 Diabetes Using Machine Learning Algorithms (Part 2): A Personalized Digital Infectious Disease Detection Mechanism
J Med Internet Res 2020;22(8):e18912
URL: https://www.jmir.org/2020/8/e18912
doi:10.2196/18912
PMID:32784179

Original Paper

A Comprehensive Evaluation of the Process of Copying a Complex Figure in Early- and Late-Onset Alzheimer Disease: A Quantitative Analysis of Digital Pen Data

Ko Woon Kim[1,2,3], MD, PhD; Sung Yun Lee[4,5], BSc; Jongdoo Choi[6,7], MSc; Juhee Chin[6,7], PhD; Byung Hwa Lee[6,7], PhD; Duk L Na[6,7,8,9,10*], MD, PhD; Jee Hyun Choi[5*], PhD

[1]Department of Neurology, School of Medicine, Jeonbuk National University Hospital, Jeonju, Republic of Korea
[2]Research Institute of Clinical Medicine of Jeonbuk National University, Jeonju, Republic of Korea
[3]Biomedical Institute of Jeonbuk National University Hospital, Jeonju, Republic of Korea
[4]Department of Physics, Pohang University of Science and Technology, Pohang, Republic of Korea
[5]Center for Neuroscience, Korea Institute of Science and Technology, Seoul, Republic of Korea
[6]Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[7]Neuroscience Center, Samsung Medical Center, Seoul, Republic of Korea
[8]Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea
[9]Stem Cell & Regenerative Medicine Institute, Samsung Medical Center, Seoul, Republic of Korea
[10]Samsung Alzheimer Research Center, Samsung Medical Center, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Jee Hyun Choi, PhD
Center for Neuroscience
Korea Institute of Science and Technology
Hwarang-ro 14-gil 5
Seongbuk-gu
Seoul, 02792
Republic of Korea
Phone: 82 29586952
Email: jeechoi@kist.re.kr

Abstract

Background: The Rey-Osterrieth Complex Figure Test (RCFT) is a neuropsychological test that is widely used to assess visual memory and visuoconstructional deficits in patients with cognitive impairment, including Alzheimer disease (AD). Patients with AD have an increased tendency for exhibiting extraordinary behaviors in the RCFT for selecting the drawing area, organizing the figure, and deciding the order of images, among other activities. However, the conventional scoring system based on pen and paper has a limited ability to reflect these detailed behaviors.

Objective: This study aims to establish a scoring system that addresses not only the spatial arrangement of the finished drawing but also the drawing process of patients with AD by using digital pen data.

Methods: A digital pen and tablet were used to copy complex figures. The stroke patterns and kinetics of normal controls (NCs) and patients with early-onset AD (EOAD) and late-onset AD (LOAD) were analyzed by comparing the pen tip trajectory, spatial arrangement, and similarity of the finished drawings.

Results: Patients with AD copied the figure in a more fragmented way with a longer pause than NCs (EOAD: *P*=.045; LOAD: *P*=.01). Patients with AD showed an increased tendency to draw the figures closer toward the target image in comparison with the NCs (EOAD: *P*=.005; LOAD: *P*=.01) Patients with AD showed the lower accuracy than NCs (EOAD: *P*=.004; LOAD: *P*=.002). Patients with EOAD and LOAD showed similar but slightly different drawing behaviors, especially in space use and in the initial stage of drawing.

Conclusions: The digitalized complex figure test evaluated copying performance quantitatively and further elucidated the patients' ongoing process during copying. We believe that this novel approach can be used as a digital biomarker of AD. In

addition, the repeatability of the test will delineate the process of executive functions and constructional organization abilities with disease progression.

(J Med Internet Res 2020;22(8):e18136) doi:10.2196/18136

KEYWORDS

alzheimer disease; Rey-Osterrieth Complex Figure; digital biomarkers; copying process

Introduction

Background

The Rey-Osterrieth Complex Figure Test (RCFT) is widely used to examine the visuoconstruction and visual memory function of patients with brain injuries or cognitive impairment such as Alzheimer disease (AD) [1]. The conventional scoring system of RCFT focuses on scoring the final finished drawing of the figure by assessing the shape and positional accuracy of its elements. There are only a few scoring methods that quantify the drawing sequences of elements [2-4]. However, patients with cognitive impairment due to brain injuries show different patterns in the drawing process and errors on the final finished drawing compared with normal controls (NC; ie, individuals with normal cognition) [5]. For example, some patients with brain injuries complete a figure by adding one detailed part after another instead of starting with an overall outline and adding local or detailed parts. Furthermore, the number, length, and speed of strokes made by patients with brain injuries may differ from that of NCs. Therefore, it is necessary to establish a scoring system using a digital device that records not only the spatial arrangement of the finished drawing but also all the details of drawing processes, including the sequence in drawing the parts of the figures. This approach will augment our understanding of the organizing strategy and executive functions of patients with brain injuries. Additionally, it will increase our understanding of the structural and design integrity of drawing figures.

A digital pen and tablet can be optimal tools for observing and acquiring data for the drawing process. Recent studies have used a digital device to evaluate the visuoconstructional abilities and executive functions of patients with AD. These researchers used a clock drawing test [6,7] or a 3D house copy test [8] to quantitatively evaluate cognitive function in patients with AD. It was possible to differentiate patients with mild cognitive impairment (MCI) and AD from NCs by assessing physical parameters such as kinetics (eg, time, velocity) or device-human interaction values (eg, pressure, number/density of strokes). Compared with NCs, even patients with MCI who received normal scores on the clock drawing test showed prolonged transition times associated with drawing performances [6]. The kinetics and sequence-based schemes from digital drawing tests have successfully demonstrated changes in cognitive progress. Previously, only 1 study has analyzed RCFT using a digital device; however, the group did not assess the drawing procedure but focused on recognizing the outline and details in relation to the conventional scoring system [9].

Objective

To implement the scheme in a digital device, we first decided to choose an abstract figure such as the RCFT rather than concrete objects such as a clock or a house. Clocks and houses elicit semantic knowledge, which primarily involves the ventral visual pathway (*what* pathway). Using complex figures on the other hand, such as the RCFT, involves the dorsal visual pathway (*where* pathway) and requires visuospatial working memory before translation into a motor program and execution of plan [10]. Second, we simplified the Rey figure to reduce not only the drawing time but also the number of strokes required to complete the figure. Arranging too many structures in a limited space may result in stroke overlap when drawing, which may prevent the digital device from identifying each stroke.

In this study, we hypothesized that the movement kinetics (pen tip trajectory of the digital pen) and digitized spatial information acquired from the simplified RCFT would differ between NCs and patients with AD. Several previous studies have reported that patients with early-onset AD (EOAD) have significant visuospatial or visuoconstructive difficulties compared with patients with late-onset AD (LOAD) [11-13]. We therefore hypothesized that patients with EOAD would show more pronounced changes in movement kinetics and digitized spatial information compared with patients with LOAD.

Methods

Recruitment of Participants

Participants were selected from those who visited the Memory Disorders Clinic at the Samsung Medical Center in Seoul, Republic of Korea, between March 1 and December 1, 2017. The study group comprised 17 patients with EOAD, 21 patients with LOAD, and 17 NC individuals. We consecutively recruited participants who satisfied the following criteria: (1) normal visual acuity, (2) >6 years of education, (3) completion of a standardized neuropsychological battery called the Seoul Neuropsychological Screening Battery (SNSB) [14,15] and the Mini-Mental State Examination (MMSE) test [16], and (4) magnetic resonance imaging (MRI) results. Participants were classified via rigorous diagnostic methods, including multiple tests (eg, blood test, MRI/positron emission tomography scans) and clinical consensus of neurologists, neuropsychologists, and radiologists. All patients with AD fulfilled the criteria proposed by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) [17]. Patients with EOAD were defined as those whose first symptoms occurred at an early age (>45 years and <65 years of age). Individuals with moderate or severe vision loss (visual acuity <0.3) or those

with very low MMSE scores with lower cutoffs of 10 or education levels <6th grade were not included in this study.

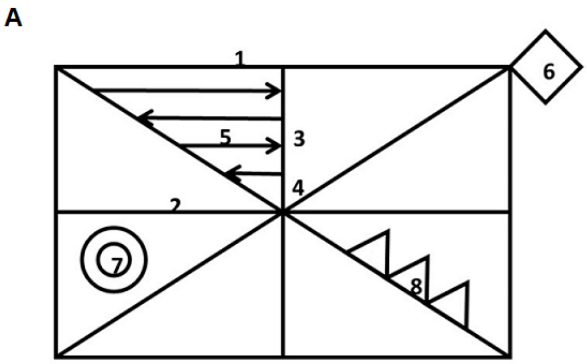
Neuropsychological Assessments

All participants underwent a standardized neuropsychological battery called the SNSB [14,15], which consists of tests for attention, language, calculation, visuospatial, memory, and frontal/executive functions. The MMSE and Clinical Dementia Rating tests were also carried out to evaluate general cognition. The standardized tests used for this study were as follows: attention was assessed using the backward digit span and letter cancellation tests; language was assessed using the Korean version of the Boston Naming Test; calculation was assessed using 3 items comprising each for addition, subtraction, multiplication, and division; visuospatial function was assessed using the RCFT; memory function was assessed using immediate and delayed recall of the Seoul Verbal Learning Test and RCFT; frontal-executive function was assessed using the phonemic and semantic Controlled Oral Word Association Test and the Stroop word/color reading test. Each score was converted into a standardized Z-score based on age- and education-adjusted norms.

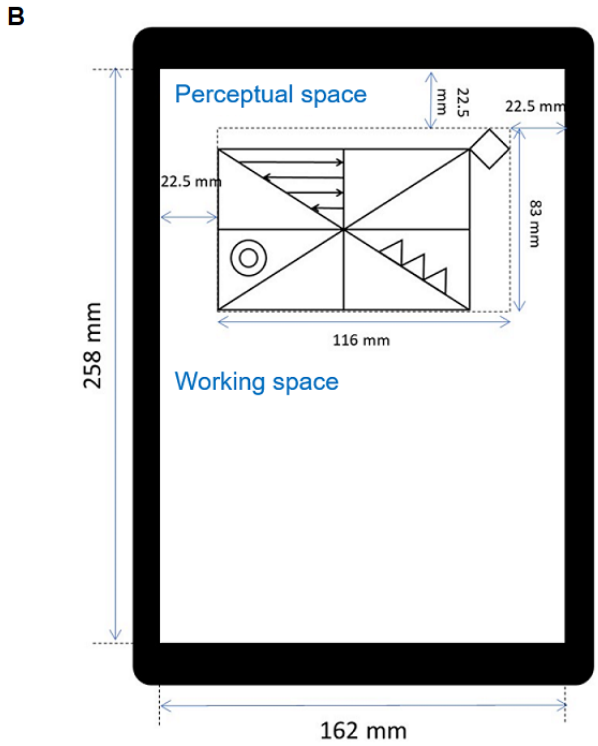
Simplified Version of the Rey-Osterrieth Complex Figure Test

We modified the original RCFT into a simpler version while maintaining the main frame of the figure. Modification was performed by 1 neurologist (experience >10 years) and 2 neuropsychologists (average experience >20 years) to balance the numbers of global and local components. Although the original RCFT comprises 18 components (4 global and 14 local components) [18], our simplified RCFT comprised 4 global and 4 local components, as illustrated in Figure 1. The large rectangle with horizontal, vertical, and diagonal crosses was maintained, but other regional or local features were simplified except for the 4 horizontal lines in the upper left panel. For the local features, the components in the original RCFT were changed as follows: (1) the diamond was replaced by a square, (2) the circle with 3 dots was replaced by double circles, (3) the 5 parallel lines were replaced by 3 triangles, (4) and the 4 parallel lines were replaced by 4 arrows. We added arrows to these lines such that the participants would be more attentive to the local features. Some overlapping lines and detailed components outside of the outlines were eliminated in the simplified version. For the global components, the side of the large triangle attached to the large rectangle and the horizontal line of the large triangle were excluded. Of the local components, the following were removed: the vertical cross, the small triangle above the large rectangle, the horizontal cross, the square attached to the large rectangle, the small rectangle, the small horizontal line above the small rectangle, the vertical line within the side of the large triangle, and the small vertical line within the large rectangle.

Figure 1. The simplified Rey-Osterrieth Complex Figure Test (RCFT). (A) The simplified RCFT consists of 4 global and 4 local components. (B) The simplified RCFT is shown on a tablet (size: 12 in; resolution: 2160 × 1440; Samsung Galaxy Book 12, Samsung Electronics) with a screen width of 162 mm and a screen height of 258 mm. The texts indicated in blue were not presented to the participants during testing.



Global		Local	
1	Large rectangle	5	Parallel arrows
2	Horizontal line	6	Square
3	Vertical line	7	Double circles
4	Diagonal cross	8	Three triangles



Each of the 8 components in the simplified RCFT was scored separately (Meyers and Meyers protocol) in terms of accuracy and placement [18]. Component scores were assigned as 2 (accurately drawn, correctly placed), 1 (accurately drawn, incorrectly placed or inaccurately drawn, correctly placed), 0.5 (inaccurately drawn, incorrectly placed but recognizable), or 0 (inaccurately drawn, incorrectly placed, unrecognizable). Thus, the possible range of raw scores was 0.0-16.0. A low score

suggested impaired visuoperceptual or visuoconstructive functions.

Experimental Apparatus (Digitized Equipment) and Drawing Procedure

The experimental equipment for measuring the performance of copying a figure consisted of an X-Y digital tablet (size: 12-inch, resolution: 2160 × 1440; Samsung Galaxy Book 12, Samsung Electronics) and a digital pen (nib diameter: 0.7 mm, pressure: 4096; S-pen, Samsung Electronics). Participants performed the simplified RCFT on a tablet in portrait orientation (Figure 1). We defined the upper half of the working area as *perceptual space* and the lower half as *working space*. The simplified RCFT was presented in the perceptual space. An eraser tool was not included in the interface to observe the basal level of copying. Before the task, the experimenter delivered the following instructions to the participants: "When the task begins, a figure will appear on the upper half of this tablet. Please copy the figure with this pen. Try to use the empty area below as much as possible. This pen does not have an eraser. You cannot change your drawing. If you made any mistakes, please disregard them and continue with carrying out the task. Take your time, and please let me know when you are finished." The trajectories of the drawings were recorded on the X-Y digital tablet at a sampling frequency of 60 Hz, and the x- and y-axes were horizontal and vertical planes, respectively.

Data Processing in Normalized Cartesian Coordinates

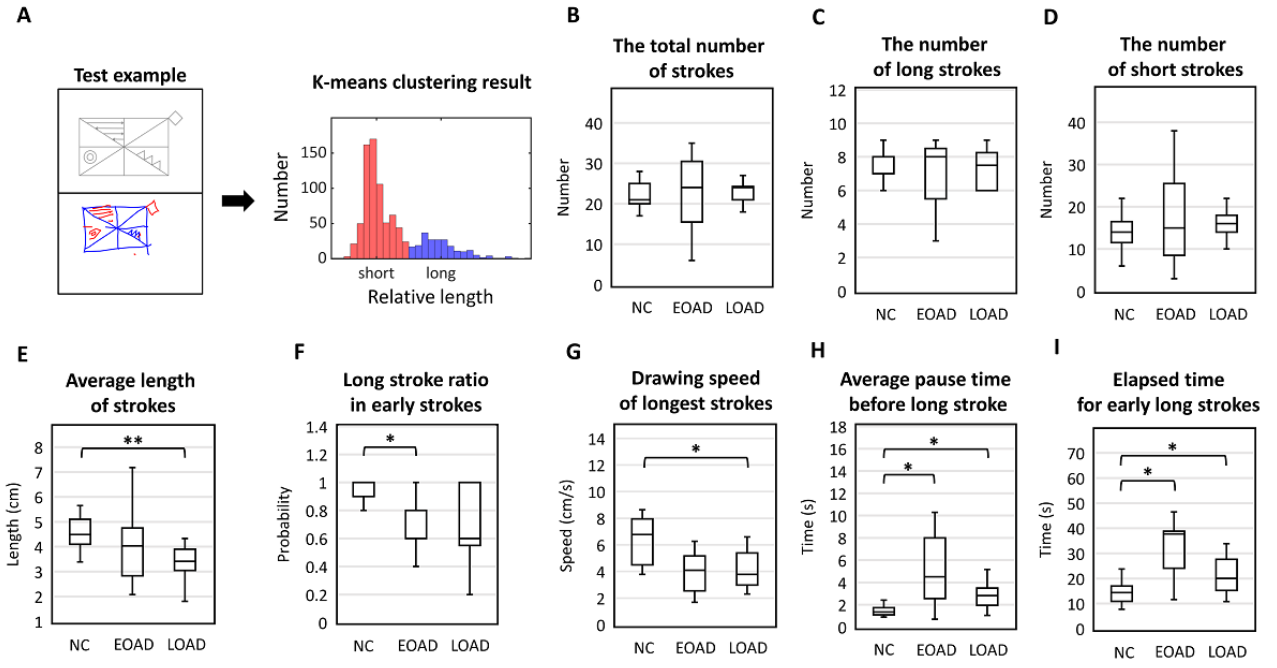
The digital pen-touch point responses were recorded by the X-Y digital tablet as Cartesian coordinates in pixel units based on placement on the screen, movement, and removal from the screen over time. The range of positions was (0, 0) to (2160, 1440), which corresponded to the range of display in pixels. We normalized the position from a range of –1 to 1 with respect to the origin (0, 0) at the center of the display. The vertical ranges of perceptual and working spaces corresponded to 0 to 1 and –1 to 0, respectively. To consider the robustness of the automatic evaluation, we focused on analyzing parameters such as pen stroke, occupied drawing area (drawing boundary), and copying results (similarity between the original figure).

We analyzed 3 features: (1) digital pen stroke, which was defined as continuous movement of the digital pen while maintaining contact with the writing tablet; (2) drawing boundary, defined as the extremum coordinates; and (3) copying results (similarity of the drawing to the original). These features were selected based on their robustness to automatic evaluation.

Analysis of Pen Strokes

Pen stroke was referred to as a trajectory created based on continuous contact of the digital pen. Specifically, pen stroke was defined as pen movements recorded from the moment the pen touched the screen to the moment the pen was lifted off the screen. To analyze stroke behaviors, the number, length, and speed were assessed. We also analyzed data after classifying strokes as long or short lines (Figure 2).

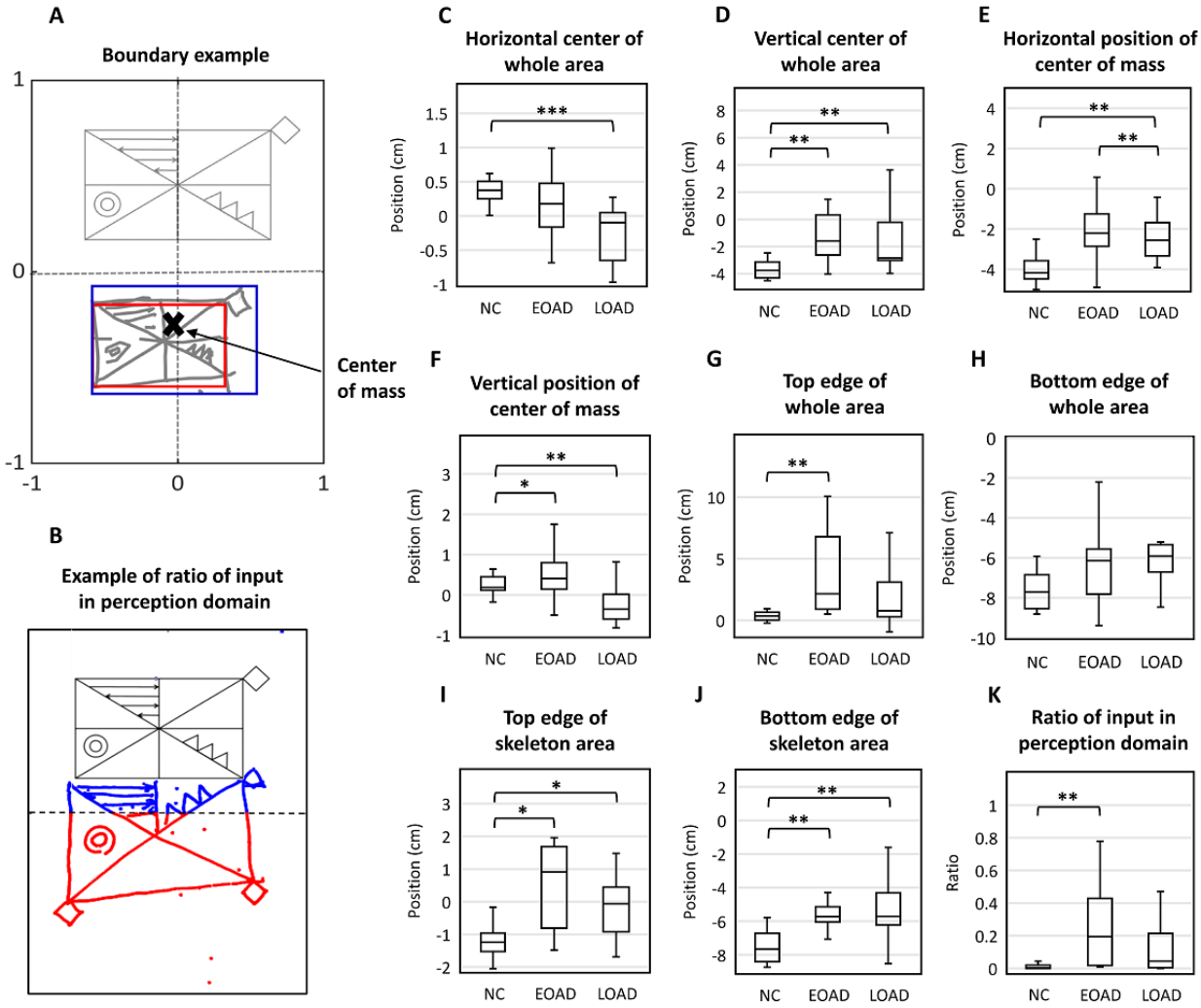
Figure 2. Pen stroke analysis. Clinical interpretation of copying performance in the Rey-Osterrieth Complex Figure Test (RCFT) was partly based on how an individual draws the image based on consecutive line creations. The number, length, and speed of strokes were analyzed. (A) Long (blue)/short (red) strokes are classified by k-means clustering; (B-D) number of strokes; (E-F) length of strokes; (G-I) speed of strokes. One and two asterisks show significant differences at $P<.05$ and $P<.01$ levels, respectively. EOAD: early-onset Alzheimer disease; LOAD: late-onset Alzheimer disease; NC: normal control.



Analysis of Drawing Areas (Spatial Arrangement of Drawings)

To evaluate the space occupancy of drawings and possible signs of unilateral neglect of space, we defined the boundary of the drawing region and measured its area and lateralization. To compare the space occupancy for all items with that of the global items (large square with diagonal axes), we defined 2 boundaries of *whole area*, which contained the whole drawing, and *skeleton area*, of that within the rectangle border (Figure 3). For ease of calculation, we defined the area as a square. The entire area included all input drawings and was determined by 4 extremums along the horizontal and vertical axes. By definition, the skeleton area was the boundary of the square created by minimizing the area following consideration of all global features. To perform automatic calculations, we determined the skeleton area using the following steps: (1) the input frequency was determined at each subregion in which the working space was divided into 10×10 subregions, (2) inputs in subregions with frequencies lower than 1 out of 4 were excluded, and (3) the extremums were obtained with the remaining inputs and defined as the skeleton area. After detection of boundaries, we obtained the center, vertex, and edge lengths of the regions. We further calculated the center of mass, , using the following equation:

Figure 3. Spatial arrangement. (A) We defined the bounded area that contains the whole drawing as the whole area (blue box) and a boundary that contains only the area inside the skeleton rectangle as the skeleton area (red box). (B) An example of the ratio of input in perceptual space: input (blue)/total input (blue + red). (C-D) Center of the whole area. (E-F) Center of mass of the whole area. (G-H) Top and bottom edge of the whole area. (I-J) Top and bottom edges of the skeletal area. (K) Closing-in phenomenon. One, two, and three asterisks show significant differences at $P<.05$, $P<.01$, and $P<.001$ levels, respectively. EOAD: early-onset Alzheimer disease; LOAD: late-onset Alzheimer disease; NC: normal control; RCFT: Rey-Osterrieth Complex Figure Test.



(1)

where m=1 if there is input at the ith pixel at or 0 if not.

Analysis of Copying Similarity

To evaluate the accuracy of copying, we performed a 2-dimensional (2D) cross-correlation analysis Figure 4. In this analysis, the input function *m* was scanned over the perceptual and working spaces with respect to the stimulus function *M*, with a value of 0 or 1 in the absence or presence of input,

respectively. The amplitude distribution of the 2D cross-correlation function, , was calculated as:

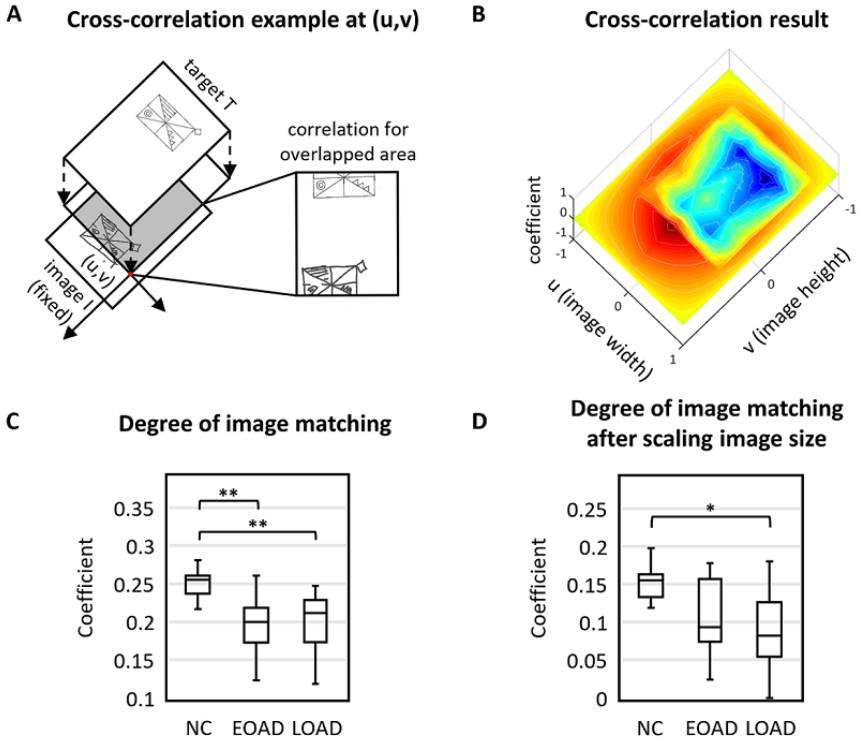
(2)

where μ_M and μ_m were the means of stimulus and input signals, respectively. The SD was computed as follows:

(3)

for $\sigma_m$and equivalently for σ_M without . All summation was performed within the overlapped area between *m* and *M*, when *m* was located at . The normalized cross-correlation function had values between −1 and +1. The maximum value of represented the similarity between the 2 patterns. Due to variation in size of the whole area between the images, we calculated the maximum correlation after rescaling along the x-axis. The coefficient of was obtained by determining its peak, which was assumed to be the center of the attention field.

Figure 4. Shape similarity between the target and the pictures drawn by the participants. (A) The Pearson correlation coefficient was used to assess cross-correlation for all possible overlap. (B). We applied 2D cross-correlation analysis and calculated its coefficient and shift in space. (C) For both AD groups, pattern matching was significantly poorer than that of the normal control (NC) group (pairwise Mann-Whitney U test: *P*=.004 for early-onset Alzheimer disease [EOAD] vs NCs and *P*=.002 for late-onset Alzheimer disease [LOAD] vs NCs). There was no statistical significance between the EOAD and LOAD groups (*P*=.86). (D) Although the amount of difference was reduced, the similarity was significantly lower for the LOAD group even after rescaling the image size (*P*=.04). There was no statistical significance for the EOAD group even after adjusting for image size. One and two asterisks show significant differences at *P*<.05 and *P*<.01 levels, respectively. RCFT: Rey-Osterrieth Complex Figure Test.



Statistical Analysis

For statistical analysis of the demographic and cognitive profile data, the Kolmogorov-Smirnov test was used to conduct normality tests. The Kruskal-Wallis test was used to examine the statistical significance between groups at a significance level of *P*=.05 because variables did not follow a normal distribution. We performed post hoc comparisons using the pairwise Mann-Whitney U test with Bonferroni correction. We used the chi-square test or Fisher exact test for categorical variables, followed by a Bonferroni post hoc analysis. To validate the simplified Rey figures, we used Pearson correlation to compare conventional RCFT scores and simplified RCFT scores.

For statistical analysis of digital pen data, the Lilliefors test was used to assess the normality and the Kruskal-Wallis test was again used to examine the statistical significance between groups at a confidence level of *P*=.05. We performed post hoc comparisons using the pairwise Mann-Whitney U test with a Dunn-Šidák correction by adjusting the *P* value of each pairwise test as $p=1-(1-p_o)^3$, where 3 reflects the number of groups. We compared the immediate/recall/recognition scores of the original RCFT and our digital metrics from the simplified RCFT using Pearson correlation to determine whether there are any data linked to memory function. All hypothesis tests were two-tailed. MATLAB R2018b was used for the calculations.

Results

Demographics and Cognitive Profiles

The demographic and cognitive profiles of the participants are summarized in Table 1. Age (P<.001) and education (P=.10) differed significantly among NCs and participants with EOAD and LOAD. The prevalence of apolipoprotein E4 carriers among NCs (0/13, 0%) was significantly lower than that among patients with EOAD (10/16, 63%) or LOAD (9/20, 45%). To investigate differences in neuropsychological scores among the 3 groups, we used Z scores based on age- and education-adjusted norms. Both patients with EOAD and LOAD showed poor performance in attention, language, visuospatial function, memory, and frontal/executive functions compared with NCs (Table 1).

Before the analysis, digital data were reviewed carefully, and data that were not appropriate were excluded from the analysis. A total of 3 participants were excluded because of mismatched digital pen data between attach/detach and movement pairs; 8 participants were excluded because of missing attach movements in coordinate data; 6 participants were excluded because of mismatched digital pen data between result image and pen movement. Digital data from a total of 11 NCs, 11 patients with EOAD, and 16 patients with LOAD were analyzed.

Table 1. Clinical characteristics.

Demographics and cognitive profiles	NC[a] (n=17)	EOAD[b] (n=17)	LOAD[c] (n=21)	Post hoc Bonferroni test (*P* value)		
				NC vs EOAD	NC vs LOAD	EOAD vs LOAD
Age (years), n (IQR)	73 (65.5-77.0)	65 (58.5-68.0)	78 (74.0-81.5)	*.011* [d]	*.005*	*<.001*
Gender						
Female:male	8:9	8:9	8:13	>.99	*<.001*	>.99
Education (years), n (IQR)	16 (10.5-16.0)	12 (9.0-13.0)	12 (11.0-16.0)	.12	.36	>.99
APOE4[e] carrier[f], n (%)	0[g] (0)	10[h]/16 (63)	9[i]/20 (45)	*.001*	*.015*	>.99
Amyloid PET[j] positive[k], n	0	15	8	*.004*	*.018*	>.99
Attention, span (IQR)						
Forward digit span	0.06 (–0.43 to 1.08)	0.12 (–1.50 to 20.99)	–0.15 (–0.59 to 0.83)	.86	>.99	>.99
Backward digit span	–0.16 (–0.55 to 1.19)	–0.82 (–1.66 to –0.03)	–0.13 (–1.07 to 0.80)	*.04*	>.99	0.41
Language, score (IQR)						
K-BNT[l]	0.27 (–1.90 to 0.86)	–0.92 (–2.52 to –0.08)	–2.60 (–3.38 to –1.61)	*.002*	*<.001*	0.15
Calculation	12 (12-12)	10 (7-12)	10 (8-12)	*.003*	*.001*	>.99
Visuospatial function, score (IQR)						
RCFT[m]: copying	0.55 (–0.70 to 0.93)	–1.72 (–7.61 to 0.00)	–1.02 (–7.78 to 0.12)	*0.008*	*0.001*	>.99
Memory, score (IQR)				*<.001*	*<.001*	
SVLT[n]: immediate recall	0.83 (0.24-1.48)	–2.25 (–2.85 to –0.95)	–1.79 (–2.45 to –0.87)			>.99
SVLT: delayed recall	0.88 (0.34-1.22)	–2.96 (–3.18 to –2.53)	–2.51 (–2.74 to –2.09)			*.008*
SVLT: recognition	0.89 (0.61-1.10)	–2.23 (–3.93 to –1.94)	–2.66 (–3.26 to –1.51)			>.99
RCFT: immediate recall	0.86 (0.61-1.10)	–2.23 (–3.93 to –1.94)	–2.66 (–3.26 to –1.51)			>.99
RCFT: delayed recall	1.36 (0.05-1.76)	–2.11 (–2.55 to –1.82)	–2.17 (–2.45 to –1.65)			>.99
RCFT: recognition	1.01 (0.04-1.47)	–2.48 (–2.75 to –2.16)	–2.36 (–2.58 to –1.97)			.56
Frontal/executive functions, score (IQR)						
COWAT[o] animal	0.23 (–0.30 to 0.98)	–2.21 (–3.14 to –1.89)	–2.36 (–2.96 to –1.17)	*<.001*	*<.001*	>.99
COWAT supermarket	0.23 (–0.78 to 1.00)	–1.60 (–2.06 to –1.32)	–1.73 (–1.99 to –1.04)	*.001*	*.001*	>.99
COWAT phonemic	0.70 (0.00-1.59)	–1.37 (–1.61 to –0.47)	–0.85 (–1.68 to 0.28)	*<.001*	*.01*	>.99
Stroop test: color	0.56 (–0.17 to 1.03)	–3.22 (–3.89 to –0.75)	–1.93 (–2.86 to –0.48)	*<.001*	*<.001*	.66
MMSE[p]	29 (28-30)	22 (15-25)	20 (17-22)	*<.001*	*<.001*	>.99
CDR[q]	0.5 (0.5-0.5)	1.0 (0.5-1.0)	1.0 (0.75-1.0)	*<.001*	*<.001*	>.99
CDR sum of box	0.5 (0.5-0.75)	5.0 (4.25-6.25)	4.5 (4.0-8.5)	*<.001*	*<.001*	>.99
GDS[r]	1.0 (0.0-4.0)	2.0 (1.0-5.0)	1.0 (0.5-7.0)	>.99	.71	>.99

[a]NC: normal control.

[b]EOAD: early-onset Alzheimer disease.

[c]LOAD: late-onset Alzheimer disease.

[d]Italicization show significant differences at $P<.05$.

[e]APOE4: apolipoprotein E4.

[f]APOE4 was analyzed in 49 patients: 13 NCs, 16 patients with EOAD, and 20 patients with LOAD. Participants with one or more copies of the ε4 allele (ie, ε2/4, ε3/4, ε4/4) were considered as ε4 carriers.

[g]n=13.

[h]n=16.

[i]n=20.

[j]PET: positron emission tomography.

[k]Amyloid PET was analyzed in 26 patients: 3 NCs, 15 patients with EOAD, and 8 patients with LOAD. Amyloid PET positivity was interpreted based on previously reported guidelines for each ligand.

[l]K-BNT: Korean version of the Boston naming test.

[m]RCFT: Rey-Osterrieth Complex Figure Test.

[n]SVLT: Seoul Verbal Learning Test.

[o]COWAT: Controlled Oral Word Association Test.

[p]MMSE: Mini-Mental State Examination.

[q]CDR: clinical dementia rating.

[r]GDS: geriatric depression scale.

Comparison Between Original and Simplified Rey-Osterrieth Complex Figure Test Scores

Table 2 shows the median and interquartile range raw copy scores of the original and simplified RCFT results from NCs, patients with EOAD, and patients with LOAD. The NCs showed significantly higher scores than the 2 AD groups in both the simplified (EOAD: $P<.001$; LOAD: $P<.001$) and original RCFT tests (EOAD: $P<.001$; LOAD: $P<.001$). There was no significant difference between patients with EOAD and patients with LOAD in the simplified or original RCFT. There was a significant linear relationship between the conventional RCFT and the simplified RCFT score (r=0.884; $P<.001$; Figure 5).

Table 2. Raw copy scores for the original and simplified Rey-Osterrieth Complex Figure Tests.

Types of copy scores	NC[a] (n=17)	EOAD[b] (n=17)	LOAD[c] (n=21)	Post hoc Bonferroni test (*P* value)		
				NC vs EOAD	NC vs LOAD	EOAD vs LOAD
Simplified RCFT[d] (0, 16; IQR)	15.5 (15-16)	12.5 (3.8-14.4)	13 (7.8-15)	*<.001* [e]	*<.001*	.69
Original RCFT (0, 36; IQR)	35 (33-35)	28 (10-32)	30 (8-31)	*<.001*	*<.001*	>.99

[a]NC: normal control.

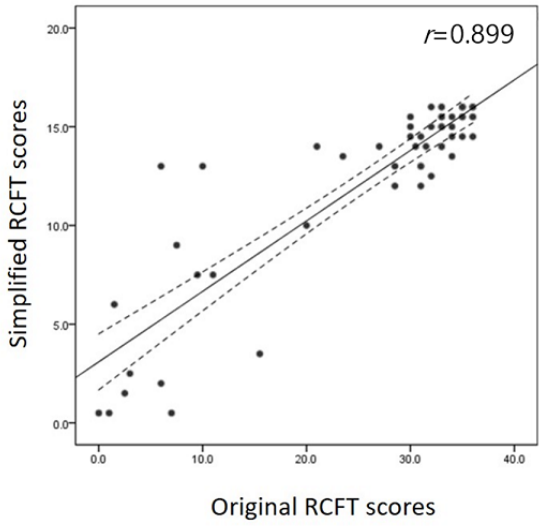
[b]EOAD: early-onset Alzheimer disease.

[c]LOAD: late-onset Alzheimer disease.

[d]RCFT: Rey-Osterrieth Complex Figure Test.

[e]Italicization show significant differences at $P<.001$.

Figure 5. Correlation between original and simplified Rey-Osterrieth Complex Figure Test (RCFT) scores. All participants performed both original and simplified RCFTs. The simplified RCFT was manually scored separately in terms of both accuracy and placement, and it complied with the Meyers and Meyers' standardized scoring of the original RCFT. Raw scores ranging from 0.0 to 16.0 can be obtained. There was a significant linear relationship between the conventional and simplified RCFT scores (r=0.889; *P*<.001).



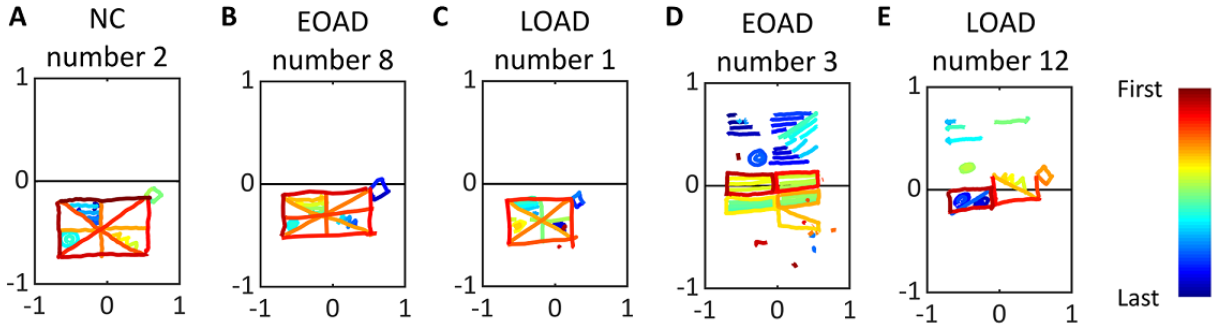
Comparison Between Original Rey-Osterrieth Complex Figure Test Scores and Digital Metrics From Simplified Rey-Osterrieth Complex Figure Test

We compared the immediate/recall/recognition scores of the original RCFT and the digital data acquired from our simplified RCFT to examine whether any of our digital data could be associated with memory function using Pearson correlation. A positive correlation (*r*=0.353; *P*=.04) was observed between immediate recall scores of the original RCFT and the coefficients of shape similarity. In addition, immediate recall scores of the original RCFT and vertical positions of the simplified RCFT on the y-axis negatively correlated as follows: vertical position of the center of the whole area (*r*=–0.341; *P*=.045), vertical position of the center of the skeletal area (*r*=–0.336; *P*=.048), vertical position of the center of the mass (*r*=–0.339; *P*=.047), vertical position of the top edge of the whole area (*r*=–0.343; *P*=.04), and location of the top edge of the skeleton area (*r*=–0.350; *P*=.04). However, there were no significant correlations between recall/recognition scores of the original RCFT and the digital data from the simplified RCFT.

Analysis of Copying Sequences

The organizing abilities of the participants were evaluated by analyzing the copying sequences. We converted the original drawing of the simplified Rey figure to pseudocolored images coded by a series of colors in the order of strokes (Multimedia Appendices 1-3). Specifically, the total number of strokes made by each participant was broken down into segments according to the order of strokes, and each segment was coded with a series of colors ranging from red to blue. For instance, when an individual made 40 strokes in total, the first 4 strokes that accounted for the first 10% (4/40) were assigned a red color and the next 10% (4/40) as yellow and so on. Most of the participants drew the global features first and then added the local features, as exemplified in Figure 6. However, some patients started drawing local features such as the double circle or the triangles, as shown in the representative figures (Figure 6) during the middle of drawing global features. In addition, 1 patient with EOAD (number 5 in Multimedia Appendix 2) completed the global structure last. Nonetheless, the tendency of drawing global elements in the beginning was preserved across all subjects. In the unorganized copy as well, participants started with the large box (Figure 6) and then moved onto the local elements.

Figure 6. Representative pseudocolored drawings for each group. Pseudocolored drawings of (A) normal control (NC) individuals, (B) patients with early-onset Alzheimer disease (EOAD), and (C) patients with late-onset Alzheimer disease (LOAD), who drew all the elements. (D) Patients with EOAD and (E) LOAD who failed to reproduce the stimulus image. The strokes were colored based on the order of the drawings.



Pen Stroke Analysis

Clinical interpretation of copying performance in the RCFT was partly based on individual performance in copying. Therefore, we assessed the number, length, and speed of sequential line strokes (Table 3).

Table 3. Digital data of simplified Rey-Osterrieth Complex Figure Tests.

Type of digital data analysis	NC[a] (n=11)	EOAD[b] (n=11)	LOAD[c] (n=16)	Post hoc Dunn-Šidák test (*P* value)		
				NC vs EOAD	NC vs LOAD	LOAD vs EOAD
Pen stroke analysis, mean (SD)						
Total number of strokes	24 (7)	26 (16)	23 (4)	>.99	>.99	>.99
Number of long strokes	14 (4)	18 (12)	16 (4)	>.99	>.99	>.99
Number of short strokes	8 (1)	9 (7)	7 (3)	>.99	>.99	>.99
Length of strokes (cm)	4.58 (0.71)	3.99 (1.48)	3.50 (0.86)	.34	*.007* [d]	.90
First 5 stroke ratios	0.91 (0.16)	0.69 (0.23)	0.68 (0.28)	.04	.07	>.99
Speed of the longest stroke (cm/s)	6.31 (1.86)	4.41 (2.76)	4.08 (1.41)	.09	*.01*	>.99
Transition time (second)	1.39 (0.60)	2.95 (1.83)	6.23 (7.54)	*.045*	*.01*	.97
Elapsed time of 5 early long strokes (second)	14.43 (4.62)	34.86 (19.39)	23.65 (14.16)	*.012*	*.04*	.32
Spatial arrangement, mean (SD)						
Whole area (cm^2)	84.17 (10.88)	108.51 (58.81)	84.01 (33.93)	>.99	>.99	>.99
Skeleton area (cm^2)	58.40 (8.79)	56.34 (20.78)	55.29 (18.93)	>.99	>.99	>.99
Horizontal center of the whole area	0.42 (0.31)	0.25 (0.67)	–0.23 (0.40)	.60	*<.001*	.11
Vertical center of the whole area	–3.70 (0.69)	–1.22 (1.84)	–1.61 (2.36)	*.005*	*.01*	.79
Horizontal position of center of mass	0.28 (0.34)	0.46 (0.64)	–0.35 (0.59)	.74	*.004*	*.01*
Vertical position of center of mass	–3.96 (0.83)	–2.05 (1.68)	1.90 (2.12)	*.02*	*.007*	.99
Top edge of the whole area	0.25 (0.62)	–3.98 (3.59)	2.37 (3.23)	*.004*	.19	.24
Bottom edge of the whole area	–7.64 (1.00)	–6.44 (1.97)	–5.59 (1.85)	.21	*.005*	.80
Top edge of the skeleton area	–1.14 (0.62)	0.89 (2.45)	1.10 (3.46)	*.03*	*.04*	.95
Bottom edge of the skeleton area	–7.64 (1.00)	–6.44 (1.97)	–5.59 (1.85)	*.008*	*.005*	>.99
Ratio of input	0.01 (0.02)	0.25 (0.26)	0.18 (0.27)	*.005*	.12	.61
Shape similarity, mean (SD)						
Cross-correlation maximum value	0.25 (0.02)	0.19 (0.05)	0.20 (0.04)	*.004*	*.002*	.86
Size-rescaled cross-correlation maximum value	0.14 (0.04)	0.11 (0.05)	0.09 (0.05)	.38	*.04*	.67

[a]NC: normal control.

[b]EOAD: early-onset Alzheimer disease.

[c]LOAD: late-onset Alzheimer disease.

[d]Italicization show significant differences at *P*<.05.

Number of Pen Strokes

The average number of strokes was 24 (SD 7) in NC individuals, 26 (SD 16) in patients with EOAD, and 23 (SD 4) in patients with LOAD. The total number of strokes did not differ across groups (Figure 2; Kruskal-Wallis test: *P*=.56).

Length of Pen Strokes

We used a union of 2 separate stroke sets with x- and y-projected lengths. The average length of the strokes was 4.58 (SD 0.71) cm for NCs, 3.99 (SD 1.48) cm for patients with EOAD, and 3.50 (SD 0.86) cm for patients with LOAD. The average length of strokes was significantly shorter in patients with LOAD than NC individuals (Figure 2; pairwise Mann-Whitney U test: *P*=.007). Extensive variation in the length of long strokes was noted in patients with EOAD.

When drawing figures, the length of strokes may be related to the constructional strategy. Longer strokes are more likely to be associated with a focus on global structure, whereas shorter lines are more likely to be associated with a focus on local features. Thus, the strokes were operationally defined as long or short using a 1-dimensional k-means++ algorithm (Figure 2). The average centroids of each long and short stroke cluster were 1.30 cm and 8.85 cm, respectively, for the x-projected length and 0.858 cm and 5.34 cm, respectively, for the y-projected length. The average ratios between the longest short stroke and the shortest long stroke were 2.20 (SD 0.60) cm for the x-projected length and 1.73 (SD 0.42) cm for the y-projected

length. The average number of long strokes was 14 (SD 4) for NCs, 18 (SD 12) for patients with EOAD, and 16 (SD 4) for patients with LOAD. The average number of short strokes was 8 (SD 1) for NCs, 9 (SD 7) for patients with EOAD, and 7 (SD 3) for patients with LOAD. The number of long or short strokes did not differ across the 3 groups (Figure 2; Kruskal-Wallis test: *P*=.58 for long strokes and *P*=.57 for short strokes). The first 5 strokes were defined as the beginning of the task. Most long strokes were drawn at the beginning of the task, but the ratio was significantly lower in patients with EOAD compared with that of NC individuals (Figure 2; pairwise Mann-Whitney U test: *P*=.04).

Speed of Pen Strokes

The longest line was used to calculate the drawing speed because it is most likely used to construct the skeleton of the figure, which is indicative of executive function. The average speed of the longest stroke was 6.31 (SD 1.86) cm/s for NCs, 4.41 (SD 2.76) cm/s for patients with EOAD, and 4.08 (SD 1.41) cm/s for patients with LOAD. Patients with LOAD were significantly slower in drawing longer lines compared with NC individuals (Figure 2; pairwise Mann-Whitney U test: *P*=.01).

When completing the task, we noticed that some patients with AD were hesitant in drawing global components. Thus, we first calculated the transition time for drawing a long line after a short line, followed by the elapsed time for the early long strokes. Transition time was defined as the time to initiate a long stroke after drawing a short stroke in all short-long sequences. The average pause time was 19.69 (SD 60.71) seconds for NCs, 14.03 (SD 26.67) seconds for patients with EOAD, and 25.35 (SD 83.81) seconds for patients with LOAD. Excluding the outlier over 60 seconds, the average pause time was 1.3 (SD 0.60) seconds for NCs, 6.23 (SD 7.54) seconds for patients with EOAD, and 2.95 (SD 1.83) seconds for patients with LOAD. Both patients with EOAD and LOAD took longer to initiate long strokes compared with NCs (Figure 2; pairwise Mann-Whitney U test: *P*=.045 for NC vs EOAD: *P*=.01 for NC vs LOAD), whereas the NC individuals showed a relatively consistent pause time. Second, the elapsed time was calculated for 5 early long strokes. Usually, it takes 5 strokes to draw the square and horizontal/vertical/diagonal lines inside the square. The average elapsed time of the 5 strokes was significantly longer for both patients with AD compared with the NC individuals (14.43, SD 4.62 seconds for NCs; 34.86, SD 19.39 seconds for patients with EOAD; and 23.65, SD 14.16 seconds for patients with LOAD). In addition, both patients with AD spent more time drawing global features compared with the NC participants (Figure 2; pairwise Mann-Whitney U test: *P*=.01 for NC vs EOAD: *P*=.04 for NC vs LOAD). None of the parameters stated above differed between patients with EOAD and LOAD.

Spatial Arrangement

Drawing Area

The working space or the drawing area may reflect the spatial arrangement abilities of the participants. We created a boundary that contained the whole drawing and defined it as the *whole area*, and another boundary that contained only the area inside the skeleton rectangle was defined as the *skeleton area* (Figure 3). The size and location of the whole area and skeletal area were analyzed for each group (Table 3). The average whole area was 84.17 (SD 10.88) cm^2 for the NCs, 108.51 (SD 58.81) cm^2 for patients with EOAD, and 84.01 (SD 33.93) cm^2 for patients with LOAD. The average skeleton area was 58.40 (SD 8.79) cm^2 for NCs, 56.34 (SD 20.78) cm^2 for patients with EOAD, and 55.29 (SD 18.93) cm^2 for patients with LOAD. The whole and skeletal areas did not differ among the 3 groups (Kruskal-Wallis test: *P*=.485 for the external boundary and *P*=.36 for the internal boundary). Significant differences were also not noted in the ratios between the whole and skeleton areas (Figure 3; Kruskal-Wallis test: *P*=.21).

Center of the Whole Area

Negative values represent a leftward or downward deviation from the center of the working space, whereas positive values represent a rightward or upward deviation from the center of the working space (Table 3). The average horizontal position of the center of the whole area was 0.42 (SD 0.31) cm for NCs, 0.25 (SD 0.67) cm for patients with EOAD, and –0.23 (SD 0.40) cm for patients with LOAD. For the horizontal position, patients with LOAD showed a significant left bias compared with NCs (Figure 3; pairwise Mann-Whitney U test: *P*<.001). The average vertical position of the center of the whole area was –3.70 (SD 0.69) cm for NCs, –1.22 (SD 1.84) cm for patients with EOAD, and –1.61 (SD 2.36) cm for patients with LOAD. For the vertical position, both AD groups showed a significant upward bias compared with NCs (Figure 3; pairwise Mann-Whitney U test: *P*=.005 for NC vs EOAD and *P*=.01 for NC vs LOAD). The position of the center of the skeleton area did not differ among the 3 groups.

Center of Mass of the Inputs

We analyzed the center of mass to determine whether the local features were skewed toward 1 side of the drawing (Figure 3). Again, negative numbers denoted leftward and downward deviations, whereas positive numbers indicated rightward and upward deviations from the center of the working space (Table 3). The average horizontal position of the center of mass was 0.28 (SD 0.34) cm for NCs, 0.46 (SD 0.64) cm for patients with EOAD, and –0.35 (SD 0.59) cm for patients with LOAD. With respect to the center of mass in the horizontal dimension, a significant left bias in patients with LOAD was observed compared with NCs and patients with EOAD (Figure 3; pairwise Mann-Whitney U test: *P*=.004 for LOAD vs NCs, *P*=.01 for LOAD vs EOAD). The average vertical position of the center of mass was –3.96 (SD 0.83) cm for NCs, –2.05 (SD 1.68) cm for patients with EOAD, and –1.90 (SD 2.12) cm for patients with LOAD. Compared with the NC individuals, the center of mass in the vertical direction was higher for both AD groups (Figure 3; pairwise Mann-Whitney U test: *P*=.02 for NC vs EOAD and *P*=.007 for NC vs LOAD).

Edge of the Whole and Skeleton Area

The invasion of the drawing into the perceptual space was assessed by analyzing the whole area of the top edge (Figure 3; Table 3). The average top edge of the whole area was 0.25 (SD 0.62) cm for NCs, –3.98 (SD 3.59) cm for patients with EOAD,

and 2.37 (SD 3.23) cm for patients with LOAD. Negative numbers denoted downward deviations, whereas positive numbers indicated upward deviations from the center of the working space. The patients with EOAD presented a significantly higher value of invasion compared with the NCs, whereas the difference between patients with LOAD and NC was not significant (Figure 3; pairwise Mann-Whitney U test: *P*=.004 for NC vs EOAD and *P*=.19 for NC vs LOAD). Regarding the bottom edge of the whole area, the patients with LOAD showed a significant upward bias compared with NCs, whereas the difference between patients with EOAD and NCs was not significant (Figure 3; pairwise Mann-Whitney U test: *P*=.21 for NC vs EOAD and *P*=.005 for NC vs LOAD).

The average top edge of the skeleton area was –1.14 (SD 0.62) cm for NCs, 0.89 (SD 2.45) cm for patients with EOAD, and 1.10 (SD 3.46) cm for patients with LOAD. Both AD groups showed a significantly higher value of invasion toward the perceptual space compared with that of NC individuals (Figure 3; pairwise Mann-Whitney U test: *P*=.03 for NC vs EOAD, *P*=.045 for NC vs LOAD). In addition, for the bottom edge of the skeleton area, both AD groups showed a significant upward bias compared with NC individuals (Figure 3; pairwise Mann-Whitney U test: *P*=.008 for NC vs EOAD and *P*=.005 for NC vs LOAD).

Closing-In Phenomenon

The closing-in phenomenon is the tendency to draw near or on the target when copying the image and is common in patients with AD or vascular dementia [19]. An upward deviation from the center and edge of the whole area may also represent the closing-in phenomenon. To achieve greater accuracy, we assessed *the ratio of input* in the perceptual space to measure the distance the drawing invaded into the perceptual space (Figure 3). Here, the term *the ratio of input* refers to the ratio of data points in the perceptual space to the total data points drawn by the participants. Figure 3 shows an example of the ratio of input in the perceptual space. Patients with EOAD (0.25, SD 0.26) showed a significant amount of invasion compared with NCs (0.01, SD 0.02), whereas the difference between patients with LOAD (0.18, SD 0.27) and NCs (0.01, SD 0.02) was not significant (Figure 3; pairwise Mann-Whitney U test: *P*=.005 for NC vs EOAD and *P*=.12 for NC vs LOAD; Table 3).

Shape Similarities Between the Target and Drawings Made by Participants

To scale the similarity between the original and copied figures, we applied a 2D cross-correlation analysis and calculated its coefficients and shifts in space. Although the conventional rating evaluates the accuracy of the shape and spatial arrangement of parts, the coefficient signifies the overall similarity between the 2 images (Table 3). To check the validity of this new similarity coefficient, we used Pearson correlation to compare the visual rating scores of the simplified RCFT and the similarity coefficients. According to the analysis, the coefficients of the shape similarity positively correlated to the visual rating (r=0.779; P<.001).

For both AD groups, the coefficient representing pattern matching was significantly lower than that of NC individuals (pairwise Mann-Whitney U test: *P*=.004 for EOAD vs NCs and *P*=.002 for LOAD vs NCs; Figure 4). However, there was no statistically significant difference between patients with EOAD and patients with LOAD (*P*=.86). Compared with the target stimulus, several participants drew images that were larger or smaller in size; thus, we normalized the copied figures and then performed a 2D cross-coefficient analysis. We found that the similarity was significantly lower in the patients with LOAD compared with that of the NC individuals even after rescaling the image sizes (*P*=.04). However, the significance of the difference between the patients with EOAD and the NC individuals disappeared after adjusting for image size (Figure 4).

Discussion

Principal Findings

This study introduced a novel analytic method that uses a digital pen and tablet to evaluate a completed drawing and movement kinetics involved in the drawing process to compare the visuoconstructional abilities of patients with AD and NC individuals. First, we created a simplified RCFT and compared the scores of the simplified Rey figure with those of the original to validate our simplified RCFT. We then compared the 2 AD groups with the NC individuals in terms of the simplified RCFT-derived digital metrics that were analyzed based on 3 aspects: (1) the number, speed, length, and sequence of pen strokes; (2) the spatial arrangement of the drawing; and (3) the similarity between the target picture and the drawing of the participants. Significant differences were noted between the NC and the 2 AD groups. Taken together, our findings suggest that movement kinetics in the drawing process and scores acquired from the finished drawings can serve as useful biomarkers to investigate visuoconstructional dysfunction in AD [20].

Validation of Simplified Rey-Osterrieth Complex Figure Test

We created a simplified RCFT to reduce drawing time and the number of strokes required to complete the picture. In the limited space, reducing the number of strokes makes it easier for the strokes to be identified by the digital device. Compared with the original, a high correlation in the simplified RCFT score was observed in both the whole and the separate groups (NC and the 2 AD groups; Figure 5; Table 2). In line with previous studies that reported strong correlations between simpler and original RCFT in normal adults [21] and patients with AD [22], our findings suggest that the simplified and original RCFTs may be comparable in evaluating visuoconstructional dysfunction in patients with AD. It was interesting to note that there were outliers among those with lower original RCFT scores: scores of 0 to 10 on the original yielded various scores of 1 to 14 on the simplified version (Figure 5). This could have occurred because patients with substantial visuospatial impairment may become overwhelmed when encountering a complex figure so that they eventually become less motivated to complete a figure. However, the same individuals would be relatively more willing to complete a figure if it was less

complicated. These findings may also have implications that our simplified figure may be more appropriate for individuals with mild to moderate than severe impairment of cognition or visuospatial abilities.

Analysis of Pen Stroke Data and Copying Sequences

Regarding the digital pen stroke data analysis, we first compared the 2 AD groups (EOAD and LOAD) with the NC individuals in terms of temporal appearance of long stokes. Overall, the AD groups used fewer long strokes than the NC group. In addition, both the AD groups drew long strokes later in the drawing process compared with that of the NC group. This suggested that the AD groups tended to draw local features earlier and used shorter strokes to finish a figure than the NC individuals. A previous study reported that more than half of the NC individuals first drew the global features of RCFT and then the details or local features subsequently [23]. Other studies suggest that piecemeal approaches in drawing are indicative of brain disease and that they also reflect the inability of an individual to process global information equivalent to that of NC individuals [18].

Additionally, our digital device enabled us to track the sequences of the drawing by color coding the strokes. In the pencil and paper version of the RCFT, neuropsychologists often track copying sequences by giving the patients a series of color pencils. For example, the red color pencil for the first 10 seconds and the yellow pencil for the next 10 seconds, and so on [2]. Similarly, in our digitized images, we assigned colors to the strokes in a certain order. We found that all NC individuals maintained a similar sequence of drawing, starting with the large box and then drawing the diagonal or orthogonal axes. Most patients who successfully copied all the elements also followed the sequence of the NC individuals. Unconventional approaches to drawing, such as drawing details before finishing the global structure or completing the global structure at the end, were identified from some of the patients. Such behavior, however, was not characterized as unimodal features but rather as a portrayal of an individual's drawing process. Color coding strokes following the completion of a test may be advantageous over using color pencils because switching color pencils will interrupt the drawing process and thus affect the overall memory processing of the participants.

The first and second methods employed in the pen stroke analysis, as described earlier, suggest that analysis using a digital device can allow quantification of the global versus the local nature of drawing in patients with AD. Regarding the anatomical substrates of global and local processing, it is known that the left hemisphere is more specialized in processing local rather than global features, and vice versa for the right hemisphere. This construct, however, is not consistent with our findings as typical AD involves the bilateral temporoparietal area in a symmetrical fashion. Rather, the *global element first and local element later* strategy may be related to the organizing abilities of individuals that can be mediated by prefrontal cortices, which is one of the brain regions affected in the early stage of AD. Another explanation could be the visual perception hypothesis. Previous studies have reported that the visual perception of normal individuals is a sequential process in which global features are perceived before local features. Another study also had patients with AD read Navon figures (large *global* digits composed of smaller *local* digits) and found that patients showed impairment in reading global figures [24]. Therefore, the preference for drawing local features from the AD groups in our study may indicate a mild form of simultanagnosia, which is part of the Balint syndrome.

Another feature we noticed from our pen stroke analysis was that not only the average length of strokes was significantly shorter in patients with LOAD compared with NC individuals, but the average speed of the longest stroke was also significantly slower in patients with LOAD compared with NC individuals (Figure 2). Age-related changes could have contributed to the results because the older the group was, the longer it took for the participants to complete a stroke (shortest to longest in time: EOAD<NC<LOAD). Statistically significant differences were only identified when comparing the NC and LOAD groups. However, we observed another movement kinetics finding that was inconsistent with our age account. Both patients with EOAD and patients with LOAD had longer pause and elapsed times to initiate long strokes compared with the time that NC individuals took (Figure 2). Therefore, decreased length/speed and increased transition and pause time might be related to Parkinsonism or psychomotor speed slowing in both groups of patients with AD. Unfortunately, we did not evaluate the Unified Parkinson's Disease Rating Scale (UPDRS) score. Therefore, although we excluded patients with Parkinsonism or degenerative disorders that accompany Parkinsonism, such as progressive supranuclear palsy or Lewy body dementia, while enrolling participants, we cannot eliminate the possibility of the Parkinsonism factored in our data.

Spatial Arrangement and Closing-In Phenomenon

On the basis of our analysis of the position of the drawing area, patients with AD tended to draw the figures closer toward the perception space containing the target figure. This upward deviation is indicative of the closing-in phenomenon. Previous studies have analyzed copied figures of the alternating square and triangle task and showed that as patients drew from left to right, the drawing approached the target, resulting in a sloping image. The angle of the slope was steeper in patients with AD and vascular dementia in comparison with that of NCs [25,26]. Other studies used the RCFT to analyze the closing-in phenomenon but relied on visual rating rather than quantification [27,28]. To the best of our knowledge, this is the first study to quantify the closing-in phenomenon using the RCFT. We analyzed the x and y coordinates of the drawing area and the input data ratio of the drawing to quantify the closing-in phenomenon (Figure 3). We have successfully demonstrated the difference between patients with AD and normal individuals. The underlying mechanisms of the closing-in phenomenon have not been fully elucidated. Previous studies have explained that the closing-in phenomenon is a part of constructional apraxia or primitive behaviors [26]. Another study analyzed the severity of closing-in as a function of figure complexity and suggested that the closing-in phenomenon is related to the patients' compensatory strategies to overcome visuospatial dysfunction or visuospatial working memory deficits [19,29,30]. The higher immediate recall scores of the original RCFT associated with

higher vertical positions of the simplified RCFT observed in this study also strongly support that the closing-in phenomenon is related to the visuospatial working memory deficit. Overall, the results of this study provide supporting evidence that using a digital device enables quantification of the closing-in phenomenon.

Shape Similarities

Our 2D cross-correlation analysis calculated a coefficient reflecting the overall similarity between the target picture and the drawings of the participants. The coefficient of both AD groups was significantly lower compared with that of the NCs, although the significance of the difference between the patients with EOAD and the NC individuals disappeared after adjusting the image size. Furthermore, there was a significant correlation between the similarity coefficients and the visual rating scores of the simplified RCFT. Therefore, automatic evaluation of similarity indices, which is not possible when using pen and paper, is another advantage of our analytic method that involves a digital device.

Early-Onset Alzheimer Disease Versus Late-Onset Alzheimer Disease

Contrary to our expectations, there was no significant difference between patients with LOAD and patients with EOAD, except that unlike patients with LOAD, patients with EOAD showed signs of leftward deviation when drawing the entire figure. Clinically, visuospatial dysfunction and apraxia are more prominent in patients with EOAD than in patients with LOAD [11,12]. A previous study also showed that 33% of patients with EOAD showed nonmemory symptoms, whereas only 6% of the patients with LOAD showed these symptoms [12]. These nonmemory symptoms included language and executive and visuoconstructional functions [31]. Another study that used RCFT to assess visuoconstructional abilities showed that the EOAD group was significantly more impaired than the LOAD group [11]. In contrast to these previous studies, there was no difference between the EOAD and the LOAD groups when analyzing various visuospatial parameters, except the horizontal position of the entire figure. The large variation in the EOAD group could have accounted for the lack of statistical significance.

Representing Cognitive Domains and Limitation

Thus far, we have addressed 3 aspects of digital metrics derived from the simplified RCFT. It is known that there are several components in human visuospatial function such as visuoperceptual, visuoconstructional, or visuospatial working memory abilities [30,32,33]. Therefore, it may be important to discuss which visuospatial abilities are related to our digital metrics. First, pen stroke data may be linked to motor skills and strategies for the construction of visual components [33]. Second, analysis of the position of the drawing area may be related to the closing-in phenomenon, which reflects visuoperceptual or visuospatial working memory deficits [30]. Third, a 2D cross-correlation analysis may be associated with general visuospatial abilities that include both perception and construction attributes of visuospatial function [32]. Additionally, the positive correlation observed between the immediate recall scores of the original RCFT and the coefficients of the shape similarity from our digital version also support the working memory deficit hypothesis.

One limitation of this study is the sample size. The relatively small sample size of the participants may restrict the validity and reliability of our results. However, to overcome this limitation, we enrolled participants who had undergone imaging and neuropsychological battery tests. The small sample size may also explain the lack of difference in visuoconstructional dysfunctions between patients with EOAD and patients with LOAD.

Conclusions

We performed an analysis using a simplified RCFT that encompassed not only the quantification of the final figure but also the drawing process, such as pen strokes, spatial arrangement, and similarities, by using digital drawing data. Analyzing the drawing sequences of global and local elements as well as the number, length, and speed of strokes may only be possible through digital devices. Furthermore, our digitized version enabled automatic quantification of the degree of similarity and position of the drawing with respect to the target picture. Therefore, our digital metrics can complement the conventional visual rating of RCFT, which has been administered via pen and paper. Future standardization studies involving many patients and NCs are warranted to investigate whether our digitized, simplified RCFT would be useful in clinical and research settings.

Acknowledgments

This work was supported by the National Research Council of Science & Technology grant by the Korean government (Ministry of Science, Information and Communications Technology and Future Planning, MSIP; No. CRC-15-04-KIST), by the National Research Foundation of Korea (2017R1A2B3012659), by the Korea Institute of Science and Technology Artificial Brain Convergence Project (2E30762), and by research funds for newly appointed professors of the Jeonbuk National University in 2019.

Authors' Contributions

KK and SL contributed equally (cofirst author). DN and JC contributed equally (cocorresponding author).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Pseudocolored drawings for normal controls.
[PNG File , 1139 KB - jmir_v22i8e18136_app1.png]

Multimedia Appendix 2
Pseudocolored drawings for patients with early-onset Alzheimer disease.
[PNG File , 1207 KB - jmir_v22i8e18136_app2.png]

Multimedia Appendix 3
Pseudocolored drawings for patients with late-onset Alzheimer disease.
[PNG File , 1642 KB - jmir_v22i8e18136_app3.png]

References

1. Cherrier MM, Mendez MF, Dave M, Perryman KM. Performance on the Rey-Osterrieth complex figure test in Alzheimer disease and vascular dementia. Neuropsychiatry Neuropsychol Behav Neurol 1999 Apr;12(2):95-101. [Medline: 10223256]
2. Lezak M. Neuropsychological Assessment. Fifth Edition. New York, USA: Oxford University Press; 2012.
3. Waber DP, Holmes JM. Assessing children's copy productions of the Rey-Osterrieth complex figure. J Clin Exp Neuropsychol 1985 Jun;7(3):264-280. [doi: 10.1080/01688638508401259] [Medline: 3998091]
4. Somerville J, Tremont G, Stern RA. The Boston qualitative scoring system as a measure of executive functioning in Rey-Osterrieth complex figure performance. J Clin Exp Neuropsychol 2000 Oct;22(5):613-621. [doi: 10.1076/1380-3395(200010)22:5;1-9;FT613] [Medline: 11094396]
5. Nazar H, Moetesum M, Ehsan S, Siddiqi I, Khurshid K, Vincent N. Classification of Graphomotor Impressions Using Convolutional Neural Networks: An Application to Automated Neuro-Psychological Screening Tests. In: Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition. 2017 Presented at: ICDAR'17; November 9-15, 2017; Kyoto, Japan. [doi: 10.1109/icdar.2017.78]
6. Müller S, Preische O, Heymann P, Elbing U, Laske C. Increased diagnostic accuracy of digital vs conventional clock drawing test for discrimination of patients in the early course of Alzheimer's disease from cognitively healthy individuals. Front Aging Neurosci 2017;9:101 [FREE Full text] [doi: 10.3389/fnagi.2017.00101] [Medline: 28443019]
7. Müller S, Herde L, Preische O, Zeller A, Heymann P, Robens S, et al. Diagnostic value of digital clock drawing test in comparison with CERAD neuropsychological battery total score for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. Sci Rep 2019 Mar 5;9(1):3543 [FREE Full text] [doi: 10.1038/s41598-019-40010-0] [Medline: 30837580]
8. Müller S, Preische O, Heymann P, Elbing U, Laske C. Diagnostic value of a tablet-based drawing task for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. J Alzheimers Dis 2017;55(4):1463-1469. [doi: 10.3233/JAD-160921] [Medline: 27858717]
9. Canham RO, Smith SL, Tyrrell AM. Automated Scoring of a Neuropsychological Test: The Rey Osterrieth Complex Figure. In: Proceedings of the 26th Euromicro Conference. 2000 Presented at: EUROMICRO'00; September 5-7, 2000; Maastricht, Netherlands. [doi: 10.1109/eurmic.2000.874519]
10. Melrose RJ, Harwood D, Khoo T, Mandelkern M, Sultzer DL. Association between cerebral metabolism and Rey-Osterrieth complex figure test performance in Alzheimer's disease. J Clin Exp Neuropsychol 2013;35(3):246-258 [FREE Full text] [doi: 10.1080/13803395.2012.763113] [Medline: 23387510]
11. Joubert S, Gour N, Guedj E, Didic M, Guériot C, Koric L, et al. Early-onset and late-onset Alzheimer's disease are associated with distinct patterns of memory impairment. Cortex 2016 Jan;74:217-232. [doi: 10.1016/j.cortex.2015.10.014] [Medline: 26694580]
12. Koedam EL, Lauffer V, van der Vlies AE, van der Flier WM, Scheltens P, Pijnenburg YA. Early-versus late-onset Alzheimer's disease: more than age alone. J Alzheimers Dis 2010;19(4):1401-1408. [doi: 10.3233/JAD-2010-1337] [Medline: 20061618]
13. Smits LL, Pijnenburg YA, Koedam EL, van der Vlies AE, Reuling IE, Koene T, et al. Early onset Alzheimer's disease is associated with a distinct neuropsychological profile. J Alzheimers Dis 2012;30(1):101-108. [doi: 10.3233/JAD-2012-111934] [Medline: 22366769]
14. Kang SH, Park YH, Lee D, Kim JP, Chin J, Ahn Y, et al. The cortical neuroanatomy related to specific neuropsychological deficits in Alzheimer's continuum. Dement Neurocogn Disord 2019 Sep;18(3):77-95 [FREE Full text] [doi: 10.12779/dnd.2019.18.3.77] [Medline: 31681443]
15. Kang Y, Jahng S, Na DL. Seoul Neuropsychological Screening Battery, Second Edition. Seoul, South Korea: Human Brain Research & Consulting Co; 2012.

16. Kang Y, Na D, Hahn S. A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients. J Korean Neurol Assoc 1997;15(2):300-308 [FREE Full text]
17. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):263-269 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.005] [Medline: 21514250]
18. Meyers J, Meyers K. Rey Complex Figure Test and Recognition Trial. Odessa, Ukraine: Psychological Assessment Resources; 1995.
19. Lee BH, Chin J, Kang SJ, Kim E, Park KC, Na DL. Mechanism of the closing-in phenomenon in a figure copying task in Alzheimer's disease patients. Neurocase 2004 Oct;10(5):393-397. [doi: 10.1080/13554790490892194] [Medline: 15788278]
20. Piau A, Wild K, Mattek N, Kaye J. Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild Alzheimer disease and implications for clinical care: systematic review. J Med Internet Res 2019 Aug 30;21(8):e12785 [FREE Full text] [doi: 10.2196/12785] [Medline: 31471958]
21. Casarotti A, Papagno C, Zarino B. Modified Taylor complex figure: normative data from 290 adults. J Neuropsychol 2014 Sep;8(2):186-198. [doi: 10.1111/jnp.12019] [Medline: 23647550]
22. de Paula JJ, Costa MV, de Andrade GD, Ávila RT, Malloy-Diniz LF. Validity and reliability of a 'simplified' version of the Taylor complex figure test for the assessment of older adults with low formal education. Dement Neuropsychol 2016;10(1):52-57 [FREE Full text] [doi: 10.1590/S1980-57642016DN10100010] [Medline: 29213432]
23. Wilson N, Batchelor J. Examining Rey complex figure test organization in healthy adults. J Clin Exp Neuropsychol 2015;37(10):1052-1061. [doi: 10.1080/13803395.2015.1075473] [Medline: 26325340]
24. Slavin MJ, Mattingley JB, Bradshaw JL, Storey E. Local-global processing in Alzheimer's disease: an examination of interference, inhibition and priming. Neuropsychologia 2002;40(8):1173-1186. [doi: 10.1016/s0028-3932(01)00225-1] [Medline: 11931921]
25. Kwak YT. 'Closing-in' phenomenon in Alzheimer's disease and subcortical vascular dementia. BMC Neurol 2004 Jan 26;4:3 [FREE Full text] [doi: 10.1186/1471-2377-4-3] [Medline: 14741055]
26. Chin J, Lee BH, Seo SW, Kim E, Suh MK, Kang SJ, et al. The closing-in phenomenon in Alzheimer's disease and vascular dementia. J Clin Neurol 2005 Oct;1(2):166-173 [FREE Full text] [doi: 10.3988/jcn.2005.1.2.166] [Medline: 20396464]
27. de Lucia N, Grossi D, Fasanaro AM, Carpi S, Trojano L. Frontal defects contribute to the genesis of closing-in in Alzheimer's disease patients. J Int Neuropsychol Soc 2013 Aug;19(7):802-808. [doi: 10.1017/S1355617713000568] [Medline: 23701672]
28. Grossi D, de Lucia N, Trojano L. Closing-in is related to apathy in Alzheimer's disease patients. J Alzheimers Dis 2015;43(3):849-855. [doi: 10.3233/JAD-141257] [Medline: 25125462]
29. Serra L, Fadda L, Perri R, Caltagirone C, Carlesimo GA. The closing-in phenomenon in the drawing performance of Alzheimer's disease patients: a compensation account. Cortex 2010 Sep;46(8):1031-1036. [doi: 10.1016/j.cortex.2009.08.010] [Medline: 19782972]
30. Ambron E, Della Sala S. A critical review of closing-in. Neuropsychology 2017 Jan;31(1):105-117. [doi: 10.1037/neu0000295] [Medline: 27442452]
31. van der Flier WM, Pijnenburg YA, Fox NC, Scheltens P. Early-onset versus late-onset Alzheimer's disease: the case of the missing APOE 4 allele. Lancet Neurol 2011 Mar;10(3):280-288. [doi: 10.1016/S1474-4422(10)70306-9] [Medline: 21185234]
32. Bauer RM. Visuospatial, visuoperceptual, visuoconstructional disorders. In: Parsons MW, Hammeke TA, editors. Clinical Neuropsychology: A Pocket Handbook for Assessment. Washington, DC: American Psychological Association; 2014.
33. Stevens A, Bernier R. Visual-motor function. In: Volkmar FR, editor. Encyclopedia of Autism Spectrum Disorders. New York, USA: Springer; 2013.

Abbreviations

AD: Alzheimer disease
EOAD: early-onset Alzheimer disease
LOAD: late-onset Alzheimer disease
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
MRI: magnetic resonance imaging
NC: normal control
RCFT: Rey-Osterrieth Complex Figure Test
SNSB: Seoul Neuropsychological Screening Battery

Edited by G Eysenbach; submitted 05.02.20; peer-reviewed by Y Jeong, K Wild; comments to author 14.03.20; revised version received 07.05.20; accepted 13.05.20; published 12.08.20.

Please cite as:
Kim KW, Lee SY, Choi J, Chin J, Lee BH, Na DL, Choi JH
A Comprehensive Evaluation of the Process of Copying a Complex Figure in Early- and Late-Onset Alzheimer Disease: A Quantitative Analysis of Digital Pen Data
J Med Internet Res 2020;22(8):e18136
URL: https://www.jmir.org/2020/8/e18136
doi:10.2196/18136
PMID:32491988

Original Paper

Use of Eye-Tracking Technology by Medical Students Taking the Objective Structured Clinical Examination: Descriptive Study

M D Grima-Murcia[1*], PhD; Francisco Sanchez-Ferrer[1*], PhD; Jose Manuel Ramos-Rincón[1], PhD; Eduardo Fernández[1], PhD

Facultad de Medicina, University Miguel Hernández, Elche, Spain

[*]these authors contributed equally

Corresponding Author:
Francisco Sanchez-Ferrer, PhD
Facultad de Medicina
University Miguel Hernández
Edificio de Departamentos, Pediatría
Avinguda de la Universitat d'Elx
Elche, 032
Spain
Phone: 34 965169538
Email: pacosanchezferrer0@hotmail.com

Abstract

Background: The objective structured clinical examination (OSCE) is a test used throughout Spain to evaluate the clinical competencies, decision making, problem solving, and other skills of sixth-year medical students.

Objective: The main goal of this study is to explore the possible applications and utility of portable eye-tracking systems in the setting of the OSCE, particularly questions associated with attention and engagement.

Methods: We used a portable Tobii Glasses 2 eye tracker, which allows real-time monitoring of where the students were looking and records the voice and ambient sounds. We then performed a qualitative and a quantitative analysis of the fields of vision and gaze points attracting attention as well as the visual itinerary.

Results: Eye-tracking technology was used in the OSCE with no major issues. This portable system was of the greatest value in the patient simulators and mannequin stations, where interaction with the simulated patient or areas of interest in the mannequin can be quantified. This technology proved useful to better identify the areas of interest in the medical images provided.

Conclusions: Portable eye trackers offer the opportunity to improve the objective evaluation of candidates and the self-evaluation of the stations used as well as medical simulations by examiners. We suggest that this technology has enough resolution to identify where a student is looking at and could be useful for developing new approaches for evaluating specific aspects of clinical competencies.

(J Med Internet Res 2020;22(8):e17719) doi:10.2196/17719

KEYWORDS

visual perception; medical education; eye tracking; objective structured clinical examination; medical evaluation

Introduction

The objective structured clinical examination (OSCE) is an evaluation of medical students that aims to assess candidates' skills and attitudes in certain clinical situations. This goal is different from that of typical written exams, which primarily evaluate knowledge. Basically, evaluating clinical competencies entails objective measurement of whether a candidate has correctly used and applied theoretical knowledge [1]. The exam is used not only for undergraduate students but also in the graduate context [2,3] and in different medical specialties [4,5]. These tests have been in use since the 1970s and have proven to be a reliable and valid tool for evaluation, even when using the same stations [6]. The OSCE is particularly important for evaluating and teaching semiology skills and training the clinical eye of students [7].

The OSCE incorporates diverse evaluation instruments, which are arranged in successive stations that simulate real clinical situations [8]. The strength of this format lies in the mixed-methods evaluation, which allows examiners to explore

three of the four levels of Miller's pyramid: to know, to know how, and to show how [3]. The number of stations ranges from 5 to 20, according to the objectives being evaluated [9].

In Spain, various medical schools have not implemented this testing method uniformly, with universities in Catalonia—as with many medical schools worldwide [10]—gaining the most experience in the more than 20 years since they began using the OSCE [11].

Recently, the National Deans Conference in Spain approved a 20-station OSCE that all sixth-year medical students must take. The general characteristics of the test are practical character; oriented toward evaluating the professional competencies of the candidate relative to the specific competencies of the medical degree, established by the Order ECI/332/2008 (published in the Official State Gazette on February 15, 2008); performed through the resolution of clinical cases; and having the objective of demonstrating clinical skills.

The OSCE explores diverse areas of evaluation through a range of methodologies. The stations include simulated and standardized patients, mannequins, short-answer questions, performance of complementary examinations adjusted to the particular case, clinical report writing, structured oral examinations (SOEs), skills and procedures, computer functions, and simulators. The OSCE entails a considerable collective effort both for candidates and examiners; however, several studies have also reported good cost-effectiveness [12] compared to traditional testing [9].

Video recording for clinical training purposes is already used successfully, including to evaluate procedures, techniques, or diagnostic tests [13,14]. However, the technique proposed in this study is different from these recordings and has substantial implications for teachers [15]. Using an eye tracker, we can view and analyze candidates through their own unique perspective, similarly to how this technology is used for training in gastroscopy or locoregional anesthesia, alone or in conjunction with conventional video [16], or in the field of radiology [17,18]. Consequently, this device makes it possible to perform a qualitative analysis of the test (also allowing examiners to self-evaluate their station) and to objectively analyze those elements that, due to the test design, may introduce subjectivity in the examination, such as the relationship with different speakers or the dynamics of evaluating complementary medical tests [19].

With this in mind, we designed a feasibility study to assess the real possibilities of eye tracking as a tool for teaching and evaluation in the OSCE, according to each station model. To the best of our knowledge, this is the first study to use eye-tracking glasses in each of the 20 OSCE stations. Our main goal is to identify whether this technology can be used in an examination as extensive as the OSCE. In addition, we analyzed the opportunities for evaluating students according to the type of station, the usefulness perceived by the teachers, and the teaching opportunities for students. The hypothesis was that this technology is useful in teaching, specifically in the OSCE.

Methods

This is a descriptive study on the use of eye-tracking glasses during the OSCE. We carried out an OSCE of 117 sixth-year medical students at the Medical School of Miguel Hernández University Spain in June 2017. The OSCE consisted of a circuit of 20 stations or situations, and the candidates had to move consecutively through all of them, spending 9 minutes on each, with 2-minute rest periods between each new station. The stations and the skill areas included were history taking, physical examination, doctor-patient communication, clinical report writing, clinical judgement, technical skills, preventive activities, and ethical-legal issues (Table 1). The students had completed a shorter OSCE in the third year.

The exam was carried out in parallel rounds of 23 candidates each. All rounds were conducted on the same day. Two rounds were held in the morning and three rounds in the afternoon (one with 25 students). Forty consultation areas were prepared in line with the needs of the specific stations, each equipped with a computer program created for candidate evaluation. The simulated patient or examiner completed a checklist of evaluation items for the skill area being assessed. Exams were continuous, with rotations communicated through a computer program with speakers. The exams lasted 4.5 hours. Candidates were not permitted to have any electronic items with them (including mobile phones or watches), and all had a white coat, stethoscope, pencil or pen, and two blank sheets of paper.

The portable eye trackers used were the Tobii Glasses 2 [13,16] (Figure 1). The system was calibrated upon start up for each wearer. This device weighs 45 g and has a 160° field of vision, with minimal vision loss with extreme eye movements. Both image and sound were recorded simultaneously. The recordings were saved on memory cards for subsequent analysis and transmitted through a wireless network to a projector or computer for real-time visualization by the teachers (recording delay of 1 second via Tobii Glasses Controller). Some of the research questions were can users recognize the purpose of each specific station; if not, what are the obstacles; what was the main focus of attention in each station; were the medical images provided easy to read and understand?

The videos for each station were edited to eliminate the waiting periods and to include only the time period from which the participants read the case being evaluated prior to entering the consultation area. The audio recordings at each station were also homogenized so that the volume levels were approximately the same for all stations.

Table 1. Station, type of station, and skills map for the objective structured clinical examination.

Station number	Station	Type	HT[a]	CE[b]	TSP[c]	CS[d]	CJ[e]	DP[f]	IR[g]	EL[h]	Total
1	MSP[i] of the digestive tract	SP[j]	65.5	24.1	0.0	10.3	0.0	0.0	0.0	0.0	100
2	MSP of the digestive tract	Report	0.0	0.0	0.0	0.0	34.8	65.3	0.0	0.0	100
3	Infectious disease pathology	SOE[k]	0.0	0.0	0.0	0.0	0.0	0.0	90.0	10.0	100
4	Emergency medicine	M-P[l]	0.0	35.0	55.0	0.0	0.0	0.0	0.0	10.0	100
5	Legal medicine	Report	0.0	0.0	0.0	0.0	0.0	38.1	61.9	0.0	100
6	Emergency	SP	0.0	96.8	0.0	3.2	0.0	0.0	0.0	0.0	100
7	MSP of the nephrourinary system	M-P	0.0	0.0	55.0	0.0	35.0	0.0	0.0	10.0	100
8	Psychiatry	SP	22.6	35.5	0.0	12.9	29.0	0.0	0.0	0.0	100
9	MSP of the musculoskeletal system	M-P	0.0	57.2	28.6	0.0	14.3	0.0	0.0	0.0	100
10	Pediatrics	SOE	0.0	0.0	0.0	0.0	42.1	0.0	47.4	10.5	100
11	Infectious disease pathology	SP	43.3	13.3	0.0	10	33.4	0.0	0.0	0.0	100
12	MSP of the endocrine system	SOE	20	0	30	0	10	0	40	0	100
13	Surgery	M-P	0.0	0.0	83.3	0.0	0.0	0.0	12.5	4.2	100
14	Gynecology and obstetrics	SP	50.0	0.0	0.0	15.0	25	0.0	10	0.0	100
15	Gynecology and obstetrics	M-P	0.0	15.8	26.3	0.0	47.4	0.0	10.5	0.0	100
16	Pediatrics	SP	38.5	30.8	0.0	11.5	11.5	0.0	7.7	0.0	100
17	MSP of the respiratory system	Report	0	0	0	0	70	30	0	0	100
18	MSP of the cardiovascular system	SP	44	36	0	12	8	0	0	0	100
19	MSP of the cardiovascular system	Report	0	0	0	0	100	0	0	0	100
20	Microbiology and legal medicine	Report	0.0	0.0	0.0	0.0	38.9	0.0	22.2	38.9	100

[a]HT: history taking.

[b]CE: clinical examination.

[c]TSP: technical skills and procedures.

[d]CS: communication skills.

[e]CJ: clinical judgement, diagnostic and therapeutic management plan.

[f]DP: disease prevention and health promotion.

[g]IR: interprofessional relationships.

[h]EL: ethical issues—legality and professionalism.

[i]MSP: medical and surgical pathology.

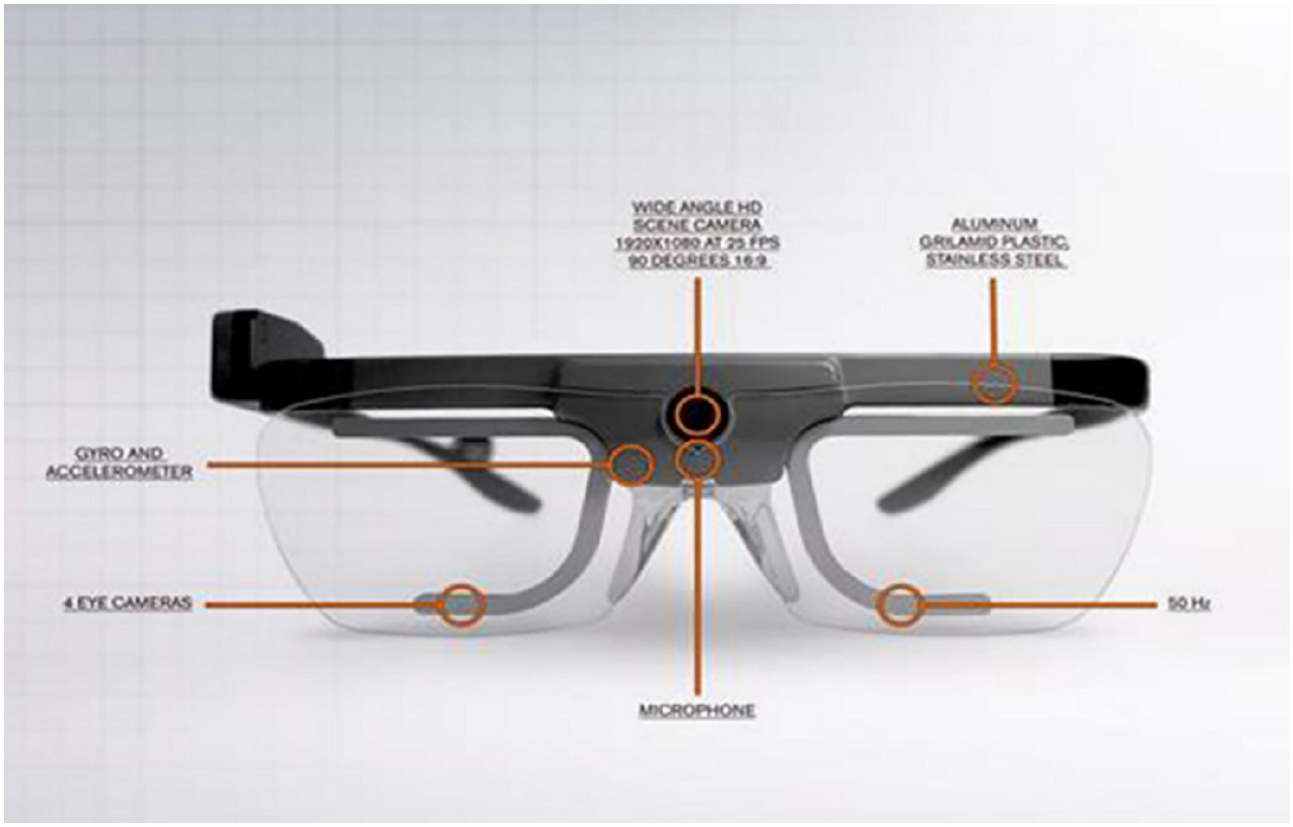
[j]SP: standardized patients.

[k]SOE: structured oral examination.

[l]M-P: mannequin-procedure.

The analysis was performed using Tobii Glasses Analysis Software and included a quantitative study of gaze points and the order that they appeared for each subject. We also used heat maps for visualization of eye tracking. Thus, we identified the areas of interest in each station and the percentage of the total area that they occupied within the image as a whole. Using these percentages, we calculated the participants' most frequent gaze locations, the number of gazes, and the visual itinerary (ie, the order that the gazes occurred). We also measured pupil dilation, but we found that it was extremely difficult to control luminance and maintain stable lighting conditions in our 20 different experimental conditions. Hence, we discovered that this particular measure was not appropriate for this type of study. Results were exported to Excel (Microsoft Corp) for statistical analysis.

Figure 1. Portable eye tracker used.



Finally, each of the station examiners (N=30) viewed the videos and completed a questionnaire to assess the usefulness of the eye tracker for each type of station in the visualization and interpretation of complementary tests, the assessment of the attitude of simulated patients and examiners, the evaluation of the candidate, the characterization of empathy and eye contact, the assessment of the design of the physical space where the exam was performed, the stimulation of ideas on training possibilities for examiners, and the external evaluation of the exam. Each of these questions was scored according to the usefulness of the device as follows: 0 not at all useful, 1 a little useful, 2 somewhat useful, and 3 very useful.

The Research Ethics Committee of Miguel Hernandez University approved this study (DMC.JRR.01.17). Verbal consent was obtained from the study participants.

Results

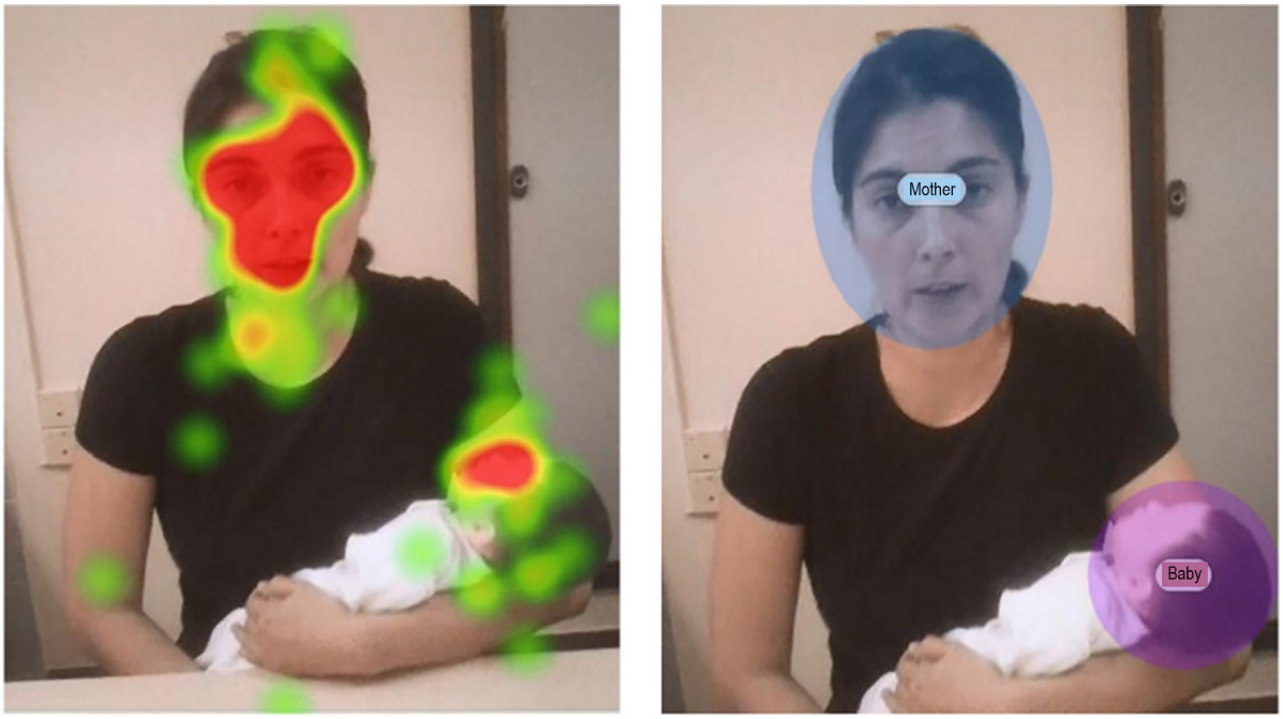
The OSCE consisted of 7 stations with simulated patients, 5 testing technical skills and procedures with or without mannequins, 5 testing candidates' abilities to draft a clinical report, and 3 SOEs on practical clinical situations. Of the 20 stations, we obtained useful recordings of 16 (80%). No recordings were made for 3 of the stations because the specific tasks involved writing reports (stations 5, 19, and 20). Furthermore, we encountered technical problems in the recordings associated with station 7 (possibly due to the battery); therefore, we discarded these data.

Eye Tracking in Stations With Simulated Patients

We evaluated 7 stations with simulated patients (stations 1, 6, 8, 11, 14, 16, and 18). The mean time for history taking was 242 (SD 28.4) seconds. In this group of stations, the eye-tracking device provided a wealth of usable data, demonstrating the extent to which the candidate *connected* on a visual level with the simulated patient. In this case, the candidate looked mainly at the patient's eyes and mouth (83.3% of the time).

In station 16, this pattern was of special interest, as it involved a simulated mother with a pediatric mannequin. The areas of visual interest were the faces of the mother and the infant (Figure 2). Of the entire image, the area of the mother's face was 12.1% of the total, while the infant's face corresponded to 5.6%. Figure 2 shows a heat map of the areas of interest (in blue and purple) for a given participant. The quantitative result for the time spent looking at both areas was 84.4%: 76.5% for the mother and 7.8% for the mannequin. Thus, of the 220 total gazes, 173 (86.5%) were focused on the mother's face, 25 (11.3 %) on the infant's face, and 22 (10.0%) on other locations. Moreover, the length of the gazes on other locations was shorter than for the two main areas of interest.

Figure 2. The areas of visual interest were the faces of the mother and the infant.



Eye Tracking in Technical Skills With or Without Mannequins

We evaluated 5 stations (1, 4, 7, 9, and 15) associated with the assessment of technical skills. Given that the students were constantly moving in these stations, for example, in the station associated with cardiopulmonary resuscitation (CPR) or in the station to evaluate suturing skills, it was difficult to analyze the results. Analysis of these data, therefore, required manual coding of the video to quantify the steadiness of the gaze [19], which was overly complex and outside of this research's scope. The subjective evaluation allowed us to observe the order used by the student in CPR and how, before each action, the student's gaze was fixed on the next task. We also identified the elements of the station that were not visualized by the student.

Eye Tracking in Image-Based Clinical Reports

In drafting clinical reports, for example, on electrocardiograms, x-rays, or other images, the eye tracker was not useful, as it only showed the candidate writing. However, in the stations evaluating a diagnostic test (for example, a chest x-ray), we were able to measure the time the candidates spent evaluating the image and the potential relationships between the focus of attention and the student's interpretation of the clinical case.

Figure 3 shows a representative example for station 19, dealing with a clinical report based on an image prompt (chest x-ray). The time spent by the medical student looking at the chest x-ray for this particular subject was 27.61 seconds. The area of interest, the hila, comprised 15.9% of the total area of the image, and the student spent 34% (9.5 seconds) of the time on this particular area (the initial gazes focused on the upper left zone). Furthermore, of the 59 total gazes, 26 were on the hila, and these gazes were also the longest.

Figure 3. The time spent and location of the gaze by the medical student looking at chest x-ray.

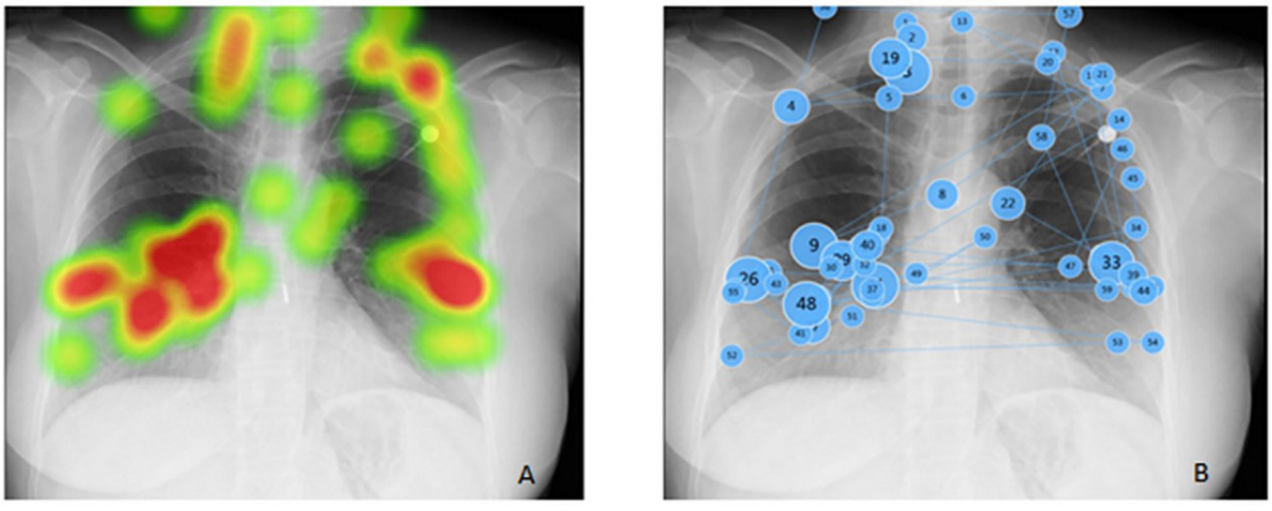
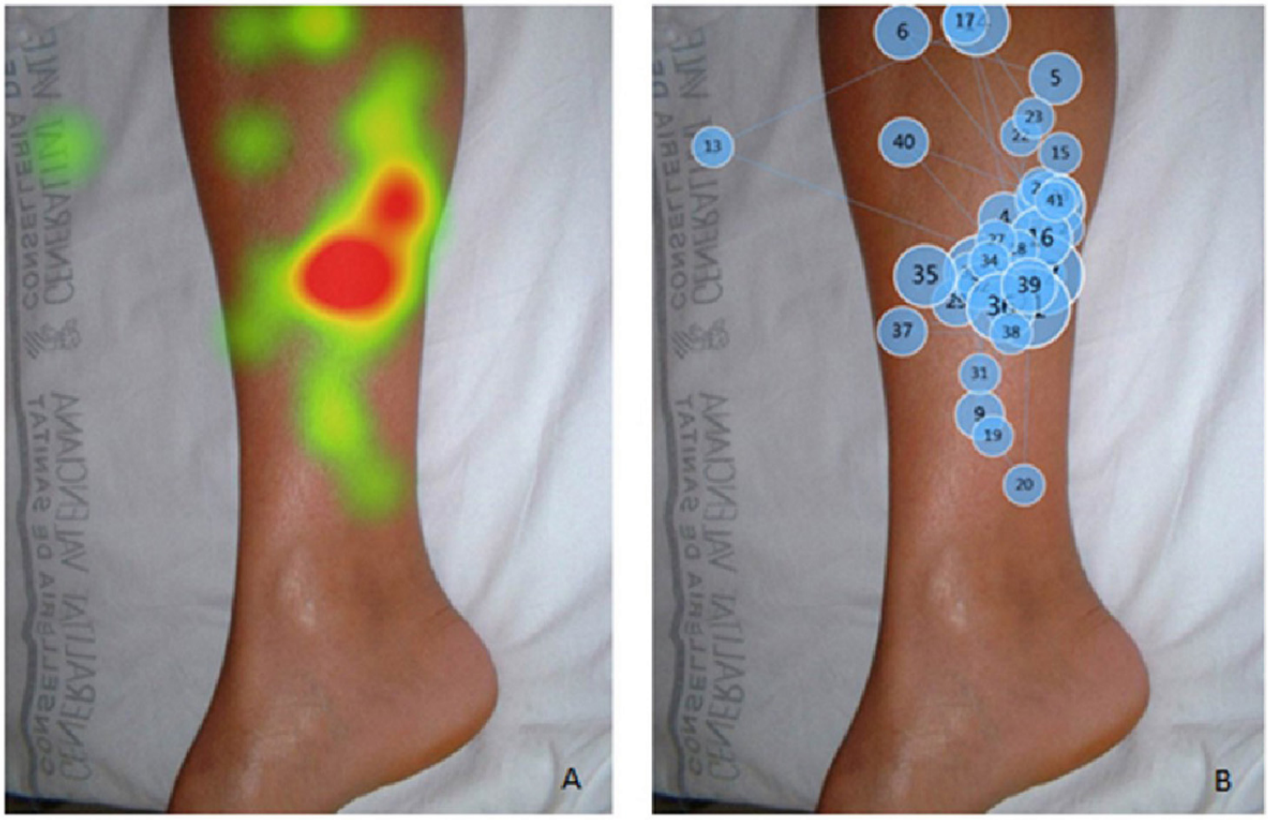


Figure 4 illustrates a different finding for station 11, which involved a simulated patient and an image of skin lesions on a lower limb. The mean time looking at this image was 17.8 seconds. In this case, we found no apparent areas of interest or systematic approach in the order of the gazes. Eye tracking was also useful for assessing the elements of the station not looked at by the candidates.

Figure 4. The time spent and location of gaze by the medical student looking at an image of skin lesions on a lower limb.



Evaluating the Usefulness of Eye Trackers According to the Station

Table 2 shows the mean subjective perceptions of the examiners regarding the potential value of a portable eye tracker in the evaluation of OSCEs. The usefulness of eye tracking appeared to be virtually nonexistent in all domains for the reports, while it was *somewhat useful* for the SOE, the evaluation of the simulated patient and examiner, and the preparation of the candidates. We did not find it useful for other areas of interest.

Table 2. Usefulness of portable eye-tracking devices in the OSCE by type of station for 30 station examiners.

Opportunities for eye tracking	Type of station, mode[a]			
	Standardized patients	Mannequin/procedure	Report	SOE[b]
Evaluation of complementary tests	0	1	2	1
Evaluation of simulated patient/examiner	3	3	0	2
Re-evaluation of candidate	2	3	0	1
Empathy/eye contact	3	1	0	1
Design of OSCE[c] consultation area	3	3	0	0
Candidate preparation for OSCE	2	3	1	2
External evaluation of OSCE	3	3	0	2

[a]Likert scale: 0 not at all useful, 1 a little useful, 2 somewhat useful, 3 very useful.

[b]SOE: structured oral examination.

[c]OSCE: objective structured clinical examination.

In contrast, the eye tracker was *very useful* for stations with standardized patient and mannequin procedures. This was true in practically every domain (with the exception of the evaluation of complementary tests and in the mannequin procedures to measure empathy or eye contact). Furthermore, portable eye tracking appeared to be of particular interest in the re-evaluation

of the candidates and in candidate preparation for the mannequin procedures. In addition, these videos have subsequently been used in teaching to provide future students of the test an insight into OSCE assessments.

Discussion

Our study shows that portable eye-tracking is an applicable and useful tool in the OSCE. In the standardized patient stations, mannequins or pictograms of the procedure were useful in many aspects. Herein is where teachers see more possibilities for their use. In contrast, in the stations where the students had to write reports, the eye tracking was not particularly useful. In addition, since this was the first time this technology was implemented in a complete OSCE, we found that preparation of batteries sufficient for the entire test is essential (recording of one station was lost due to battery issues).

OSCE testing has now been implemented in all the medical schools in Spain, with some programs dating back more than 20 years [11]. The effectiveness of the OSCE has also been demonstrated in various studies [6,9,12]; although, the exams being evaluated are heterogeneous in terms of design, number of stations, length, appropriateness of the simulated patients, the number of elements for assessment in each station, and the scales used to grade candidates' behavior or attitudes [6,16]. Students value more practical teaching, which includes the OSCE [20].

The use of portable eye-tracking technology has been introduced in medicine more recently [13,15,16,21-24], and many unanswered questions remain, particularly in the field of medical teaching and evaluation. In this context, research using eye tracking is widespread, but research on the use of eye tracking for assessing student performance is scarce. To our knowledge, this is the first study to be undertaken in this field.

Our preliminary results show a wide range of possibilities for future research. The technology may help evaluators to objectively measure the empathy shown by the candidate in the stations with simulated patients by characterizing the features of the candidate's gaze (eg, establishment of eye contact or steadiness). Although the OSCE is intended as an objective test, some elements are assessed with a certain degree of subjectivity, such as the scores associated with the candidate's treatment of patients, nonverbal communication, and conversation and empathy [25,26]. The eye tracker with recording represents an improvement over external video recording in that it enables measurement of the amount of time the candidate maintains eye contact with the patient, providing a novel and objective way to evaluate this item. Furthermore, the video is highly useful to professors, allowing them to evaluate the performance of the simulated patient (adherence to the script) and to obtain a comprehensive vision of the station from the candidate's perspective. This functionality also allows teaching staff to visualize which elements are used in the station, as some are not seen by the candidates due to the object's location.

Until now, in the stations assessing candidates' interpretation of image prompts (electrocardiogram, x-ray, photos, etc), we could not know whether candidates responded correctly to the questions as a result of an adequate systematic approach that included a revision of all relevant points or whether their responses were due to chance or previous knowledge of the answer [27]. However, using the eye tracker and the subsequent analysis, we have a better idea of the approach candidates use to reach their conclusion and the basis of their response in the image provided. In this case, studies on eye trackers have evaluated candidates' interpretation of an electrocardiogram [24] or other medical images [22,23]. However, further research should bear in mind that gaze does not necessarily have a direct correlation with a complex cognitive process.

The usefulness of eye tracking varied according to the station, whether these dealt with image-based clinical reports, images shown to the candidate, simulated patients (as a method to quantify doctor-patient empathy), or procedures or mannequins in the re-evaluation of the candidate. Thus, in the stations using images, we believe that it may be more practical to use optical tracking by means of a device placed directly on the monitor showing the image. As in other studies [17,18], this would allow a more accurate and straightforward quantitative analysis, in which more participants could take part.

On the other hand, the OSCE is stressful for medical students. To help them to prepare this evaluation, universities have created a study guide [28]. In our case, the videos obtained from the eye-tracking glasses were helpful to prepare the OSCE for future students.

We are fully aware that this is a proof-of-concept study and has important limitations. Furthermore, we were not able to compare the results of several students, as eye tracking is an expensive technology, and we only had a few devices available. Nevertheless, our preliminary results suggest that portable eye-tracking devices offer a number of opportunities in the field of OSCE evaluation, including the design, set up, and self-evaluation of the examiners at each station. Moreover, these devices may also provide insight into methods to improve the evaluation of candidates during the exams, particularly in the stations assessing subjective elements. Nevertheless, further studies are still needed.

Acknowledgments

The authors are grateful for the collaboration of the people who participated in the study. The authors disclosed no receipt of any financial support for the research. The data sets used during or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MDGM contributed to the design, data acquisition, data analysis, and interpretation; critically revised the manuscript; gave final approval; and agrees to be held accountable for all aspects of the work, ensuring integrity and accuracy. FSF contributed to the design, data acquisition, data analysis, and interpretation; critically revised the manuscript; gave final approval; and agrees to be held accountable for all aspects of the work, ensuring integrity and accuracy. EF contributed to the design, data acquisition, data analysis, and interpretation; critically revised the manuscript; gave final approval; and agrees to be held accountable for all aspects of the work, ensuring integrity and accuracy. JMRR contributed to the conception, design, and interpretation; drafted and critically revised the manuscript; gave final approval; and agrees to be held accountable for all aspects of the work, ensuring integrity and accuracy.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sample video captured during the investigation.
[MP4 File (MP4 Video), 43770 KB - jmir_v22i8e17719_app1.mp4]

References

1. Regehr G, Freeman R, Robb A, Missiha N, Heisey R. OSCE performance evaluations made by standardized patients: comparing checklist and global rating scores. Acad Med 1999 Oct;74(10 Suppl):S135-S137. [doi: 10.1097/00001888-199910000-00064] [Medline: 10536618]
2. Hamui M, Ferreira JP, Paganini A, Torres F, Ossorio MF, Yulitta H, et al. [The use of the objective structured clinical examination for paediatric residents in the City of Buenos Aires]. Rev Chil Pediatr 2016;87(4):274-278 [FREE Full text] [doi: 10.1016/j.rchipe.2016.02.003] [Medline: 26987274]
3. Dornan T, Bundy C. What can experience add to early medical education? Consensus survey. BMJ 2004 Oct 09;329(7470):834 [FREE Full text] [doi: 10.1136/bmj.329.7470.834] [Medline: 15472265]
4. Ellman MS, Putnam A, Green M, Pfeiffer C, Bia M. Demonstrating medical student competency in palliative care: development and evaluation of a new objective structured clinical examination station. J Palliat Med 2016 Jul;19(7):706-711. [doi: 10.1089/jpm.2015.0462] [Medline: 27249323]
5. Mangold KA, Jeffers JM, Burns RA, Trainor JL, Unti SM, Eppich W, et al. An objective structured clinical examination to improve formative assessment for senior pediatrics residents. J Grad Med Educ 2015 Sep;7(3):470-474 [FREE Full text] [doi: 10.4300/JGME-D-14-00777.1] [Medline: 26457159]
6. Baig LA, Violato C. Temporal stability of objective structured clinical exams: a longitudinal study employing item response theory. BMC Med Educ 2012 Dec 07;12:121 [FREE Full text] [doi: 10.1186/1472-6920-12-121] [Medline: 23216816]
7. Toro-Huamanchumo CJ, Toro-Huamanchumo C, Arce-Villalobos L, Arce-Villalobos L. The clinical eye: a need to improve the teaching of semiology in undergraduate medical education. Int J Med Students 2014 Oct 25;2(3):144-145. [doi: 10.5195/ijms.2014.186]
8. Prislin MD, Fitzpatrick CF, Lie D, Giglio M, Radecki S, Lewis E. Use of an objective structured clinical examination in evaluating student performance. Fam Med 1998 May;30(5):338-344. [Medline: 9597531]
9. Mondal R, Sarkar S, Nandi M, Hazra A. Comparative analysis between objective structured clinical examination (OSCE) and conventional examination (CE) as a formative evaluation tool in Pediatrics in semester examination for final MBBS students. Kathmandu Univ Med J (KUMJ) 2012;10(37):62-65. [doi: 10.3126/kumj.v10i1.6917] [Medline: 22971865]
10. Elfaki OA, Al-Humayed S. Medical students' perception of OSCE at the Department of Internal Medicine, College of Medicine, King Khalid University, Abha, KSA. J Coll Physicians Surg Pak 2016 Feb;26(2):158-159. [Medline: 26876409]
11. Kronfly Rubiano E, Ricarte Díez JI, Juncosa Font S, Martínez Carretero JM. [Evaluation of the clinical competence of Catalonian medicine schools 1994-2006. Evolution of examination formats until the objective and structured clinical evaluation (ECOE)]. Med Clin (Barc) 2007 Dec 01;129(20):777-784. [doi: 10.1157/13113768] [Medline: 18093480]
12. Carpenter JL. Cost analysis of objective structured clinical examinations. Acad Med 1995 Sep;70(9):828-833. [Medline: 7669163]
13. Sánchez-Ferrer ML, Grima-Murcia MD, Sánchez-Ferrer F, Hernández-Peñalver AI, Fernández-Jover E, Sánchez Del Campo F. Use of eye tracking as an innovative instructional method in surgical human anatomy. J Surg Educ 2017;74(4):668-673. [doi: 10.1016/j.jsurg.2016.12.012] [Medline: 28126379]
14. Koh RYI, Park T, Wickens CD, Ong LT, Chia SN. Differences in attentional strategies by novice and experienced operating theatre scrub nurses. J Exp Psychol Appl 2011 Sep;17(3):233-246. [doi: 10.1037/a0025171] [Medline: 21942313]
15. Meining A, Atasoy S, Chung A, Navab N, Yang GZ. "Eye-tracking" for assessment of image perception in gastrointestinal endoscopy with narrow-band imaging compared with white-light endoscopy. Endoscopy 2010 Aug;42(8):652-655. [doi: 10.1055/s-0029-1244233] [Medline: 20539975]

16. Harrison TK, Kim TE, Kou A, Shum C, Mariano ER, Howard SK, ADAPT (Anesthesiology-Directed Advanced Procedural Training) Research Group. Feasibility of eye-tracking technology to quantify expertise in ultrasound-guided regional anesthesia. J Anesth 2016 Jun;30(3):530-533. [doi: 10.1007/s00540-016-2157-6] [Medline: 26980475]
17. Kok EM, Jarodzka H, de Bruin ABH, BinAmir HAN, Robben SGF, van Merriënboer JJG. Systematic viewing in radiology: seeing more, missing less? Adv Health Sci Educ Theory Pract 2016 Mar;21(1):189-205 [FREE Full text] [doi: 10.1007/s10459-015-9624-y] [Medline: 26228704]
18. van der Gijp A, Ravesloot CJ, Jarodzka H, van der Schaaf MF, van der Schaaf IC, van Schaik JPJ, et al. How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. Adv Health Sci Educ Theory Pract 2017 Aug;22(3):765-787 [FREE Full text] [doi: 10.1007/s10459-016-9698-1] [Medline: 27436353]
19. Jarodzka H, Boshuizen HPA, Kirschner PA. Cognitive skills in medicine. In: Lanzer P, editor. Catheter-Based Cardiovascular Interventions. Berlin, Heidelberg: Springer; 2013.
20. Nekkanti S, Manjunath S, Mahtani A, Meka A, Rao T. A survey based feedback analysis of the current medical teaching methodology and trends in medical research practice in a South Indian Medical Institute. Int J Med Students 2018 Apr 30;6(1):6-17. [doi: 10.5195/ijms.2018.3]
21. Kok EM, Jarodzka H. Before your very eyes: the value and limitations of eye tracking in medical education. Med Educ 2017 Jan;51(1):114-122. [doi: 10.1111/medu.13066] [Medline: 27580633]
22. Warren AL, Donnon TL, Wagg CR, Priest H. Educational interventions to improve cytology visual diagnostic reasoning measured by eye tracking. J Vet Med Educ 2015;42(1):69-78. [doi: 10.3138/jvme.0514-056R] [Medline: 25497363]
23. O'Meara P, Munro G, Williams B, Cooper S, Bogossian F, Ross L, et al. Developing situation awareness amongst nursing and paramedicine students utilizing eye tracking technology and video debriefing techniques: a proof of concept paper. Int Emerg Nurs 2015 Apr;23(2):94-99. [doi: 10.1016/j.ienj.2014.11.001] [Medline: 25434606]
24. Breen CJ, Bond R, Finlay D. An evaluation of eye tracking technology in the assessment of 12 lead electrocardiography interpretation. J Electrocardiol 2014;47(6):922-929. [doi: 10.1016/j.jelectrocard.2014.08.008] [Medline: 25200901]
25. Cömert M, Zill JM, Christalle E, Dirmaier J, Härter M, Scholl I. Assessing communication skills of medical students in objective structured clinical examinations (OSCE)--a systematic review of rating scales. PLoS One 2016;11(3):e0152717 [FREE Full text] [doi: 10.1371/journal.pone.0152717] [Medline: 27031506]
26. Moineau G, Power B, Pion AJ, Wood TJ, Humphrey-Murto S. Comparison of student examiner to faculty examiner scoring and feedback in an OSCE. Med Educ 2011 Feb;45(2):183-191. [doi: 10.1111/j.1365-2923.2010.03800.x] [Medline: 21166691]
27. Monteiro SD, Walsh A, Grierson LEM. OSCE circuit performance effects: does circuit order influence scores? Med Teach 2016;38(1):98-100. [doi: 10.3109/0142159X.2015.1075647] [Medline: 26421751]
28. Allard M, Lafleur A, Richard E, Lebouthillier A, Vailles C. How medical students edited an OSCE study guide and why should you? Int J Med Students 2018 May 09;6(2):78-82. [doi: 10.5195/ijms.2018.248]

Abbreviations

CPR: cardiopulmonary resuscitation
OSCE: objective structured clinical examination
SOE: structured oral examination

Edited by G Eysenbach; submitted 07.01.20; peer-reviewed by M Gaman, S Panda, J Wilkinson, V Osadchiy, K Alipasandi; comments to author 10.03.20; revised version received 02.05.20; accepted 13.05.20; published 21.08.20.

Please cite as:
Grima-Murcia MD, Sanchez-Ferrer F, Ramos-Rincón JM, Fernández E
Use of Eye-Tracking Technology by Medical Students Taking the Objective Structured Clinical Examination: Descriptive Study
J Med Internet Res 2020;22(8):e17719
URL: http://www.jmir.org/2020/8/e17719/
doi:10.2196/17719
PMID:32821060

Original Paper

Exploring the Use of Evidence From the Development and Evaluation of an Electronic Health (eHealth) Trial: Case Study

Monika Jurkeviciute[1], MSc; Henrik Eriksson[1], MSc, PhD

Centre for Healthcare Improvement, Chalmers University of Technology, Gothenburg, Sweden

Corresponding Author:
Monika Jurkeviciute, MSc
Centre for Healthcare Improvement
Chalmers University of Technology
Vera Sandbergs allé 8
Gothenburg,
Sweden
Phone: 46 766061558
Email: monika.jurkeviciute@chalmers.se

Abstract

Background: Evidence-based practice refers to building clinical decisions on credible research evidence, professional experience, and patient preferences. However, there is a growing concern that evidence in the context of electronic health (eHealth) is not sufficiently used when forming policies and practice of health care. In this context, using evaluation and research evidence in clinical or policy decisions dominates the discourse. However, the use of additional types of evidence, such as professional experience, is underexplored. Moreover, there might be other ways of using evidence than in clinical or policy decisions.

Objective: This study aimed to analyze how different types of evidence (such as evaluation outcomes [including patient preferences], professional experiences, and existing scientific evidence from other research) obtained within the development and evaluation of an eHealth trial are used by diverse stakeholders. An additional aim was to identify barriers to the use of evidence and ways to support its use.

Methods: This study was built on a case of an eHealth trial funded by the European Union. The project included 4 care centers, 2 research and development companies that provided the web-based physical exercise program and an activity monitoring device, and 2 science institutions. The qualitative data collection included 9 semistructured interviews conducted 8 months after the evaluation was concluded. The data analysis concerned (1) activities and decisions that were made based on evidence after the project ended, (2) evidence used for those activities and decisions, (3) in what way the evidence was used, and (4) barriers to the use of evidence.

Results: Evidence generated from eHealth trials can be used by various stakeholders for decisions regarding clinical integration of eHealth solutions, policy making, scientific publishing, research funding applications, eHealth technology, and teaching. Evaluation evidence has less value than professional experiences to local decision making regarding eHealth integration into clinical practice. Professional experiences constitute the evidence that is valuable to the highest variety of activities and decisions in relation to eHealth trials. When using existing scientific evidence related to eHealth trials, it is important to consider contextual relevance, such as location or disease. To support the use of evidence, it is suggested to create possibilities for health care professionals to gain experience, assess a few rather than a large number of variables, and design for shorter iterative cycles of evaluation.

Conclusions: Initiatives to support and standardize evidence-based practice in the context of eHealth should consider the complexities in how the evidence is used in order to achieve better uptake of evidence in practice. However, one should be aware that the assumption of fact-based decision making in organizations is misleading. In order to create better chances that the evidence produced would be used, this should be addressed through the design of eHealth trials.

(J Med Internet Res 2020;22(8):e17718) doi:10.2196/17718

KEYWORDS

evidence-based practice; evidence use; eHealth; evaluation; evaluation use

Introduction

Evidence-based medicine has taken a central role in health care, aiming to increase the quality of clinical practice. In the medical domain, it is conceptualized as building clinical decisions on credible research evidence, professional experience and judgement, and patient preferences [1-3]. This trend has also risen in the evaluation and implementation of information and communication technologies (ICT) in health care (electronic health [eHealth]) [4,5]. Similar to conventional medicine, decision making in the implementations of eHealth solutions should "rely on explicit evidence derived from rigorous studies on what makes systems clinically acceptable, safe, and effective – not on basic science or experts alone" [6]. Hence, evidence should have utility in these decisions (ie, it should be usable and used). However, there is a growing concern that scientific evidence on whether eHealth works and is safe to use is not sufficiently used when forming health care policies and practice [4,7-11].

Evidence produced by evaluations dominates the discourse related to the evidence-based practice of eHealth implementations [7,12,13]. This emphasis and the strategies to support the use of this type of evidence can be seen through the scholarly discussions and sound methodological base developed in the form of evaluation guidelines, standard measures, and evaluation frameworks [7,13,14]. However, evidence-based practice includes additional types of evidence, such as professional experience and judgement, existing scientific evidence from other research, and patient preferences [1,2,7]. The use of these types of evidence generated through testing and implementing eHealth solutions is underexplored, leading to a lack of supporting strategies.

When the expectation is to use the evaluation evidence in making decisions regarding clinical implementations of eHealth [5], it refers to instrumental use, which is the direct use of the information in decision making and taking action, in order to change existing practice [15-19]. When evidence is not used instrumentally, it is referred to as a waste of resources and efforts contributing to the phenomenon of "pilotism" (remaining in a pilot state and not taken to integration) [20,21]. However, previous research has identified a number of other ways of evidence use [15,17-19,22,23]. Conceptual use refers to a nondirect use of information and perspectives to enhance understanding. Strategic or symbolic use happens when the evidence is brought up to support or confront an existing idea or decision. To the best of our knowledge, the different ways of evidence use (instrumental, conceptual, or symbolic) in the context of eHealth have not been addressed by previous research. Discussions limited to instrumental use potentially provide a too narrow view of actual evidence use in practice.

Furthermore, the considered users of evidence in the context of eHealth are usually limited to policy makers or health care professionals. However, eHealth is a multi-disciplinary field [14], and there might be more beneficiaries of the evidence. Exploring the types of evidence and the ways different actors use it can be worthwhile to support the uptake of evidence in the context of eHealth.

The purpose of this study was to analyze how different types of evidence (such as evaluation outcomes [including patient preferences], professional experiences obtained within the development and evaluation of an eHealth trial, and existing scientific evidence from other research) are used by diverse stakeholders. An additional aim was to identify barriers to the use of evidence and ways to support its use.

Methods

Context

This study was built on a case of a multinational and interdisciplinary European Union–funded project (for anonymity reasons, called "Alpha" in this paper). It was a nonpharmacological eHealth trial aimed at improving quality of life and increasing the independence of elderly with mild cognitive impairment and mild dementia. The Alpha trial introduced a case manager role and an ICT platform that consisted of web-based physical and cognitive exercise programs and an activity monitoring device to be used at home. As such, innovations were introduced on both the service model and technological levels.

The Alpha approach was implemented and tested in 4 countries. The project involved 8 partners: 4 care centers (in the aforementioned countries), 2 research and development companies that provided the eHealth solutions, and 2 science institutions. The trial lasted for 6 months. The evaluation included a number of variables, such as clinical efficacy, quality of life, patient adherence to technology, patient and health care professional's satisfaction, process effectiveness, and a cost-benefit analysis. Clinical professionals were asked to collect patient data using a number of standardized and custom-made questionnaires, as well as to register some data in the registries in Excel files. The evaluation was performed at the end of the project and was finalized in June 2018. During the evaluation, several project partners were charged analyzing the different variables. As it frequently happens in multinational projects, the evaluation had to overcome several practical circumstances such as ethical approvals and systems integration issues that delayed patient recruitment and the start of the trial. This created a situation in which the trial time had to be shortened for some patients, resulting in a smaller dataset than planned.

Data Collection

The evidence considered in this study included evaluation results (including patient preferences) and professional experiences from the Alpha trial as well as existing scientific evidence from other research. Data were collected through 9 semistructured interviews with all the partners involved in the Alpha project (4 care centers, 2 research and development companies that provided the eHealth solutions, and 2 science institutions) that were conducted 8 months after the evaluation was concluded. The stakeholders were delimited to the partners of the Alpha project, since they had deep knowledge and experience from the trial and they were the primary candidates to consider using the evidence developed. However, the funding institution did not require the project partners to use evidence from the Alpha project.

The interviewee selection followed the principles of purposive sampling [24] and involved the key members of the Alpha teams in every country (see Table 1). At least one interview per partner was conducted. The interviewees were in a position to either use the evidence directly in making decisions that change practice (clinical, technological, scientific) or to decide whether the evidence is worthy of suggesting or presenting to the decision makers in the organizations. The positions of the interviewees and the industry of their work are presented in Table 1. All the interviews lasted one hour, were conducted via Skype, were recorded, and were transcribed.

Table 1. Interview respondents.

Stakeholders	Interviewee occupation
Care centers	
Care center 1	Clinical neuropsychologist
Care center 2	Quality director
Care center 2	Senior physician
Care center 3	Head of eHealth[a] research
Care center 4	Project manager
Research and development companies	
Research and development company 1	Coordinator of the eHealth group
Research and development company 2	Project manager and scientific coordinator
Science institutions	
Science institution 1	Director of the research center
Science institution 2	Scientific coordinator

[a]eHealth: electronic health.

The interviews followed a guide structured around the components of evidence use [18,19,25], such as evidence users, types of impact, evidence already used (and useful) depending on the agenda of the stakeholder, agenda or purpose, quality of research, methodological credibility, relevance and timing for the organization to use the evidence, presentation of the results, and future plans in relation to the evidence.

Data Analysis

For every stakeholder, we analyzed the following: (1) the types of activities and decisions that were made after the Alpha project ended; (2) the types of evidence (evaluation results from the trial [including patient preferences], professional experiences, or existing scientific evidence from other research) that were used for those activities and decisions, based on the definitions of evidence [1]; (3) in what way the evidence was used (instrumental, conceptual, symbolic) [15,17-19,22]; and (4) barriers to the use of evidence. Instrumental use of evidence was assumed if the evidence obtained from the trial had a direct impact on practice decisions. Conceptual use of evidence was assumed when the evidence from the trial was used indirectly in ways that impacted the understanding, attitudes, and knowledge of the stakeholders but did not cause a change in practice. Symbolic use of evidence was concluded when the stakeholder had used evidence in confirming previous decisions. If necessary, the data were validated with professionals from the Alpha partners.

The findings were grouped by the types of activities and decisions made by the stakeholders using evidence from Alpha. For each activity or decision, the use of evidence is discussed.

Results

At 8 months after the Alpha project ended, partners in the project (stakeholders) used evidence in the following decisions and activities: (1) integrating or abandoning the Alpha approach in clinical practice, (2) publishing results from the Alpha study, (3) applying for new research funding, (4) supporting regional policymaking, (5) improving technology, and (6) teaching students and health care professionals. Next, the ways that evidence were used by the stakeholders in every decision and activity are explained.

Integrating the Alpha Approach Into Clinical Practice

At the time of this study, 2 of the 4 care centers had decided to integrate the Alpha approach into clinical practice (care centers 1 and 2). In these care centers, the decision was mainly informed by the evidence from professional experiences. The health care professionals decided to adopt the Alpha approach based on their experiences with usability and adherence to the technology, as well as on patient satisfaction with the service. At the time of the decision, the evaluation results were not available yet. However, the professionals relied on their experience and the existing scientific evidence from other research that demonstrated that technologies and care models similar to Alpha can be beneficial for the patients targeted. The existing research was quite explicit on the benefits of physical and cognitive exercise (with and without the help of technology) for patients with cognitive impairments.

> *Specific clinical data are not a reason to not try to implement this model. Existing research can provide us such information. <…> Data related to adherence*

> *are enough to be interpreted as a useful model for these patients.* [Care center 1, clinical neuropsychologist]

The decisions to adopt the Alpha approach in care centers 1 and 2 were facilitated by the fact that resources for implementation were available from regional policies supporting and financing care models like Alpha. Care centers 1 and 2 planned to perform deeper statistical analyses on the clinical outcomes, cost, and savings. If a positive effect was found, the results would be disseminated, which would support their decision to integrate the Alpha approach in practice. Hence, in the case of care centers 1 and 2, professional experiences and existing scientific evidence from other research were used instrumentally (directly in decision making and action), while evaluation evidence was used symbolically (to strengthen the already taken decision).

> *When the decision was made, we didn't have any results yet. <...> Our experience and preliminary data showed that this model is quite good. <…> Managers trusted our previous evaluations of similar models and thought that it will be the same. <…> We have to redo the economic evaluation to see how much it actually costs and how much we save.* [Care center 2, quality director]

Care center 3 planned to use the evaluation results to make a decision regarding adopting the Alpha approach in its clinical practice. The organization planned to present the evaluation results to the board and express a need for an eHealth solution like that tested in the Alpha project. Once approval from the board is obtained, the technology can be purchased and integrated in clinical practice. In this case, using the evaluation results in decision making for practice improvement indicates instrumental use.

> *Once we demonstrate that the results are OK, we are in a position to escalate it to decision makers, and we are able to incorporate it in our organization.* [Care center 3, Head of eHealth research]

Care center 4 decided to abandon the Alpha approach. Professional experiences were the primary influence on this decision. The concept was abandoned when the staff realized, over the course of the trial, how many resources the new model requires when applying it within the context of care center 4. It was deemed not the right time for the concept to be adopted in the organization. After the project finished, staff's experiences in the care process of Alpha were presented to management. In the case of care center 4, professional experiences were used instrumentally (directly in decision making and action).

> *We didn't pay so much attention to the results of the evaluation. We looked at what does it mean to work with patients in a situation like that. <…> For management, the descriptive conclusions were more interesting than the analysis.* [Care center 4, project manager]

Publishing Results of the Alpha Project

Publishing the results from the Alpha project in scientific outlets was initiated by almost all the partners of Alpha (except care center 4). The care centers used the clinical outcomes, quality of life, patient and employee satisfaction, and cost data in the scientific publications. The research and development companies used the adherence data and the feedback from patients and health care professionals related to their specific technology. Publishing the evaluation results helped these companies strengthen their image by demonstrating a case of the technology application in a real setting. In addition to the evaluation results, the partners used the existing scientific evidence from other research to build a case for research.

Since scientific publishing did not directly relate to decision making in practice, such use of evaluation evidence and existing evidence from other research was deemed conceptual.

Applying for New Research Funding

Most of the partners (care centers 1, 2, and 3; science institutions; and research and development companies) used Alpha evaluation results and professional experiences when applying for further research funding. Evaluation outcomes, experiences, and lessons learned in Alpha provided the basis for the case and allowed ideas to be built that could be applied in the next project. In such a case, Alpha evaluation results and professional experiences did not change local practices but increased the knowledge and understanding of the partners that subsequently helped to develop better research ideas and improve the design of the future studies. Hence, such use of evaluation evidence and professional experiences was deemed conceptual.

Supporting Regional Policy

Science institutions 1 and 2 and care centers 1, 2, and 3 presented Alpha results in regional policymaking activities as a concrete local example of an eHealth-supported care model within their regions. For this purpose, science institution 2 used the managerial and economic evidence from the evaluation of Alpha to demonstrate the possible impact of the eHealth solution on the local care facility. In addition, science institutions 1 and 2 used the health care professionals' feedback and perceptions on working with Alpha in their local contexts to demonstrate local applicability. Since the Alpha case served as an example and was not meant to make decisions based on its results, such use of evaluation evidence and professional experiences in policymaking was deemed conceptual.

> *When you discuss a real case here in <region>, these messages are stronger than to discuss cases in <another country> or to say that literature says these things are useful.* [Science institution 2, scientific coordinator]

Improving eHealth Technology

The evaluation results of Alpha helped research and development company 2 in making decisions to improve its technology. The company focused on the patients' and health care professionals' feedback and preferences in relation to its technology and initiated actions to improve it. Since the evaluation evidence was used directly for decisions and action, we classified such use of evidence as instrumental.

Teaching Students and Health Care Professionals

Professional experiences with Alpha were used by science institution 1 in teaching students and health care professionals. Science institution 1 relied on the professionals' feedback and perceptions on working with Alpha in their local contexts, as it demonstrates local applicability. Since the Alpha case served as an example and decisions were not meant to be made based on its results, such use of professional experiences in teaching was deemed conceptual.

> *We used the experience with the care models as an example in our courses. <…> We use it for practitioners as a subject to discuss and reflect upon.* [Science institution 1, director of the research center]

Table 2 describes situations of using the evidence 8 months after the project ended.

Table 2. Evidence use by different stakeholders.

Decisions taken and activities	Use of different types of evidence in making the decision or performing the activity		
	Alpha evaluation results	Professional experience with Alpha	Existing scientific evidence from other research
Care center 1			
Adopt Alpha approach	Symbolic	Instrumental	Instrumental
Publish results	Conceptual	No use observed	Conceptual
Support regional policy	No use observed	Conceptual	No use observed
Care center 2			
Adopt Alpha approach	Symbolic	Instrumental	Instrumental
Publish results	Conceptual	No use observed	Conceptual
Apply for research funding	Conceptual	Conceptual	No use observed
Support regional policy	No use observed	Conceptual	No use observed
Care center 3			
Present Alpha approach to decision makers for full implementation	Instrumental (planned)	Instrumental (planned)	Instrumental (planned)
Publish results	Conceptual	No use observed	No use observed
Apply for research funding	Conceptual	Conceptual	No use observed
Support regional policy	No use observed	Conceptual	No use observed
Care center 4			
Abandon Alpha approach	No use observed	Instrumental	No use observed
Science institution 1			
Publish results	Conceptual	No use observed	Conceptual
Teach students and health care professionals	No use observed	Conceptual	No use observed
Support regional policy	No use observed	Conceptual	No use observed
Apply for research funding	Conceptual	Conceptual	No use observed
Science institution 2			
Support regional policy	Conceptual	Conceptual	No use observed
Apply for research funding	Conceptual	Conceptual	No use observed
Publish results	Conceptual	Conceptual	Conceptual
Research and development company 1			
Publish results	Conceptual	No use observed	Conceptual
Apply for research funding	Conceptual	Conceptual	No use observed
Research and development company 2			
Improve technology	Instrumental	Instrumental	No use observed
Publish results	Conceptual	No use observed	Conceptual

Barriers to the Use of Evidence From the Alpha Trial

The first barrier to the use of evidence was related to the number of variables included in the Alpha evaluation. In most of the project locations, the scope of evaluation was deemed too extensive (it included several variables related to clinical efficacy, quality of life, patient adherence to technology, patient and health care professionals' satisfaction, a number of variables to assess process effectiveness, and a cost-benefit analysis). The interviewees indicated that the time needed to collect this amount of data for every patient was too long. The clinicians had to fit the data collection into their routine work during meetings with the patients. Consequently, the clinicians were making choices about which data to collect at a particular time. Such trade-offs between data collection for the project and time spent with a patient affected the completeness of data collected and consequently the quality of evidence produced in the evaluation.

> *When you want to monitor a lot of variables, it is directly related to the time you need to spend with the patients. <…> The target should be to optimize how we collect the variables and information using not that much time.* [Care center 3, Head of eHealth research]

The second barrier to the use of evidence was related to the Alpha evaluation design when comparing before-after situations. Some of the interviewees with a clinical background disagreed that eHealth integration decisions can be purely based on hard facts and not on evidence from the evaluation when making such a decision. According to these respondents, novel eHealth-supported care models tested during trials are complex dynamic systems within the local context that vary, have differences in culture, and are affected by social interaction. Since these care models cannot be assumed as stable controlled systems, the assumption of stability in the traditional before-after measurement design of an eHealth trial provides less valuable information for eHealth integration decisions to improve practice. Additionally, such an evaluation design comparing before-after situations does not maximize the potential to enhance local learning. The interviewees suggested that people's experiences with an eHealth solution and process measures, both captured continuously, could lead to iterative adaptation and adjustment between the eHealth solution and the context. Such iterative evaluation would provide higher value in these eHealth integration decisions to improve practice.

> *If we own the evaluation, we would take repeated measurements for improvement efforts and enhanced learning, rather than traditional approaches.* [Care center 2, quality director]

Discussion

Principal Findings

In this study, we analyzed how different types of evidence, generated through the development and evaluation of an eHealth trial, are used by diverse stakeholders. This work demonstrated that evidence from eHealth trials is used in more ways than decision making regarding clinical integration or policymaking. In addition, different stakeholders can use the evidence for scientific publishing and dissemination, eHealth technology improvement, research funding applications, and teaching.

We found that professional experiences seem to have more influence over decisions regarding eHealth integration into clinical practice than formal evaluations and research. Learning whether and how an eHealth solution could fit within the local context provides key information for local decision making. If the design of an eHealth trial fails to create conditions for professionals to gain experiences, it might prevent learning and obtaining evidence that are crucial to increase the success rate of eHealth trials in their post-pilot phase and reduce "pilotism" [5,20]. Moreover, professional experiences from eHealth trials provide evidence that is valuable for the greatest variety of activities that include disseminating knowledge in various formats, policymaking, teaching, and providing feedback for technology improvement. Evaluation evidence might mostly be valuable for scientific publishing. Existing scientific evidence from other research is another type of evidence that can help to make decisions when it comes to integrating eHealth solutions into clinical practice. However, contextual relevance (eg, location, disease) matters.

To support the use of evaluation evidence, one could consider assessing a few, rather than a large number of, variables during an evaluation. It can help ensure the quality of data collected, preventing from making trade-offs between the time required and quality of the data. Additionally, shorter iterative cycles of evaluation could create better possibilities for health care professionals to gain more experience and use it as evidence in decision making regarding integration of eHealth solutions. Professional experiences could enhance evidence when relevant professionals are included and accumulate experience during eHealth trials. Such an approach could increase the degree of learning and chances that the eHealth solution would be integrated into practice and reach sustainability. Existing scientific evidence collected from previously conducted projects or initiatives in the same location as an eHealth solution is implemented can support decisions regarding eHealth integrations better than scientific evidence obtained from other locations. Research evidence produced in other contexts can be problematic for direct translation into making such decisions. However, it has value for scientific dissemination.

Limitations

The findings of this study were based on a specific innovation research project funded by the European Union. The interviewee sample was delimited to the partners of the project, since they had deep knowledge of the project and its results and were in the favorable position to use the evidence obtained. However, perspectives of the funding agency, industry, or governments could be a valuable avenue for further research. Similarly, evidence produced in other settings and study designs could provide a different view of the use of evidence. Furthermore, the study captured the situation 8 months after the Alpha project was finished. However, the use of evidence might be more extensive in later stages due to the so-called "gestation period" [25].

Comparison With Prior Work

Previous research on evidence-based practice usually described clinical implementations and policymaking [5,7] as evidence use in the context of eHealth. By focusing on a wider ecosystem of stakeholders than the traditional focus on health care providers and policymakers, our work identifies more uses of evidence, such as scientific publishing and dissemination, eHealth product and service improvement, applying for research funding, and teaching. Furthermore, we analyzed the use of additional types of evidence [1] such as professional experiences in addition to the traditional focus on evidence generated by evaluations and research [7,12,13].

Our study shows that the typical discourse on the instrumental use (in making decisions) of evidence (and lack of it) [5] does not sufficiently reflect the actual use of evidence in the context of eHealth. Evidence also serves the stakeholders conceptually (increasing knowledge and understanding) and symbolically (supporting already taken decisions) [15,17-19,22]. This suggests that evidence created through eHealth trials can have utility beyond decision making in clinical implementations of eHealth solutions or policymaking. Therefore, evidence that is not used instrumentally might not be a waste of resources and effort (the problem of so-called "pilotism"), as frequently judged by scholars [5,20]. Viewing the use of evidence through different users and the ways of using it can reveal the actual ways in which evidence from a trial is used, help commissioning bodies have realistic expectations on the influence trials can have, and help to design better trials.

Our study indicates that experiences in eHealth trials matter more than facts when it comes to how evidence is used for eHealth integration. The reason for this could be that "hard facts" from evaluation are difficult to straightforwardly implement in a complex health care reality [26,27]. Although it is attractive to think about organizations as rationale organisms and systems [5,28], organizations are rather characterized by decisions and actions that are far from rational. Instead, personal incentives, organizational culture, and other more subjective fields come into play when explaining how decisions are made and organizations develop [27,29,30]. Hence, we believe that the designers of eHealth trials could benefit from understanding how learning and continuous improvement are created and how knowledge development occurs in contemporary health care organizations. In other words, these fields could explain why opinions and subjective knowledge are more important than "hard facts" from an evaluation.

We identified a number of strategies to support the use of evidence in practice in the context of eHealth trials. First, we suggest focusing on a smaller number of variables during evaluation, to ensure the quality of collected data. Such a strategy is contradictory to the ever-expanding evaluation frameworks that include a growing number of variables (eg, [31,32]) and arguments that evaluations should aim to capture as wide an array of outcomes as possible [13]. Second, alternative designs to eHealth evaluation have been discussed in prior research (eg, [14,33]). Our study showed that such designs leading to shorter cycles of evaluation and enabling learning, iteration, and forming experiences that can support decisions could be more beneficial in improving the practice of different stakeholders [34-36]. Failure to produce timely results for decision making is among the barriers to the use of evidence identified by previous research [26].

Conclusions

We conclude that various stakeholders (such as care centers, research and development companies, science institutions) benefit from evidence differently. Therefore, the delimited focus of research on decision making in clinical settings or policymaking does not capture the actual beneficiaries and realities of evidence use. In addition, when making decisions regarding improving practice, stakeholders do not necessarily rely on the factual evidence produced by evaluations. We conclude that, in the context of eHealth, professional experiences seem to have more influence over decisions than formal evaluations and research. Hence, we suggest that scientific and practical discussions around evidence-based practice in the context of eHealth should include all sorts of evidence (evaluation evidence, professional experiences, existing scientific evidence from other research, and patient preferences). Additionally, it could be beneficial to have an in-depth view on how the evidence is used. This could help expand the conventional focus on clinical settings or policymaking and the direct use of evidence produced by research or evaluation when making decisions. Initiatives to support and standardize evidence-based practice in the context of eHealth should take these complexities into consideration to achieve better uptake of evidence in practice.

Acknowledgments

The Alpha project was funded by the European Union's Horizon 2020 research and innovation program. The funding organization took no part in designing this study, collecting and analyzing the data, or in preparing this manuscript. The authors would like to express gratitude to the Alpha research consortium for providing data for this article. Written consent to publish was obtained from the project partners whose quotes have been used in this study.

Conflicts of Interest

None declared.

References

1. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992 Nov 04;268(17):2420-2425. [doi: 10.1001/jama.1992.03490170092032] [Medline: 1404801]
2. Greenhalgh T, Howick J, Maskrey N, Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? BMJ 2014 Jun 13;348(jun13 4):g3725-g3725 [FREE Full text] [doi: 10.1136/bmj.g3725] [Medline: 24927763]
3. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. The Lancet 2017 Jul;390(10092):415-423. [doi: 10.1016/s0140-6736(16)31592-6]
4. Ammenwerth E, Rigby M, editors. Evidence-based health informatics: Promoting safety and efficiency through scientific methods and ethical policy. Amsterdam, Netherlands: IOS Press; May 20, 2016.
5. Rigby M, Ammenwerth E, Beuscart-Zephir M, Brender J, Hyppönen H, Melia S, et al. Evidence Based Health Informatics: 10 Years of Efforts to Promote the Principle. Yearb Med Inform 2018 Mar 05;22(01):34-46. [doi: 10.1055/s-0038-1638830]
6. Wyatt J. Evidence-based Health Informatics and the Scientific Development of the Field. In: Ammenwerth E, Rigby M, editors. Evidence-based health informatics: Promoting safety and efficiency through scientific methods and ethical policy. Amsterdam, Netherlands: IOS Press; May 20, 2016:14-24.
7. Rigby M, Magrabi F, Scott P, Doupi P, Hypponen H, Ammenwerth E. Steps in Moving Evidence-Based Health Informatics from Theory to Practice. Healthc Inform Res 2016 Oct;22(4):255-260 [FREE Full text] [doi: 10.4258/hir.2016.22.4.255] [Medline: 27895956]
8. Koppel R. Is Healthcare Information Technology Based on Evidence? Yearb Med Inform 2018 Mar 05;22(01):07-12. [doi: 10.1055/s-0038-1638826]
9. Cohen G, Schroeder J, Newson R, King L, Rychetnik L, Milat AJ, et al. Does health intervention research have real world policy and practice impacts: testing a new impact assessment tool. Health Res Policy Syst 2015 Jan 01;13(1):3 [FREE Full text] [doi: 10.1186/1478-4505-13-3] [Medline: 25552272]
10. Alla K, Oprescu F, Hall WD, Whiteford HA, Head BW, Meurk CS. Can automated content analysis be used to assess and improve the use of evidence in mental health policy? A systematic review. Syst Rev 2018 Nov 15;7(1):194 [FREE Full text] [doi: 10.1186/s13643-018-0853-z] [Medline: 30442191]
11. Patrick J. The Validity of Personal Experiences in Evaluating HIT. Appl Clin Inform 2017 Dec 16;01(04):462-465. [doi: 10.4338/aci-2010-10-ie-0058]
12. Haux R. Preface. In: Ammenwerth E, Rigby M, editors. Evidence-based health informatics: Promoting safety and efficiency through scientific methods and ethical policy. Amsterdam, Netherlands: IOS Press; May 20, 2016:i-xv.
13. Ammenwerth E. Evidence-based Health Informatics: How Do We Know What We Know? Methods Inf Med 2018 Jan 22;54(04):298-307. [doi: 10.3414/me14-01-0119]
14. Ossebaard HC, Van Gemert-Pijnen L. eHealth and quality in health care: implementation time. Int J Qual Health Care 2016 Jun;28(3):415-419. [doi: 10.1093/intqhc/mzw032] [Medline: 27029590]
15. Alkin MC, Taut SM. Unbundling evaluation use. Studies in Educational Evaluation 2002 Mar;29(1):1-12. [doi: 10.1016/s0191-491x(03)90001-0]
16. Leviton LC, Hughes EF. Research On the Utilization of Evaluations. Eval Rev 2016 Jul 26;5(4):525-548. [doi: 10.1177/0193841x8100500405]
17. Rich R. Uses of social science information by federal bureaucrats: Knowledge for action versus knowledge for understanding. In: Weiss C. Using social research in public policy making. Lexington MA: Lexington Books; 1977:199-211.
18. Nutley S, Walter I, Davies H. Using evidence: How research can inform public services. Bristol, England: The Policy Press; 2007.
19. Alkin MC, King JA. Definitions of Evaluation Use and Misuse, Evaluation Influence, and Factors Affecting Use. Am J Eval 2017 Aug 04;38(3):434-450. [doi: 10.1177/1098214017717015]
20. Andreassen HK, Kjekshus LE, Tjora A. Survival of the project: a case study of ICT innovation in health care. Soc Sci Med 2015 May;132:62-69. [doi: 10.1016/j.socscimed.2015.03.016] [Medline: 25795426]
21. Urueña A, Hidalgo A, Arenas AE. Identifying capabilities in innovation projects: Evidences from eHealth. Journal of Business Research 2016 Nov;69(11):4843-4848. [doi: 10.1016/j.jbusres.2016.04.041]
22. Weiss CH. The Interface between Evaluation and Public Policy. Evaluation 2016 Jul 24;5(4):468-486. [doi: 10.1177/135638909900500408]
23. Estabrooks C, Squires J, Strandberg E, Nilsson-Kajermo K, Scott S, Profetto-McGrath J, et al. Towards better measures of research utilization: a collaborative study in Canada and Sweden. J Adv Nurs 2011 Aug;67(8):1705-1718. [doi: 10.1111/j.1365-2648.2011.05610.x] [Medline: 21477116]
24. Bryman A, Bell E. Business research methods. Oxford, England: Oxford University Press; 2011.
25. Feinstein ON. Use of Evaluations and the Evaluation of their Use. Evaluation 2016 Jul 24;8(4):433-439. [doi: 10.1177/13563890260620621]
26. Hammersley M. The myth of research-based policy and practice. London, England: Sage; 2015.
27. Freeman AC, Sweeney K. Why general practitioners do not implement evidence: qualitative study. BMJ 2001 Nov 10;323(7321):1100-1102 [FREE Full text] [doi: 10.1136/bmj.323.7321.1100] [Medline: 11701576]
28. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]

29. Argyris C, Schön DA. Organizational learning: A theory of action perspective. Reading, MA: Addison-Wesley; 1978.
30. Crossan M. Chris Argyris and Donald Schön's Organizational Learning: There is no silver bullet. AMP 2003 May;17(2):38-39. [doi: 10.5465/ame.2003.10025187]
31. Eslami Andargoli A, Scheepers H, Rajendran D, Sohal A. Health information systems evaluation frameworks: A systematic review. Int J Med Inform 2017 Jan;97:195-209. [doi: 10.1016/j.ijmedinf.2016.10.008] [Medline: 27919378]
32. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: mast. Int J Technol Assess Health Care 2012 Jan;28(1):44-51. [doi: 10.1017/S0266462311000638] [Medline: 22617736]
33. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014 Feb 19;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
34. Provost LP. Analytical studies: a framework for quality improvement design and analysis. BMJ Qual Saf 2011 Apr 30;20 Suppl 1(Suppl 1):i92-i96 [FREE Full text] [doi: 10.1136/bmjqs.2011.051557] [Medline: 21450782]
35. Moen R, Norman C. Evolution of the PDCA cycle. 2009 Presented at: 7th ANQ Congress; September 17, 2009; Tokyo, Japan p. 1-11. [doi: 10.1002/9781118785317.weom100162]
36. Deming W. The new economics for industry, government, education. Cambridge, MA: MIT press; 2018.

Abbreviations

eHealth: electronic health.
ICT: Information and communication technologies

Edited by G Eysenbach; submitted 07.01.20; peer-reviewed by A Persson, H Durrani; comments to author 17.03.20; revised version received 21.05.20; accepted 13.06.20; published 28.08.20.

Please cite as:
Jurkeviciute M, Eriksson H
Exploring the Use of Evidence From the Development and Evaluation of an Electronic Health (eHealth) Trial: Case Study
J Med Internet Res 2020;22(8):e17718
URL: http://www.jmir.org/2020/8/e17718/
doi:10.2196/17718
PMID:32857057

Original Paper

Experiences of a National Web-Based Heart Age Calculator for Cardiovascular Disease Prevention: User Characteristics, Heart Age Results, and Behavior Change Survey

Carissa Bonner[1], MPH, PhD; Natalie Raffoul[2], BPharm, Grad Cert Pharm Med, MHM; Tanya Battaglia[3], BA; Julie Anne Mitchell[2], BA, MPH, Grad Dip, RN; Carys Batcup[1], BSc, MSc; Bill Stavreski[3], BEc, MPPM

[1]Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[2]National Heart Foundation of Australia, Sydney, Australia

[3]National Heart Foundation of Australia, Melbourne, Australia

Corresponding Author:
Carissa Bonner, MPH, PhD
Sydney School of Public Health
Faculty of Medicine and Health
University of Sydney
Rm 226A, Edward Ford Building A27
Sydney,
Australia
Phone: 61 2 9351 7125
Email: carissa.bonner@sydney.edu.au

Abstract

Background: Heart age calculators are used worldwide to engage the public in cardiovascular disease (CVD) prevention. Experimental studies with small samples have found mixed effects of these tools, and previous reports of population samples that used web-based heart age tools have not evaluated psychological and behavioral outcomes.

Objective: This study aims to report on national users of the Australian heart age calculator and the follow-up of a sample of users.

Methods: The heart age calculator was launched in 2019 by the National Heart Foundation of Australia. Heart age results were calculated for all users and recorded for those who signed up for a heart age report and an email follow-up over 10 weeks, after which a survey was conducted. CVD risk factors, heart age results, and psychological and behavioral questions were analyzed using descriptive statistics and chi-square tests. Open responses were thematically coded.

Results: There were 361,044 anonymous users over 5 months, of which 30,279 signed up to receive a heart age report and 1303 completed the survey. There were more women (19,840/30,279, 65.52%), with an average age of 55.67 (SD 11.43) years, and most users knew blood pressure levels (20,279/30,279, 66.97%) but not cholesterol levels (12,267/30,279, 40.51%). The average heart age result was 4.61 (SD 4.71) years older than the current age, including (23,840/30,279, 78.73%) with an older heart age. For the survey, most users recalled their heart age category (892/1303, 68.46%), and many reported lifestyle improvements (diet 821/1303, 63.01% and physical activity 809/1303, 62.09%). People with an older heart age result were more likely to report a doctor visit (538/1055, 51.00%). Participants indicated strong emotional responses to heart age, both positive and negative.

Conclusions: Most Australian users received an older heart age as per international and UK heart age tools. Heart age reports with follow-up over 10 weeks prompted strong emotional responses, high recall rates, and self-reported lifestyle changes and clinical checks for more than half of the survey respondents. These findings are based on a more engaged user sample than previous research, who were more likely to know blood pressure and cholesterol values. Further research is needed to determine which aspects are most effective in initiating and maintaining lifestyle changes. The results confirm high public interest in heart age tools, but additional support is needed to help users understand the results and take appropriate action.

(J Med Internet Res 2020;22(8):e19028) doi:10.2196/19028

KEYWORDS

heart age; risk communication; cardiovascular disease prevention; eHealth; behavior change

Introduction

Heart age calculators are increasingly popular worldwide as a way to engage the public in cardiovascular disease (CVD) risk assessment [1]. Methods vary but heart age is generally calculated by comparing a person's absolute risk of a heart attack or stroke in the next 5-10 years with a person with *ideal* risk factor levels, such as a nonsmoker with 120 mm Hg systolic blood pressure. If any risk factor is higher than ideal (eg, 140 mm Hg systolic blood pressure), then the result is an older heart age [2]. Some calculators also allow younger heart age if risk factors are below the ideal threshold (eg, 110 mm Hg systolic blood pressure).

Heart age calculators are often used as motivational tools to raise personal awareness about CVD risk factors and prompt follow-up action. Millions of people have used web-based heart age calculators. An international Unilever campaign engaged 2.7 million users from 13 countries in 2009 to 2011 using a Framingham model–based heart age calculator [3], and a more recent UK version based on QRISK reached 1.4 million hits with almost 600,000 complete users over 5 months [4]. In New Zealand, a 5-year Framingham version was used to promote clinical guidelines [2], and health organizations in the United States and China have released population estimates of heart age [5,6]. Australia launched a heart age calculator as part of a national consumer awareness campaign in February 2019, which engaged approximately 1.6 million users over 16 months.

With an increasing number of heart age calculators becoming available on the web, it is important to note that the same person can get a very different heart age result depending on which calculator is used [1]. This depends on the underlying absolute risk model (eg, Framingham vs QRISK), the ideal thresholds set for risk factors (eg, systolic blood pressure 120 mm Hg), and restrictions in the way that absolute risk is converted to heart age (eg, allowing younger heart age or not). It is important to note that an older heart age is not the same as high absolute risk where clinical guidelines would recommend medication; it is an alternative risk communication format that indicates at least one elevated risk factor compared with the ideal levels set. It is therefore possible to have a low absolute risk but an older heart age, for example, a young woman with cholesterol levels above the ideal level.

With so much variability in the way these calculators are set up, heart age is not recommended as a clinical assessment tool for medication decisions [1]. However, they may engage users to consider the personal relevance of risk factors and lifestyle changes or to seek a more accurate risk assessment with their doctor [7]. Particularly in younger adults, low absolute risk may conceal a high relative risk of developing CVD, and heart age calculators can be useful in communicating the long-term consequences of an individual's lifestyle and associated risk factors [8]. Some trials have found that using a heart age tool improves risk factor control compared with standard care [9], and direct comparisons of heart age with absolute risk in the same interactive format have found greater emotional responses to heart age, but this may not necessarily translate into behavior change [10-12]. Therefore, additional support is needed to help users understand the results and take appropriate action based on heart age calculators.

Existing research on the effect of heart age calculators is divided into small experimental samples that were randomized (which show mixed results overall) [13] and large population samples where users have been simply described. This study aimed to draw these 2 areas together by reporting on the users of a new Australian heart age calculator, followed by lifestyle change outcomes in a smaller sample of users who signed up to receive a report and further support by email. Previous reports describing the general population's use of heart age calculators have not evaluated behavioral outcomes and include repeated or less serious uses of the tool (eg, just testing the tool or trying it out for someone else).

Methods

Materials

After conducting an environmental scan of international heart age calculators, the National Heart Foundation of Australia (NHFA) created the Australian version based on Framingham model algorithms. The calculator was developed with funding from an unrestricted and unconditional grant from Amgen, who did not contribute in any way to the development. Some adjustments were made in line with Australian guidelines (eg, ideal levels set at 120 mm Hg for systolic blood pressure and <4 mmol/L for total cholesterol), which resulted in some changes to the weightings for some gender or age groups in the published model. These were tested and discussed with a committee including general practitioners (GPs) and cardiologists to ensure that the calculator would not potentially lead to treatment based on single risk factors. A pop-up message prompted users to see a doctor if the blood pressure or cholesterol level was plausible but considered high risk in accordance with Australian guidelines, for example, "Total cholesterol above 7.5 mmol/L puts you at high risk of having a heart attack or stroke. Please see your doctor as soon as possible about your cholesterol." Implausible values prompted a different message about the range required, for example, "Please enter a number between 2 and 10.5." Heart age was calculated once a plausible value was entered. No adjustments were made to account for higher risk populations (eg, Aboriginal and Torres Strait Islander Peoples) because of a lack of clear evidence. The minimum reported heart age was set at <35 years and the maximum at ≥85 years. If blood pressure or cholesterol levels were not known, a population average was used based on the relevant 5-year age group in the National Health Survey data from 2011 to 2012 [14].

The resulting web-based heart age calculator was intended for people aged 35 to 75 years without existing CVD. The following information was collected: age, sex, family history of premature heart disease, smoking status, height, weight, diabetes status, blood pressure, cholesterol, and whether or not users were taking medication for high blood pressure. Users who did not know their blood pressure or cholesterol were informed that a population average would be used. The result was presented as the user's heart age and whether it was younger, the same as, or older than their current age. Users were encouraged to provide

their email address to obtain a more detailed report and those in the target age group were recommended to see a doctor for a heart health check for absolute risk assessment. The tool is available on the NHFA website [15]. Figure 1 shows example screenshots, and Figure 2 shows example reports.

Figure 1. Example screenshots from heart age calculator (eg, 54 year old male smoker, family history, diabetes and average blood pressure/cholesterol).

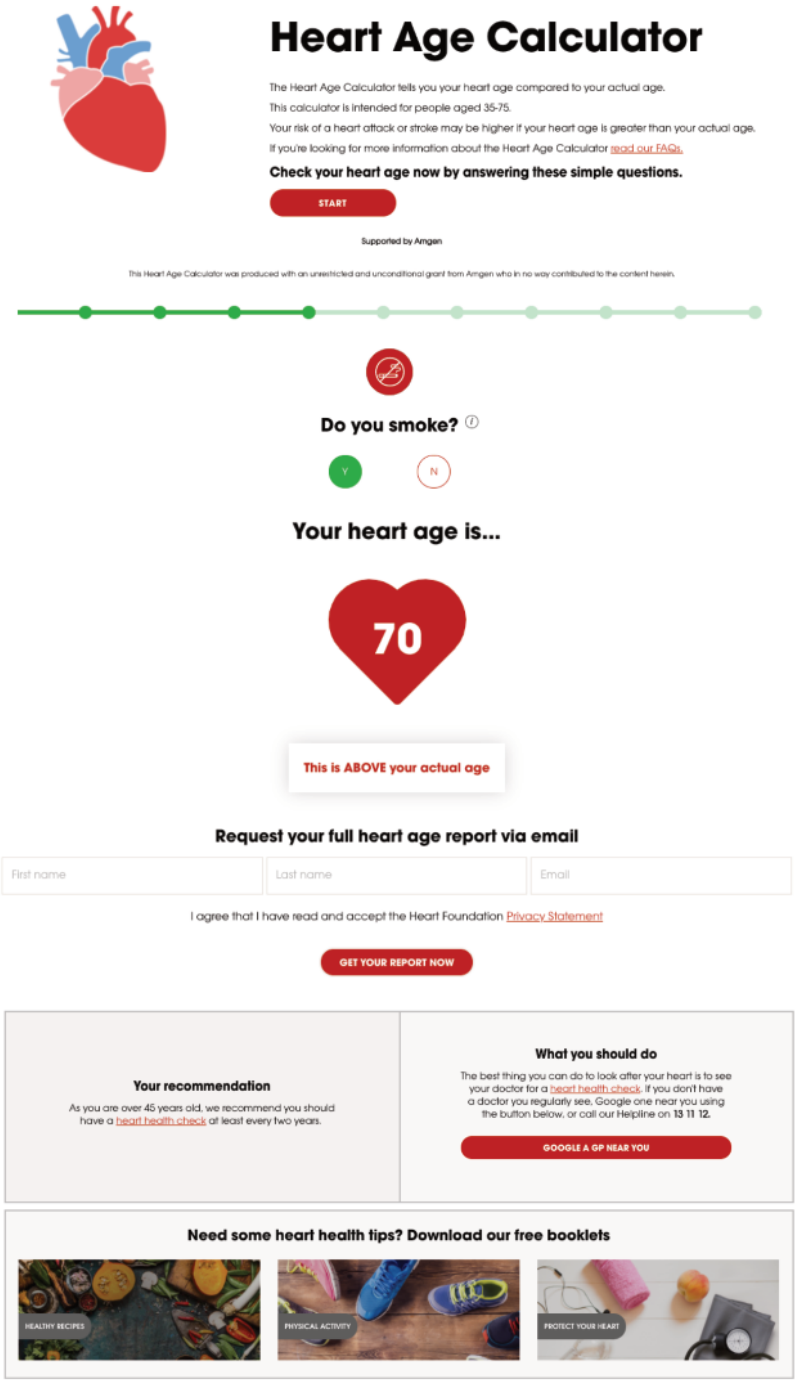


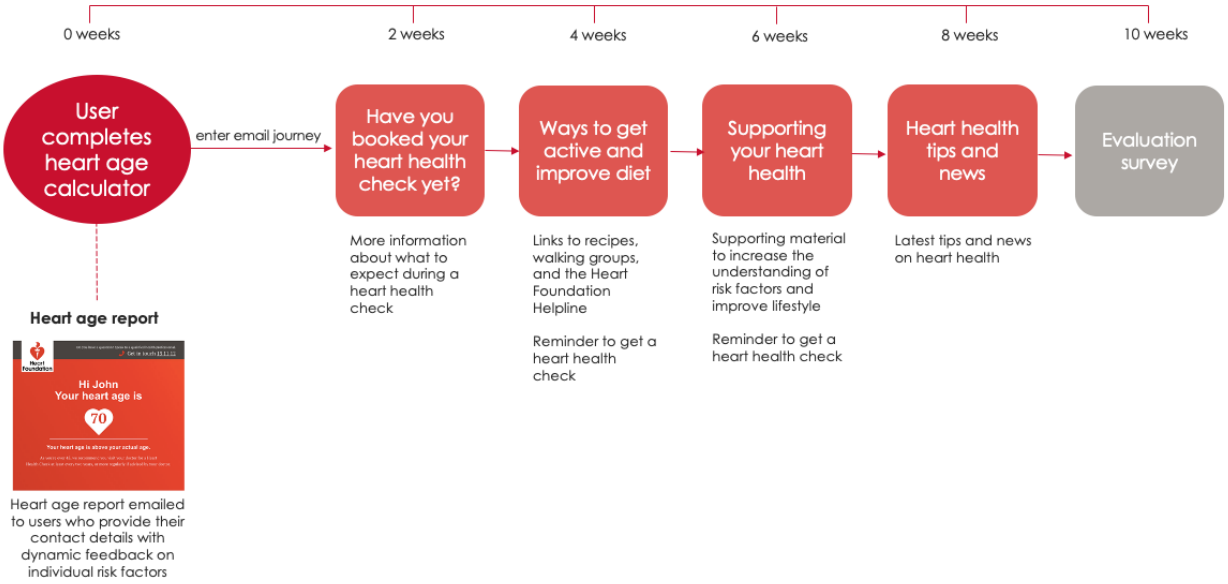
Figure 2. Example heart age calculator report (eg, 54 year old male smoker, family history, diabetes, and average blood pressure/cholesterol).

Procedure

The Australian heart age calculator was launched in February 2019 as part of an NHFA consumer awareness campaign involving mass marketing and media interventions. This *Serial Killer* campaign aimed to boost public awareness around heart disease as Australia's leading cause of death and sought to increase the personal relevance of the condition to all Australian adults. Other campaign objectives included advocating for a range of Heart Foundation federal election requests, including Medicare-funded heart health checks. The web-based heart age calculator URL was included as a *call to action* for this campaign. After completing the heart age calculator, users were prompted to sign up to receive a detailed report via email, further explaining their heart age results (Figure 2). Users were then automatically signed up to a 10-week email journey consisting of fortnightly emails prompting eligible patients to see their GP for a heart health check and providing general advice on healthy

eating, exercise, and heart health (Figure 3). The email journey included various behavior change techniques [16], namely: credible source, prompts/cues, goal setting, information about health consequences, salience of consequences, instruction on how to perform the behavior, social support, and material incentive (reward for completing a lifestyle challenge via an app). At 10 weeks, users were asked to participate in a follow-up survey to evaluate psychological and behavioral outcomes, including recall, positive and negative emotional responses, information seeking, lifestyle change, and clinical checks. Survey respondents entered a draw to win 1 of 3 gift cards. An open response question was also included to evaluate general reactions to heart age. Multimedia Appendix 1 provides all survey questions.

Figure 3. Heart age calculator email journey flowchart.



Participants

The calculator could be completed by Australians aged 35-75 years without CVD, in accordance with the target group for CVD risk assessment in Australia.

Analysis

User data were cleaned to remove duplicates based on internet protocol addresses or email addresses, and users who completed the heart age calculator between February 19 and July 31, 2019, were included in the final data set. CVD risk factors (for all anonymous users), heart age results (for those who requested a report by email), and psychological and behavioral questions (for survey respondents) were linked to the original heart age calculator results. Statistical analysis was performed using IBM SPSS Statistics version 26 (IBM Corp) statistical software package (TB). Descriptive statistics are reported with numbers and percentages for the 3 samples, and exploratory comparisons between age, gender, and heart age category groups in the survey sample were performed using chi-square tests, where a value of $P<.05$ was considered statistically significant. Free text responses to the heart age result were coded using a framework analysis approach where themes were first identified from the data using an inductive approach (C Batcup and C Bonner). We then applied a theoretical framework deductively to organize the themes (C Bonner), before all data were coded under the categories of expectation, experience, risk perception, evaluation, and action (C Batcup). The framework was developed in a previous qualitative study using think-aloud methods to understand how participants use and react to heart age calculators [7]. A sample of 10% was double coded, and discrepancies were resolved through discussion (C Batcup and C Bonner). All authors contributed to the interpretation of the results.

Ethical Approval

An exemption letter was provided by the University of Sydney Human Research Ethics Committee, as the study involved an analysis of existing anonymized data, originally obtained by the NHFA for internal evaluation purposes.

Results

CVD Risk Factors for Anonymous Users, Those Who Requested a Report, and Survey Respondents

Overall, data were obtained from 361,044 anonymous heart age calculator users (CVD risk factors only), 30,279 users who provided email addresses to request a report (heart age results) and 1303 survey respondents (psychological and behavioral questions). Figure 4 shows a sample flowchart, and Table 1 provides a summary of risk factors for the 3 samples. The anonymous user sample was younger (mean 49.37, SD 11.79 years) with a higher proportion of smokers (35,503/361,044, 9.83%), and fewer knew their blood pressure level

(178,281/361,044, 49.38%) and cholesterol level (59,013/361,044, 16.35%), were on blood pressure–lowering medication (64,464/361,044, 17.85%), and reported a family history (123,680/361,044, 34.26%). Of those who provided their email to receive the report and follow-up, there were 19,840 (19,840/30,279, 65.52%) women and 10,439 (10,439/30,279, 34.48%) men, with 80.40% (24,348/30,279) in the target age for heart health checks for the general population (45-75 years), and a mean age of 55.67 (SD 11.43) years. In terms of modifiable risk factors, 6.46% (1957/30,279) of users reported smoking, 40.51% (12,267/30,279) knew their cholesterol level, 66.97% (20,279/30,279) knew their blood pressure level, and 26.26% (7950/30,279) were taking blood pressure–lowering medication. Less than half of the users (12,844/30,279, 42.42%) reported a family history of heart disease, and only 7.56% (2290/30,279) of users reported having a diagnosis of diabetes. The survey respondent sample was older (mean age 60.43, SD 10.15 years), with a lower proportion of smokers (39/1303, 2.99%); and more knew their blood pressure level (961/1303, 73.75%), knew their cholesterol level (585/1303, 44.90%), were on blood pressure–lowering medication (413/1303, 31.70%), and reported a family history (587/1303, 45.04%).

Figure 4. Sample flowchart.

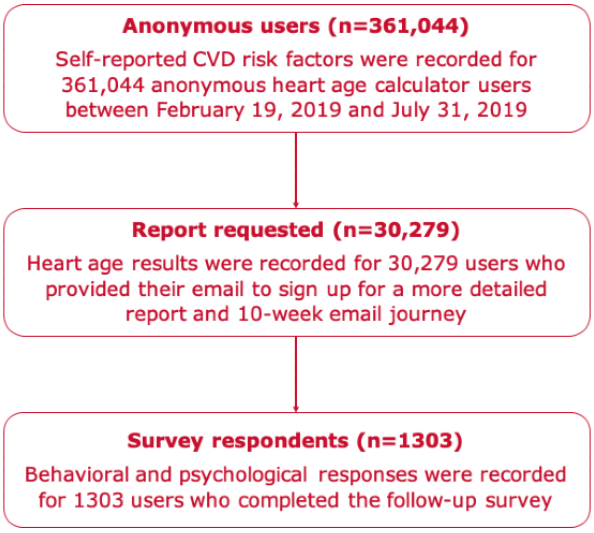


Table 1. Risk factors by heart age calculator user sample.

CVD[a] risk factors	Anonymous users (n=361,044)	Report requested (n=30,279)	Survey respondents (n=1303)
Gender, n (%)			
Female	221,278 (61.29)	19,840 (65.52)	867 (66.54)
Male	139,766 (38.71)	10,439 (34.48)	436 (33.46)
Age (years)			
Mean (SD)	49.37 (11.79)	55.67 (11.43)	60.43 (10.15)
Range, n (%)			
35-44	144,430 (40.00)	5931 (19.59)	112 (8.60)
45-54	88,945 (24.63)	7015 (23.17)	216 (16.58)
55-64	83,313 (23.08)	9809 (32.40)	469 (35.99)
65-75	44,356 (12.29)	7524 (24.85)	506 (38.83)
Smoker, n (%)	35,503 (9.83)	1957 (6.46)	39 (2.99)
Family history of CVD, n (%)	123,680 (34.26)	12,844 (42.42)	587 (45.04)
Diabetes, n (%)	20,606 (5.71)	2290 (7.56)	89 (6.83)
Taking BP[b] medication, n (%)	64,464 (17.85)	7950 (26.26)	413 (31.70)
Know BP level, n (%)	178,281 (49.38)	20,279 (66.97)	961 (73.75)
Know cholesterol level, n (%)	59,013 (16.35)	12,267 (40.51)	585 (44.90)

[a]CVD: cardiovascular disease.

[b]BP: blood pressure.

Heart Age Results for Users Who Requested a Report

Overall, heart age was on average 4.61 years older than current age, including 78.73% (23,840/30,279) with older heart age and 13.75% (4163/30,279) with younger heart age. Heart age results were significantly different by age group (χ^2_6=1601.1; *P*<.001) and gender (χ^2_2=445.2; *P*<.001). Those aged 44-54 years were the most likely group to receive a younger heart age for both women (1380/4640, 29.74%) and men (266/2375, 11.20%), whereas those aged 64-75 years were the most likely group to receive an older heart age result for women (4442/4812, 92.31%) and men (2440/2712, 89.97%). Women were almost twice as likely to receive a younger heart age result than men overall (3242/19,840, 16.34% vs 921/10,439, 8.82%).

Psychological and Behavioral Outcomes for the Survey Respondent Sample

Compared with the total sample that requested a report, survey respondents had a slightly higher proportion of people with an older heart age result (1055/1303, 80.97% vs 23,840/30,279, 78.73%) and a slightly lower proportion of people with a younger heart age result (155/1303, 11.90% vs 4163/30,279, 13.75%), but the rates were similar.

Table 2 summarizes the psychological and behavioral outcomes for the survey respondent sample. Of those who completed the survey 10 weeks after their initial result, most (892/1303, 68.46%) were able to correctly recall their heart age category as being younger, equal to, or older than their current age. This was similar for younger (104/155, 67.09%) and older (735/1055, 69.67%) heart age results, but significantly lower for equal heart age results (53/93, 56.99%; χ^2_2=6.5; *P*=.04). More than one-fourth of users reported feeling a strong positive emotional response (507/1303, 38.91% very motivated and 324/1303, 24.87% very optimistic), and a lower proportion reported strong negative emotions (167/1303, 12.82% very anxious and 160/1303, 12.28% very worried). Compared with those with a younger/equal heart age, users who received an older heart age report were more likely to feel very anxious (159/1055, 15.07% vs 8/248, 3.2% reporting a great deal or a lot; χ^2_1=25.2; *P*<.001) or worried (151/1055, 14.31% vs 9/248, 3.63%; χ^2_1=21.3; *P*<.001). They were less likely to feel optimistic about their result (229/1055, 21.71% vs 95/248, 38.3%; χ^2_1=29.6; *P*<.001), but motivation levels were similar (406/1055, 38.48% vs 101/248, 40.7%; χ^2_1=0.4; *P*=.52).

In terms of lifestyle behavior, more than half of the survey respondents reported improvements in their diet (821/1303, 63.01%) and physical activity (809/1303, 62.09%), with just under half reporting weight loss (643/1303, 49.35%). Almost one-third of users reported reducing stress (412/1303, 31.62%) and alcohol intake (406/1303, 31.16%). Of those who smoked, 48% (19/39) reported reductions. Some lifestyle change behaviors were reported at higher rates for those with older compared with younger/equal heart age, including diet (680/1055, 64.45% vs 141/248, 56.8%; χ^2_1=5.0; *P*=.03) and weight loss (537/1055, 50.90% vs 106/248, 42.7%; χ^2_1=5.4; *P*=.02).

For outcomes relating to clinical risk assessment, almost half of the users had already seen their GP (621/1303, 47.66%), and one-fourth reported receiving a heart health check (362/1303, 27.78%) in the 10 weeks since receiving their heart age report. Higher proportions had obtained specific clinical tests, with three-fourths of the users checking blood pressure level and more than half obtaining blood tests for cholesterol (737/1303, 56.56%) and diabetes or sugar levels (697/1303, 53.49%). People with an older heart age result were more likely to have visited their doctor (538/1055, 51.00% vs 83/248, 33.4%; χ^2_1=24.7; *P*<.001) or had a heart health check (314/1055, 29.76% vs 48/248, 19.3%; χ^2_1=10.8; *P*<.001), compared with those with a younger or equal heart age.

Table 2. Heart age calculator user outcomes after 10 weeks for survey respondents.

Outcomes	All survey respondents (n=1303), n (%)	Older heart age (n=1055), n (%)	Younger or equal heart age (n=248), n (%)
Psychological			
Recall of correct heart age category	892 (68.4)	735 (69.6)	157 (63.3)
Very motivated (a great deal/a lot)	507 (38.9)	406 (38.4)	101 (40.7)
Very optimistic (a great deal/a lot)	324 (24.8)	229 (21.7)	95 (38.3)
Very anxious (a great deal/a lot)	167 (12.8)	159 (15.0)	8 (3.2)
Very worried (a great deal/a lot)	160 (12.2)	151 (14.3)	9 (3.6)
Spoke to family about familial history	555 (42.5)	466 (44.1)	89 (35.8)
Found out more information	787 (60.4)	669 (63.4)	118 (47.5)
Told others about the calculator	492 (37.7)	397 (37.6)	95 (38.3)
Lifestyle change			
Increased physical activity	809 (62.0)	668 (63.3)	141 (56.8)
Lost weight	643 (49.3)	537 (50.9)	106 (42.7)
Improved diet	821 (63.0)	680 (64.4)	141 (56.8)
Reduced or quit smoking	19 (48.7)	19 (48.7)	0 (0.0)
Reduced stress	412 (31.6)	333 (31.56)	79 (31.85)
Limited alcohol intake	406 (31.1)	332 (31.4)	74 (29.8)
Clinical risk assessment			
Saw general practitioner	621 (47.6)	538 (51.0)	83 (33.4)
Had a heart health check up	362 (27.7)	314 (29.7)	48 (19.3)
Had a blood pressure check	976 (74.9)	809 (76.6)	167 (67.3)
Had a blood test for cholesterol	737 (56.5)	619 (58.6)	118 (47.5)
Had a test for diabetes or sugar levels	697 (53.4)	587 (55.6)	110 (44.3)

Qualitative Responses to Heart Age for the Survey Respondent Sample

The 1077 open response comments were coded and organized into 5 themes from a previous qualitative study on the process of heart age calculator use [7]: (1) participants' expectations of what the result would show, (2) their experience of seeing the result, (3) what they understood about their risk based on the results, (4) their evaluation of the result as a credible source of information, and (5) actions they were prompted to take. These themes and their subthemes are shown in Table 3, along with sample quotes. Those with older heart age tended to show more concern about the result and considered it as an indication of ill health or a need for change, although some thought it was not a problem or disregarded the result. Those with younger or equal heart age described feeling happier about their result and considered their health to be good whereas some wanted it to be lower. Many participants also provided reasons for why they received their result, citing a variety of factors such as fitness levels, genetics, and poor overall health. Common actions prompted by the heart age calculator included arranging a GP consultation, changing diet, or increasing physical activity.

Table 3. Themes identified in open responses to heart age results.

Themes and subthemes	Example quotes
Expectations	
Perception of lifestyle	• "I'm a bit unsure why as I exercise regularly, don't smoke only drink occasionally, within normal weight range"
Information from doctor	• "I had only just had an appointment with my cardiologist and he said my heart is very good"
Experience	
Happy or fine with result	• "I'm on the right track"
Surprise at result	• "Surprised and puzzled as to true meaning"
Concerned or disappointed	• "I was quite shocked and worried"
Defensive at result	• "How is it possible as I had the best possible score therefore everybody must be above"
Focus on age or being old	• "It still feels old!"
No impression	• "I didn't think that it had any relevance to me"
Risk perception	
Indicates good health	• "I assumed it meant that my heart was probably in good condition for my age"
Indicates health issues	• "I have a higher than average chance of having a heart attack or a stroke"
Unsure of meaning	• "Don't really know"
Inconsistent with heart age category	• "That I am healthier than average (older heart age result)." • "I am unhealthy (younger heart age result)"
Interpretation reflects heart age category	• "Heart is older than my age (older heart age result)." • "My heart is in better health than it's [SIC] actual age (younger heart age result)"
Current or heart age discrepancy	• "79 was extremely scary for a 67 year old (older heart age result)." • "It was only one year younger, so it was good, but not great (younger heart age result)"
Evaluation	
Incorrect or mistrust result	• "The assessment tool was too simplistic to be reliable"
Expected result	• "I was aware that this would probably be the case"
Risk factors too limited	• "I was annoyed as the questions were quite limited and did not take account lifestyle and medications"
Family history or genetics	• "I thought it was elevated because of my family history because I otherwise take good care of my health"
Explain result	• "I was not eating properly and exercising enough"
Action	
No motivation to change	• "I'm on track with my general health"
Need to change	• "I thought it meant I had to do some work to get it back to my right age or lower"
See a doctor	• "That I needed to see a doctor"
Reflection on life	• "An aged heart that hasn't been well taken care of. A wake up call to nuture [*SIC*] it and the rest of me"

Discussion

Principal Findings

This paper is the first report of a national Australian sample of heart age calculator users. It contributes to the broader heart age outcomes literature with a larger sample of population users, who requested a report with follow-up to support behavior change over a 10-week period. In line with other tools used internationally and in the United Kingdom [3,4], the majority of users who requested a heart age report received an older heart age, but many people did not know their cholesterol or blood pressure levels; therefore, the risk assessment was often based on the population average. More than half of the survey respondents reported lifestyle changes after using the heart age calculator, and many reported seeking further information, including clinical checks to receive a more reliable risk assessment, particularly if they received an older heart age. Interestingly, changes were also reported by those who received younger and equal heart age results. This aligns with prior qualitative research, showing that the process of using heart age calculators can prompt the consideration of lifestyle changes regardless of the actual result [7]. This finding could be because of the process of receiving a heart age result regardless of its value or alternatively the survey sample could have been more motivated in general, that is, they were interested in lifestyle changes before the heart age result. However, a previous randomized study found psychological differences between the heart age and control groups that were similar for younger and older heart age results, suggesting that there is something about receiving this risk format that does prompt different reactions regardless of the result itself [10].

The Australian heart age calculator website has been accessed by a large number of people, with 1.3 million users engaged in the first year (internal figures from the NHFA). This paper shows that older people, those more likely to know their risk factors and/or take medication and nonsmokers, were more likely to engage further in health promotion activities via a digital follow-up report. This points to the need for additional strategies to engage people with unknown risk factors and some high-risk groups. Alternative biological age concepts such as *lung age* may be more effective for engaging specific groups, such as younger smokers [17]. The high prevalence of web-based health risk calculators (eg, CVD, diabetes, and cancer [18-20]) shows that this is a popular marketing strategy or call to action, which may be effective if targeted to the right audience and backed up with behavior change support programs. Systematic reviews show that risk communication can increase intentions to change health-related behavior, and effects on behavior can be enhanced by addressing several aspects of risk perception and repeated communication [21,22]. However, additional behavior change techniques may be needed to bridge the intention-behavior gap and maintain changes over time, such as action plans that incorporate implementation intentions [23,24].

As found in previous experiments comparing heart age to absolute risk [10,11], this sample reported high recall of the heart age result category and strong emotional responses to concepts such as worry about older heart age or optimism. Those who received equal heart age were less likely to remember this, suggesting that positive responses to young heart age or negative responses to older heart age may reinforce the result and aid recall. Previous research has also raised the issue of credibility [7,10], which was reflected in the thematic analysis of open responses. Users had many questions about the role of additional risk factors, conflicting information from health professionals, and the reliability of the web-based assessment, particularly when the result was unexpected. This led some participants to question the usefulness of the heart age calculator but prompted others to seek clinical assessments or lifestyle changes. For heart age calculator developers, it may be important to explain how and why different risk factors are used for those who want more explanation and to clearly state the need to see a doctor for a more accurate risk assessment.

The launch of the Australian heart age calculator was part of a broader campaign to address barriers to absolute CVD risk assessment, including lobbying for federal government funding of clinical heart health checks. More than half of the survey respondents reported having seen their GP in the 10 weeks since finding out their heart age, and one-fourth of users reported receiving a heart health check. Most users were eligible for a full CVD risk assessment with their GP in line with clinical guidelines targeting those aged 45-74 years. The barriers to engage otherwise healthy adults in preventive health checks are complex, covering all 3 broad determinants of behavior change: capability (eg, lack of knowledge and awareness), opportunity (eg, time and access constraints), and motivation (eg, aversion to preventive medicine) [25,26]. Heart age calculators may be particularly useful for addressing awareness issues and motivating people to see their doctor for a more accurate clinical assessment, but this needs to be supported by the broader health system to address opportunity barriers. In Australia, a combination of strategies has led to more than 100,000 Australians receiving a heart health check from their GP under the Medicare Benefits Scheme in the 12 months since this heart age calculator was launched [27].

Further research is needed to determine whether the behavioral outcomes of heart age calculators can be improved by linking it to additional behavior change strategies known to improve lifestyle change (eg, action planning) [16] and whether absolute risk formats used in clinical practice can be equally engaging [28]. There is very little research comparing different labels for the general concept of *biological age*, but one study has found differences in the way that young people interpret *heart age* compared with *fitness age* even when the same numerical age result is used [29]. Different target populations may respond best to different labels, and it is important to consider potential harms from misunderstandings as well as the potential for positive behavior changes. Further investigation is also needed to address the information needs of people with lower health literacy, who have fewer skills required to access, understand, and act on health information [30]. Different interactive tools may be needed for different patient populations to enable informed consent about CVD management options, such as patient decision aids with actionable values clarification

exercises to help people weigh lifestyle approaches compared with medication recommended by a doctor [31].

Strengths and Limitations

The main strength of this study is the analysis of a more engaged sample than other national/international heart age user reports (excluding repeat and nonserious users), but the survey respondents are likely biased in terms of motivation and have different characteristics to the broader samples in this study. As there was no randomization, we could not determine causation or efficacy of heart age over other risk communication methods or the length of follow-up required for sustained lifestyle change. The descriptive data available could not be used to determine whether the heart age result itself caused behavior change or whether it simply promoted engagement with further behavior change strategies.

In conclusion, the results confirm high public interest in heart age tools as a way to engage people in the target age for CVD risk assessment and prevention activities, with the potential to prompt clinical risk assessments and lifestyle changes for many users. Supporting the initial heart age result with more detailed reports to explain the results and evidence-based behavior change techniques may improve the effectiveness of these tools.

Acknowledgments

This paper was not specifically funded, but C Bonner was supported by a National Health and Medical Research Council/ NHFA fellowship and C Batcup was supported by an NHFA Vanguard Grant for a separate but related study involving a heart age consumer engagement tool led by C Bonner.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Follow up survey items.
[PDF File (Adobe PDF File), 122 KB - jmir_v22i8e19028_app1.pdf]

References

1. Bonner C, Bell K, Jansen J, Glasziou P, Irwig L, Doust J, et al. Should heart age calculators be used alongside absolute cardiovascular disease risk assessment? BMC Cardiovasc Disord 2018 Feb 7;18(1):19 [FREE Full text] [doi: 10.1186/s12872-018-0760-1] [Medline: 29409444]
2. Wells S, Kerr A, Eadie S, Wiltshire C, Jackson R. 'Your heart forecast': a new approach for describing and communicating cardiovascular risk? Heart 2010 May 27;96(9):708-713. [doi: 10.1136/hrt.2009.191320] [Medline: 20424153]
3. Neufingerl N, Cobain MR, Newson RS. Web-based self-assessment health tools: who are the users and what is the impact of missing input information? J Med Internet Res 2014 Sep 26;16(9):e215 [FREE Full text] [doi: 10.2196/jmir.3146] [Medline: 25261155]
4. Patel RS, Lagord C, Waterall J, Moth M, Knapton M, Deanfield JE. Online self-assessment of cardiovascular risk using the joint British societies (JBS3)-derived heart age tool: a descriptive study. BMJ Open 2016 Sep 28;6(9):e011511 [FREE Full text] [doi: 10.1136/bmjopen-2016-011511] [Medline: 27683512]
5. Yang Q, Zhong Y, Ritchey M, Cobain M, Gillespie C, Merritt R, et al. Vital signs: predicted heart age and racial disparities in heart age among US adults at the state level. MMWR Morb Mortal Wkly Rep 2015 Sep 4;64(34):950-958 [FREE Full text] [doi: 10.15585/mmwr.mm6434a6] [Medline: 26335037]
6. Shi R, Lan Y, Yu W. A feasibility study on 10-year CVD risk assessment as a primary prevention tool for cardiovascular disease. Value Health 2017;20(5):A275. [doi: 10.1161/circoutcomes.10.suppl_3.051]
7. Bonner C, Jansen J, Newell BR, Irwig L, Glasziou P, Doust J, et al. I don't believe it, but I'd better do something about it: patient experiences of online heart age risk calculators. J Med Internet Res 2014 May 5;16(5):118-129 [FREE Full text] [doi: 10.2196/jmir.3190] [Medline: 24797339]
8. Jackson R, Kerr A, Wells S. 'Should we reconsider the role of age in treatment allocation for primary prevention of cardiovascular disease?' No, but we can improve risk communication metrics. Eur Heart J 2017 May 21;38(20):1548-1552. [doi: 10.1093/eurheartj/ehw322] [Medline: 27436864]
9. Grover SA, Lowensteyn I, Joseph L, Kaouache M, Marchand S, Coupal L, Cardiovascular Health Evaluation to Improve Compliance and Knowledge Among Uninformed Patients (CHECK-UP) Study Group. Patient knowledge of coronary risk profile improves the effectiveness of dyslipidemia therapy: the CHECK-UP study: a randomized controlled trial. Arch Intern Med 2007 Nov 26;167(21):2296-2303. [doi: 10.1001/archinte.167.21.2296] [Medline: 18039987]
10. Bonner C, Jansen J, Newell BR, Irwig L, Teixeira-Pinto A, Glasziou P, et al. Is the 'heart age' concept helpful or harmful compared to absolute cardiovascular disease risk? An experimental study. Med Decis Making 2015 Nov;35(8):967-978. [doi: 10.1177/0272989X15597224] [Medline: 26251465]

11. Soureti A, Hurling R, Murray P, van Mechelen W, Cobain M. Evaluation of a cardiovascular disease risk assessment tool for the promotion of healthier lifestyles. Eur J Cardiovasc Prev Rehabil 2010 Oct;17(5):519-523. [doi: 10.1097/HJR.0b013e328337ccd3] [Medline: 20195154]
12. Witteman HO, Fuhrel-Forbis A, Wijeysundera HC, Exe N, Dickson M, Holtzman L, et al. Animated randomness, avatars, movement, and personalization in risk graphics. J Med Internet Res 2014 Mar 18;16(3):294-313 [FREE Full text] [doi: 10.2196/jmir.2895] [Medline: 24642037]
13. Kulendrarajah B, Grey A, Nunan D. How effective are 'age' tools at changing patient behaviour? A rapid review. BMJ Evid Based Med 2020 Apr;25(2):1-2. [doi: 10.1136/bmjebm-2019-111244] [Medline: 31558486]
14. Australian Bureau of Statistics, Australian Government. Microdata: Australian Health Survey, National Health Survey, 2011-12. 2011. URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4324.0.55.001Main+Features492011-12?OpenDocument [accessed 2020-04-26]
15. Heart Age Calculator. National Heart Foundation of Australia. 2019. URL: https://www.heartfoundation.org.au/heart-age-calculator [accessed 2020-07-30]
16. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
17. Parkes G, Greenhalgh T, Griffin M, Dent R. Effect on smoking quit rate of telling patients their lung age: the Step2quit randomised controlled trial. Br Med J 2008 Mar 15;336(7644):598-600 [FREE Full text] [doi: 10.1136/bmj.39503.582396.25] [Medline: 18326503]
18. Waters EA, Sullivan HW, Nelson W, Hesse BW. What is my cancer risk? How internet-based cancer risk assessment tools communicate individualized risk estimates to the public: content analysis. J Med Internet Res 2009 Jul 31;11(3):e33 [FREE Full text] [doi: 10.2196/jmir.1222] [Medline: 19674958]
19. Bonner C, Fajardo MA, Hui S, Stubbs R, Trevena L. Clinical validity, understandability, and actionability of online cardiovascular disease risk calculators: systematic review. J Med Internet Res 2018 Feb 1;20(2):e29 [FREE Full text] [doi: 10.2196/jmir.8538] [Medline: 29391344]
20. Fajardo MA, Balthazaar G, Zalums A, Trevena L, Bonner C. Favourable understandability, but poor actionability: an evaluation of online type 2 diabetes risk calculators. Patient Educ Couns 2019 Mar;102(3):467-473. [doi: 10.1016/j.pec.2018.10.014] [Medline: 30389187]
21. Sheeran P, Harris PR, Epton T. Does heightening risk appraisals change people's intentions and behavior? A meta-analysis of experimental studies. Psychol Bull 2014 Mar;140(2):511-543. [doi: 10.1037/a0033065] [Medline: 23731175]
22. Sheridan SL, Viera AJ, Krantz MJ, Ice CL, Steinman LE, Peters KE, Cardiovascular Health Intervention Research and Translation Network Work Group on Global Coronary Heart Disease Risk. The effect of giving global coronary risk information to adults: a systematic review. Arch Intern Med 2010 Feb 8;170(3):230-239. [doi: 10.1001/archinternmed.2009.516] [Medline: 20142567]
23. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
24. Bélanger-Gravel A, Godin G, Amireault S. A meta-analytic review of the effect of implementation intentions on physical activity. Health Psychol Rev 2013 Mar;7(1):23-54. [doi: 10.1080/17437199.2011.560095]
25. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
26. Harte E, MacLure C, Martin A, Saunders CL, Meads C, Walter FM, et al. Reasons why people do not attend NHS health checks: a systematic review and qualitative synthesis. Br J Gen Pract 2018 Jan;68(666):e28-e35 [FREE Full text] [doi: 10.3399/bjgp17X693929] [Medline: 29203682]
27. Requested Medicare Items Processed From April 2019 to April 2020. Services Australia - Statistics - Item Reports. URL: http://medicarestatistics.humanservices.gov.au/statistics/do.jsp?PROGRAM=%2Fstatistics%2Fmbs_item_standard_report&DRILL=ag&group=17%2C19&VAR=services&STAT=count&RPT_FMT=by+state&PTYPE=month&START_DT=201904&END_DT=202004 [accessed 2020-06-24]
28. Bonner C, Fajardo MA, Doust J, McCaffery K, Trevena L. Implementing cardiovascular disease prevention guidelines to translate evidence-based medicine and shared decision making into general practice: theory-based intervention development, qualitative piloting and quantitative feasibility. Implement Sci 2019 Aug 30;14(1):86 [FREE Full text] [doi: 10.1186/s13012-019-0927-x] [Medline: 31466526]
29. van der Pol-Harney E, Bonner C, Turner R, McCaffery K. The Effects of Communicating CVD Risk as 'Fitness Age'. In: Proceedings of the 14th International Congress of Behavioral Medicine. 2016 Presented at: ISBN'16; December 7-10, 2016; Melbourne, Australia p. 7-10.
30. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Prom Int 2000;15(3):259-267. [doi: 10.1093/heapro/15.3.259]
31. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4:CD001431 [FREE Full text] [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]

Abbreviations

BP: blood pressure
CVD: cardiovascular disease
GP: general practitioner
NHFA: National Heart Foundation of Australia

Edited by G Eysenbach; submitted 01.05.20; peer-reviewed by D Silveira, E Neter; comments to author 12.06.20; revised version received 24.06.20; accepted 25.06.20; published 07.08.20.

Please cite as:
Bonner C, Raffoul N, Battaglia T, Mitchell JA, Batcup C, Stavreski B
Experiences of a National Web-Based Heart Age Calculator for Cardiovascular Disease Prevention: User Characteristics, Heart Age Results, and Behavior Change Survey
J Med Internet Res 2020;22(8):e19028
URL: https://www.jmir.org/2020/8/e19028
doi:10.2196/19028
PMID:32763875

Original Paper

Communicating Uncertainty in Written Consumer Health Information to the Public: Parallel-Group, Web-Based Randomized Controlled Trial

Roland B Büchter[1], MPH; Cornelia Betsch[2], PhD; Martina Ehrlich[1], Dr oec troph; Dennis Fechtelpeter[1], MSc; Ulrich Grouven[1], Dr rer biol hum; Sabine Keller[1], Dip Soz; Regina Meuer[1], MPH; Constanze Rossmann[2], PhD; Andreas Waltering[1], MD

[1]Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany

[2]Media and Communication Science, University of Erfurt, Erfurt, Germany

Corresponding Author:
Roland B Büchter, MPH
Institute for Quality and Efficiency in Health Care (IQWiG)
Im Mediapark 8
Cologne, 50670
Germany
Phone: 49 221 35685 ext 413
Email: roland.buechter@iqwig.de

Abstract

Background: Uncertainty is integral to evidence-informed decision making and is of particular importance for preference-sensitive decisions. Communicating uncertainty to patients and the public has long been identified as a goal in the informed and shared decision-making movement. Despite this, there is little quantitative research on how uncertainty in health information is perceived by readers.

Objective: The aim of this study was to examine the impact of different uncertainty descriptions regarding the evidence for a treatment effect in a written research summary for the public.

Methods: We developed 8 versions of a research summary on a fictitious drug for tinnitus with varying degrees (Q1), sources (Q2), and magnitudes of uncertainty (Q3). We recruited 2099 members of the German public from a web-based research panel. Of these, 1727 fulfilled the inclusion criteria and were randomly presented with one of these research summaries. Randomization was conducted by using a centralized computer with a random number generator. Web-based recruitment and data collection were fully automated. Participants were not aware of the purpose of the study and alternative presentations. We measured the following outcomes: perception of the treatment effectiveness (primary), certainty in the judgement of treatment effectiveness, perception of the body of evidence, text quality, and intended decision. The outcomes were self-assessed.

Results: For the primary outcome, we did not find a global effect for Q1 and Q2 (P=.25 and P=.73), but we found a global effect for Q3 (P=.048). Pairwise comparisons showed a weaker perception of treatment effectiveness for the research summary with 3 sources of uncertainty compared to the version with 2 sources of uncertainty (P=.04). Specifically, the proportion of the participants in the group with 3 sources of uncertainty that perceived the drug as possibly beneficial was 9% lower than that of the participants in the group with 2 sources of uncertainty (92/195, 47.2% vs 111/197, 56.3%, respectively). The proportion of the participants in the group with 3 sources of uncertainty that considered the drug to be of unclear benefit was 8% higher than that of the participants in the group with 2 sources of uncertainty (72/195, 36.9% vs 57/197, 28.9%, respectively). However, there was no significant difference compared to the version with 1 source of uncertainty (P=.31). We did not find any meaningful differences between the research summaries for the secondary outcomes.

Conclusions: Communicating even a large magnitude of uncertainty for a treatment effect had little impact on the perceived effectiveness. Efforts to improve public understanding of research are needed to improve the understanding of evidence-based health information.

Trial Registration: German Clinical Trials Register DRKS00015911, https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00015911

International Registered Report Identifier (IRRID): RR2-10.2196/13425

(J Med Internet Res 2020;22(8):e15899) doi:10.2196/15899

KEYWORDS

uncertainty; consumer health information; decision making

Introduction

Uncertainty pervades health care and is integral to evidence-informed decision making. Helping patients and consumers to understand and deal with uncertainty has been identified as one of the goals of the shared decision making and informed choices movement [1]. Understanding uncertainty is of particular importance for preference-sensitive decisions, that is, when there is a close trade-off between benefits and harms and patient values and preferences are highly variable. However, there are many difficulties in communicating uncertainty. Information providers often have to decide which of the many sources of uncertainty are the most relevant to patients. Examples include risk of bias, statistical uncertainty (imprecision), and lack of generalizability (indirectness). Selection is required to prevent information overload, which can prompt people to base their decisions on heuristics rather than on evidence [2]. Furthermore, communication of uncertainty may also have detrimental effects, for example, by hampering understanding or decreasing the credibility of the information provider [3,4].

Quantitative research on how to communicate uncertainty regarding the benefits and harms of treatments to patients and the public is limited. A systematic review by the Agency for Healthcare Research and Quality identified 8 controlled studies with 9 comparisons, including 6 randomized trials [5]. Four of these studies examined quantification of statistical uncertainty, 4 studied different ways of communicating net benefit, and 1 addressed uncertainty arising from the use of a surrogate outcome. These studies were very heterogeneous in terms of the context (cancer screening, treatment decision making, etc), interventions (written information, drug fact boxes, multifaceted interventions etc), and outcomes (risk perception, decision making, etc).

From our experience in providing evidence-based health information to the German public, enabling informed decisions often requires providing information on other types of uncertainty. For example, a patient with subacromial pain syndrome may wonder whether surgery using subacromial decompression may help with his/her condition. Two randomized, sham-controlled studies have shown no benefit of this treatment. However, few people in these trials had a hooked type III acromion. Thus, there is uncertainty whether this subgroup of patients may benefit from such a treatment. Whether such information is presented or not may impact readers in different ways and affect their choices. We are not aware of any studies investigating whether the perception of treatment effectiveness depends on the degree, type, and magnitude of uncertainty presented in written health information.

In order to test the main way in which we communicate uncertainty, we conducted a study addressing the following 3 questions.

Q1. Degree of uncertainty: Do members of the public perceive treatment effects differently depending on the certainty of the treatment effect (higher versus lower), which are expressed by different wordings (“studies show” or “studies indicate”)?

Q2. Type of uncertainty: Do members of the public interpret uncertain treatment effects differently depending on the type of uncertainty (publication bias, indirectness, imprecision)?

Q3. Number of sources of uncertainty: Is there an additive effect when multiple sources of uncertainty are presented (1-3 sources of uncertainty)?

We hypothesized that higher degrees of uncertainty expressed through different wordings or a larger magnitude of uncertainty reduce the perceived effectiveness of a treatment, the certainty in this judgement, and the intention to use the treatment. We further hypothesized that uncertainties due to vested interests/publication bias reduce perceived effectiveness to a larger extent than other sources of uncertainty. Lastly, we hypothesized that the ratings of text quality decrease with an increasing magnitude of uncertainty.

We investigated the 3 research questions by using 8 variations of a research summary for a fictitious drug for tinnitus that was set in a hypothetical treatment decision scenario. We designed the experiment as a randomized superiority study with 8 parallel groups allocated in an equal ratio (between-group design). This trial has been registered in the German Clinical Trials Register (DRKS00015911). The study protocol has been published elsewhere and includes the technical details of the study conduct [6].

Methods

Recruitment

We recruited members of the German public from a web-based research panel. Panel members were eligible if they were at least 18 years of age and able to read and write German. There were no other restrictions. We used a quota to ensure equal representation of different age groups (below and above 45 years) and sex. Once a quota cell was full, enrolment for this quota was closed.

Study Procedure

After enrolment, the participants were provided with a short introduction to the study and an informed consent sheet. Panel members who agreed to participate were asked for information on age, sex, and educational degree based on the German school system (none/basic secondary/higher secondary/general entry qualification for university/university degree). We then asked them to imagine having tinnitus and having unsuccessfully tried several treatments. Participants were then randomly presented with one of the 8 versions of the research summary. These contained brief information on the medical condition and a short summary of the evidence for a fictitious new tinnitus medication

("Oroxil"). After the presentation of the research summaries, we collected data on the different outcomes by using a questionnaire developed for the purpose of this study. We allowed participants to refer to the research summary as needed while answering the questions. Before the conclusion of the experiment, participants were asked about their profession and history of tinnitus.

Interventions

We developed 8 variations of the research summary based on standards and use of language in providing evidence-based health information to consumers through Germany's statutory health website [7]. According to the objectives of the study, we systematically modified the research summary regarding the degree of uncertainty, the sources of uncertainty, and the magnitude of uncertainty. None of the other parts of the research summaries were altered between variations, including the numbers.

For Q1, we compared 3 alternative wordings for the expression of uncertainty. Version A describes a *certain* treatment effect and version B a *possible but not a certain* treatment effect (indication of effect). Version B1 was identical to B, but it included an additional statement on the *need for further research*. The wordings are based on the methods for the assessment of treatment benefits developed by the German Institute for Quality and Efficiency in Health Care (IQWiG) [8].

For Q2, we drew on the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) framework to categorize different sources of uncertainty. According to GRADE, uncertainty can arise from risk of bias, (unexplained) inconsistency, indirectness (such as lack of generalizability or use of a surrogate outcome), imprecision, and other threats to validity such as publication bias or vested interests [9]. We therefore compared 3 additional research summaries describing publication bias/vested interests (B2), indirectness (B3), and imprecision (B4). We operationalized these sources of uncertainty by using everyday language (Table 1). We also included the variation B1 into this comparison.

For Q3, we compared a combination of 1, 2, or 3 sources of uncertainty (B4 vs B42 containing B4 and B2 vs B432 containing B4, B3, and B2).

This resulted in 8 variations of the research summary, 2 of which were used in 2 statistically independent comparisons (Table 1). An exemplary version of the research summary is provided in the Multimedia Appendix 1; this appendix shows the original German version that was translated into English for this publication. Multimedia Appendix 2 shows the original German version of the research summaries.

Table 1. Variations of the research summaries used to examine the 3 overarching research questions (translated from German).

Questions, Group identifier	Variations examined	Variations in the text
Q1: Degree of uncertainty		
A	Effect shown	Studies show that Oroxil can reduce tinnitus.
B	Indication of effect	Studies indicate that Oroxil may reduce tinnitus.
B1	Indication of effect with general explanation	Studies indicate that Oroxil may reduce tinnitus. (…) The pros and cons of Oroxil cannot be fully judged, however. This requires further research.
Q2: Type of uncertainty		
B1	Indication of effect with general explanation	As above
B2	Publication bias/vested interests	Studies indicate that Oroxil may reduce tinnitus. (…) The pros and cons of Oroxil cannot be fully judged, however. The reason for this is that the company that developed the drug has not published all of the studies on Oroxil.
B3	Indirectness (population)	Studies indicate that Oroxil may reduce tinnitus. (…) The pros and cons of Oroxil cannot be fully judged, however. The reason for this is that people who took part in the study were exposed to loud noises at work. It is uncertain whether the results also apply to other people.
B4	Imprecision (small sample size)	Studies indicate that Oroxil may reduce tinnitus. (…) The pros and cons of Oroxil cannot be fully judged, however. The reason for this is that only a small number of people took part in the studies.
Q3: Number of sources of uncertainty		
B4	Imprecision (small sample size)	As above
B42	Publication bias/vested interests and imprecision	Studies indicate that Oroxil may reduce tinnitus. (…) The pros and cons of Oroxil cannot be fully judged, however. The reason for this is that only a small number of people took part in the studies. Furthermore, the company that developed the drug has not published all of the studies on Oroxil.
B432	Publication bias/vested interests and imprecision and indirectness	Studies indicate that Oroxil may ease tinnitus. (…) The pros and cons of Oroxil cannot be fully judged, however. The reason for this is that the studies were small. Furthermore, people who took part in the study were exposed to loud noises at work. It is uncertain whether the results also apply to other people. Lastly, the company that developed the drug has not published all studies on Oroxil.

Measurements

Our primary outcome was the perception of treatment effectiveness. Secondary outcomes were subjective certainty in the judgement of treatment effectiveness, perception of the body of evidence, intended decision, and perception of text quality. The outcomes were self-assessed through a web-based questionnaire. The outcome measures were developed and pretested by the author team if not stated otherwise.

The perception of treatment effectiveness was measured with 1 item on an ordinal scale (*How do you judge the effectiveness of Oroxil?*) with 5 possible answers: (1) it is proven that Oroxil can help, (2) Oroxil may possibly help, (3) it is unclear whether Oroxil helps, (4) Oroxil may not help, and (5) Oroxil definitely does not help.

Subjective certainty in the judgement of treatment effectiveness (*How certain do you feel in making this judgement?*) was measured on a 5-point Likert scale ranging from *not certain at all* to *very certain*. As this relates to the first question on the perception of treatment effectiveness, data on this item were gathered immediately after answering the first question.

The perception of the body of evidence was measured with a 6-item semantic differential scale with each item measured on a 5-point Likert scale. Participants were asked to rate the body of evidence as certain to uncertain, reliable to unreliable, valid to not valid, generalizable to not generalizable, excellent to poor, and trustworthy to untrustworthy.

The intended decision was measured using 1 item (*How would you decide?*) measured on a 5-point Likert scale with 2 poles: (1) definitely not take Oroxil and (2) definitely take Oroxil.

The perception of text quality was measured with a 9-item semantic differential scale based on previous literature [10,11]. The construct includes the following items measured on a 5-point Likert scale: interesting to uninteresting, balanced to 1-sided, comprehensible to incomprehensible, credible to

incredible, clear to unclear, well done to not well done, professional to unprofessional, appealing to not appealing, and respectable to not respectable.

For the secondary outcomes *perception of the body of evidence* and *text quality*, we combined the items of each of the scales into 1 index by averaging their values, where a higher value indicates better perception of the body of evidence or text quality, respectively. The internal consistency was good with a Cronbach alpha of .81 for both indices.

We piloted a paper-and-pencil-version of the questionnaire and 2 versions of the research summary in a convenience sample of 40 students to test and optimize the reliability of the constructs, comprehensibility of the instructions, the stimuli, and the questions and to gather data for the sample size calculation.

Randomization and Data Collection

Participants were recruited through a national web-based access panel run by a professional survey firm that provides incentives to participants (Bilendi). The data collection was run by the Survey Centre Bonn (uzbonn–Gesellschaft für empirische Sozialforschung und Evaluation), a spinout company of the Center for Evaluation and Methods at the University of Bonn. Unicom Intelligence survey software (formerly IBM SPSS Data Collection) was used for randomization and data collection [6]. Panel members were allocated after they had answered the demographic questions and were computer-checked for eligibility. As the experiment was web-based by using an automated and centralized computer system, the allocation sequence was concealed from the investigators and the data collectors. Participants were not aware of the purpose of the study and were not told about alternative presentations.

Sample Size Calculation and Statistical Analysis

We calculated the sample size for the present study based on an effect size of Cohen f=0.15 derived from the pretest, a significance level of 5%, a power of 90%, and adjusted for use of a nonparametric test. This resulted in 159 participants per group. To allow some dropouts, we decided to randomize 1500 participants, equaling an average of 187.5 participants per group. The details of the sample size calculations are presented in the study protocol [6].

For the statistical analysis of the primary and secondary outcomes, we conducted Kruskal-Wallis tests to test for overall differences between the groups within each of our 3 primary study questions. We chose to use a nonparametric test to account for the types of scales used (ordinal scaling or unequal distances between items). In case of a significant overall difference, we conducted a multiple testing procedure to perform pairwise comparisons between the groups by means of the Dwass-Steel-Critchlow-Fligner multiple comparison analysis, which is based on pairwise 2-sample Wilcoxon comparisons. All statistical tests were 2-sided and performed using a 5% significance level. While we used nonparametric tests, data for the primary outcomes are presented as the proportion of responders for each possible answer and as means and standard deviations for the secondary outcomes in order to be more informative. To test the robustness of the findings, we repeated all the analyses by using analysis of variance (ANOVA). For this, we scored the primary outcome as a 5-point Likert scale in line with the other outcomes.

All participants were analyzed in the originally assigned group. Because we collected outcome data immediately after the presentation of the research summaries and because panel members were required to finish the questionnaire to receive their incentive, we had no major concerns regarding missing data. Therefore, we did not plan to employ any imputation methods and conducted all analyses on the data available. The statistical analysis was conducted by a statistician from the Medical Biometry Department at IQWiG with SAS version 9.4 (SAS Institute Inc).

As the experiment was web-based and participants came from a panel that provides incentives for participation, there was a risk that some participants would only participate to collect their incentives and not provide proper answers. As a measure of quality assurance, we excluded data from participants who answered all the questions in less than 2 minutes, spent less than 20 seconds on the page displaying the research summary, and spent less than 1.5 minutes between reading the research summary and completing the questionnaire (so-called speeders). These time limits were determined by test readings. We also excluded participants who provided the answers in the same row for the matrix questions, that is, when more than 1 item was displayed on the screen (so-called straightliners). This can be considered a conservative approach to exclude data by using a high threshold for implausibility. The exclusion of data from speeders and straightliners was planned a priori. We conducted sensitivity analyses by including these data to test the robustness of the main results.

Ethical Review

A formal ethical approval for this study was waived by the institutional review board of the University of Erfurt owing to the negligible risk to the participants and because the study did not involve collection of identifiable data (EV-20180921). All participants gave their written informed consent to use and share their data for scientific purposes. Only anonymized nonidentifiable data that do not enable identification of the individual participants were collected and analyzed (Multimedia Appendix 3 shows the CONSORT-EHEALTH checklist).

Results

Sociodemographic Data

A total of 2099 invited panel members were assessed for eligibility: 354 participants did not qualify owing to full quotas, 18 declined to participate, and 94 were excluded after the randomization because the data were deemed invalid (93 straightliners or speeders and 1 empty questionnaire). The sociodemographic characteristics of these participants excluded from the analysis owing to implausible data are presented in Multimedia Appendix 4. There is no visible pattern of postrandomization exclusions across groups (range 3%-7%). The final sample comprised 1633 participants. Data on the primary and secondary outcomes were available for 95%-97% of these participants, depending on the outcome (Figure 1). There were no notable differences between the number of

participants with incomplete data among the groups (range 3%-8%). The number of participants randomized to each group and their sociodemographic characteristics are presented in Table 2.

Figure 1. Study flow diagram.

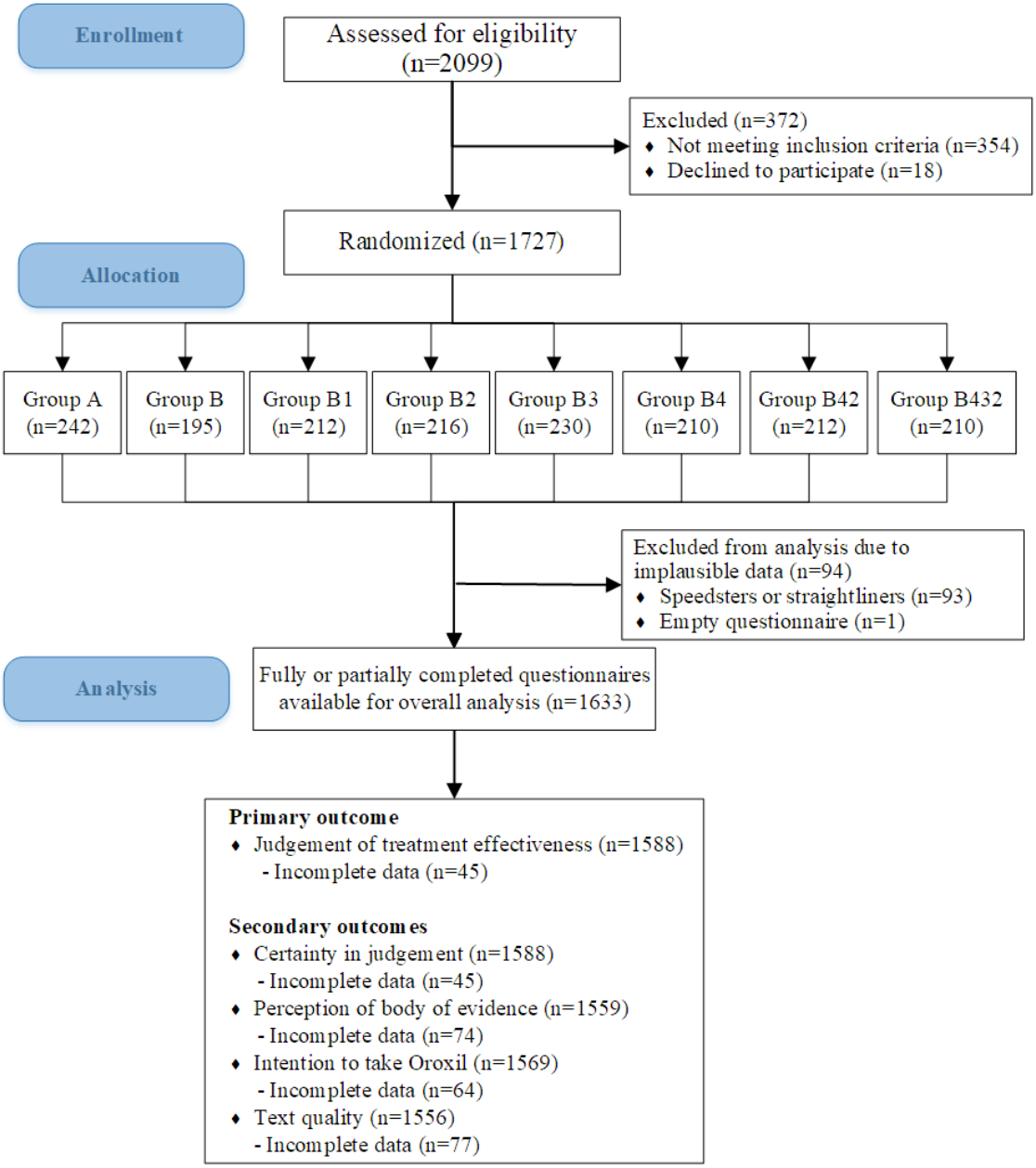


Table 2. Sociodemographic characteristics of the participants.

Data	Group A[a], n=242	Group B[b], n=195	Group B1[c], n=212	Group B2[d], n=216	Group B3[e], n=230	Group B4[f], n=210	Group B42[g], n=212	Group B432[h], n=210
Exclusions and missing data for the primary outcome, n (%)								
Excluded due to invalid data[i]	17 (7.0)	8 (4.1)	12 (5.7)	14 (6.5)	8 (3.5)	14 (6.7)	9 (4.2)	12 (5.7)
Missing data[j]	3 (1.2)	6 (3.1)	7 (3.3)	12 (5.6)	3 (1.3)	5 (2.4)	3 (1.4)	6 (2.9)
Analyzed, n (%)[j]								
	222 (91.7)	181 (92.8)	193 (91.0)	190 (88.0)	219 (95.2)	191 (91.0)	197 (92.9)	195 (92.9)
Demographic characteristics								
Mean age (years), (SD)	46 (13.7)	45 (13.5)	46 (13.3)	47 (13.5)	46 (13.7)	46 (13.9)	45 (13.9)	47 (14.2)
Men, n (%)	105 (46.7)	91 (48.7)	90 (45.0)	103 (51.0)	106 (47.7)	101 (51.5)	96 (48.0)	117 (58.2)
Women, n (%)	120 (53.3)	96 (51.3)	110 (55.0)	99 (49.0)	116 (52.3)	95 (48.5)	104 (52.0)	84 (41.8)
Educational degree, n (%)								
None	0 (0)	1 (0.5)	1 (0.5)	2 (1.0)	0 (0)	0 (0)	0 (0)	0 (0)
Basic secondary	30 (13.3)	23 (12.3)	30 (15.0)	23 (11.4)	27 (12.2)	23 (11.7)	23 (11.5)	29 (13.7)
Higher secondary	74 (32.9)	73 (39.0)	65 (32.5)	82 (40.6)	67 (30.2)	63 (32.1)	80 (40.0)	68 (32.2)
General entry qualification for university	54 (24.0)	44 (23.5)	48 (24.0)	44 (21.8)	68 (30.6)	48 (24.5)	46 (23.0)	68 (32.2)
University degree	67 (29.8)	46 (24.6)	56 (28.0)	51 (25.2)	60 (27.0)	62 (31.6)	51 (25.5)	46 (21.8)
History of tinnitus, n (%)[k]								
Currently symptomatic	31 (14.3)	30 (16.9)	25 (13.4)	22 (11.8)	22 (10.2)	23 (12.2)	27 (14.1)	30 (15.7)
Previously symptomatic	37 (17.1)	21 (11.9)	27 (14.5)	37 (19.9)	28 (13.0)	33 (17.5)	26 (13.5)	24 (12.6)
No history of tinnitus	149 (68.7)	126 (71.2)	134 (72.0)	127 (68.3)	166 (76.9)	133 (70.4)	139 (72.4)	137 (71.7)
Profession, n (%)[k]								
Medical	17 (7.8)	13 (7.3)	20 (10.8)	19 (10.2)	18 (8.3)	18 (9.5)	26 (13.5)	18 (9.4)
Nonmedical	200 (92.2)	164 (92.7)	166 (89.2)	167 (89.8)	198 (91.7)	171 (90.5)	166 (86.5)	173 (90.6)

[a]Effect shown.

[b]Indication of effect.

[c]Indication of effect with general explanation.

[d]Publication bias/vested interests.

[e]Indirectness (population).

[f]Imprecision (small sample size).

[g]Publication bias/vested interests and imprecision.

[h]Publication bias/vested interests and imprecision and indirectness.

[i]Speeders, straightliners, and empty questionnaires.

[j]Regarding primary outcome.

[k]Information not provided by 79 participants.

Perception of the Treatment Effectiveness (Primary Outcome)

There were no overall statistical differences between the groups for Q1 (*degree of uncertainty*, *P*=.25). Numerically, the data show a slightly larger proportion of participants considering the drug to be of proven benefit in group A (*effect shown*) compared to that in group B (*indication of effect*) and B1 (*indication of effect and need for further research*), while a slightly smaller proportion of respondents considered it to be of unlikely benefit in group A (Table 3).

There were also no overall statistical differences between the groups for Q2 (*type of uncertainty*, *P*=.73). Numerically, condition B3 (*indirectness*) shifted some participants from considering the treatment to be clearly beneficial to considering it to be possibly beneficial (Table 3).

For Q3 (*magnitude of uncertainty*), there was a significant global effect on the perception of treatment effectiveness (*P*=.048).

Pairwise comparisons showed a weaker perception of the effectiveness for the research summary with 3 sources of uncertainty compared to the version with 2 sources of uncertainty (*P*=.04). This is reflected in the answers: the proportion of the participants in the group with 3 sources of uncertainty that perceived the drug as possibly beneficial was 9% lower than that of the participants in the group with 2 sources of uncertainty (92/195, 47.2% vs 111/197, 56.3%, respectively), and the proportion of the participants in the group with 3 sources of uncertainty that considered the drug to be of unclear benefit was 8% higher than that of the participants in the group with 2 sources of uncertainty (72/195, 36.9% vs 57/197, 28.9%, respectively). Furthermore, with 3.6% (7/195) of the responses, the group with 3 sources of uncertainty was the only group with a considerable number of participants perceiving the drug as being unbeneficial (Table 3). However, there was no difference between the version with 1 and the version with 3 sources of uncertainty on pairwise comparison (*P*=.31). The results from the sensitivity analysis using ANOVA were consistent with those of the Kruskal-Wallis tests for all global comparisons. The results from the sensitivity analyses, including straightliners and speeders, were consistent with the main findings.

Table 3. Results of the primary outcome (perception of treatment effectiveness).

Questions (Q), Group identifier	Proportion of respondents' answers, n (%)					*P* value[a]
	Benefit proven	Possible benefit	Benefit unclear	Benefit unlikely	No Benefit	
Q1: Degree of uncertainty						.25
A[b], n=222	33 (14.9)	115 (51.8)	63 (28.4)	10 (4.5)	1 (0.5)	
B[c], n=181	20 (11.1)	92 (50.8)	53 (29.3)	14 (7.7)	2 (1.1)	
B1[d], n=193	19 (9.8)	103 (53.4)	55 (28.5)	14 (7.3)	2 (1.0)	
Q2: Type of uncertainty						.73
B1, n=193	19 (9.8)	103 (53.4)	55 (28.5)	14 (7.3)	2 (1.0)	
B2[e], n=190	23 (12.1)	96 (50.5)	59 (31.1)	12 (6.3)	0 (0.0)	
B3[f], n=219	13 (5.9)	124 (56.6)	64 (29.2)	16 (7.3)	2 (0.9)	
B4[g], n=191	16 (8.4)	99 (51.8)	60 (31.4)	15 (7.9)	1 (0.5)	
Q3: Number of sources of uncertainty						.048
B4, n=191	16 (8.4)	99 (51.8)	60 (31.4)	15 (7.9)	1 (0.5)	
B42[h], n=197	17 (8.6)	111 (56.3)	57 (28.9)	12 (6.1)	0 (0.0)	
B432[i], n=195	12 (6.2)	92 (47.2)	72 (36.9)	12 (6.2)	7 (3.6)	

[a]Kruskal-Wallis test for global effect.

[b]Effect shown.

[c]Indication of effect.

[d]Indication of effect with general explanation.

[e]Publication bias/vested interests.

[f]Indirectness (population).

[g]Imprecision (small sample size).

[h]Publication bias/vested interests and imprecision.

[i]Publication bias/vested interests and imprecision and indirectness.

Secondary Outcomes

The variations of the research summary had little impact on the secondary outcomes. There was a significant overall effect of the *type of uncertainty* on the perception of the body of evidence (*P*=.01). In pairwise comparison, the description for imprecision (B4) had a slightly larger effect on the *perceived limitations in the body of evidence* than the general statement that more research is needed (B1) (*P*=.01). Furthermore, there was a significant global effect on the *text quality* (*P*=.03). However, pairwise comparisons did not show statistical significance (*P*=.06 for the difference between B2 [*publication bias/vested interests*] and B4 [*imprecision*]). The results of the sensitivity analysis using ANOVA were consistent with those of the Kruskal-Wallis tests for the secondary outcomes. The results of the sensitivity analyses including straightliners and speeders were consistent with the main findings. The detailed results of the secondary outcomes are presented in Table 4.

Table 4. Results of the Kruskal-Wallis tests for global effects of the secondary outcomes.

Secondary outcome, mean (SD)	Q1: Degree of uncertainty				Q2: Type of uncertainty					Q3: Number of sources of uncertainty			
	A[a]	B[b]	B1[c]	*P* value	B1	B2[d]	B3[e]	B4[f]	*P* value	B4	B42[g]	B432[h]	*P* value
Certainty in judgement	3.60 (0.96)	3.49 (1.01)	3.49 (0.92)	.34	3.49 (0.92)	3.46 (0.92)	3.44 (0.93)	3.61 (0.98)	.22	3.61 (0.98)	3.47 (0.96)	3.49 (0.90)	.23
Perception of body of evidence	3.13 (0.76)	3.16 (0.81)	3.14 (0.77)	.80	3.14 (0.77)	3.00 (0.76)	3.00 (0.79)	2.89 (0.79)	.01	2.89 (0.79)	3.02 (0.75)	2.85 (0.68)	.09
Intention to take Oroxil	3.32 (1.18)	3.26 (1.18)	3.30 (1.23)	.91	3.30 (1.23)	3.26 (1.04)	3.26 (1.14)	3.24 (1.17)	.93	3.24 (1.17)	3.34 (1.05)	3.20 (1.10)	.51
Text quality	3.51 (0.75)	3.59 (0.80)	3.60 (0.80)	.33	3.60 (0.80)	3.44 (0.75)	3.58 (0.76)	3.60 (0.78)	.03	3.60 (0.78)	3.63 (0.70)	3.52 (0.68)	.23

[a]Effect shown.

[b]Indication of effect.

[c]Indication of effect with general explanation.

[d]Publication bias/vested interests.

[e]Indirectness (population).

[f]Imprecision (small sample size).

[g]Publication bias/vested interests and imprecision.

[h]Publication bias/vested interests and imprecision and indirectness.

Discussion

Summary of Results

For the primary outcome (*perception of effectiveness)*, we did not find any statistically significant differences between alternative wordings for different *degrees of uncertainty* (Q1) and between different *sources of uncertainty* (Q2). However, there was a significant effect of the *number of sources of uncertainty* (Q3). As the differences between the groups in Q3 were small and the *P* value was just below the significance level, we believe that this result should still be interpreted cautiously. Furthermore, we did not see a dose-response effect of the number of sources of uncertainty on the perception of effectiveness. A possible explanation is that indirectness, which was added in the version of the research summary with 3 sources of uncertainty, is of particular importance to people. This seems plausible, as it is a tangible source of uncertainty, easy to understand, and relates to them personally. However, this type of uncertainty did not decrease the perception of treatment effectiveness in Q2 to a statistically significant degree.

The different presentations of uncertainty had no meaningful effect on the secondary outcomes, including the intention to take Oroxil. While the *body of evidence* was perceived as slightly more limited in the research summary describing imprecision, the difference between the groups was small and may be a result of the wordings used to measure the outcome (certain, reliable, etc).

Possible Explanations

Psychological research suggests that discounting cues (eg, referring to a study with reduced credibility) prior to the provision of information can reduce the impact of that information [12]. Thus, our results may not apply to information that presents uncertainty earlier. If positioning had an effect, 1 possible solution in written health information would be to open the evidence section with a statement on the quality of the evidence and to make this as clear as possible, instead of presenting it at the end.

The results may also be explained by the neutral choice of language in the research summaries, thereby resulting in subtle differences between some of the variations. While stronger wordings may be needed to convey the uncertainties, we do not generally recommend the use of a more partial style of language as this would counteract the aims of providing balanced information and supporting informed decision making. This said, there are exceptions to this rule, for example, in case of highly implausible treatments such as homeopathy or new high-risk treatments lacking good evidence. On a more critical note, use of words may be unsuitable to convey the quality of evidence in general. While we identified little experimental research addressing similar questions, 1 study with physicians did not show an effect of different wordings such as “might” or “suggest” on conveying the strength of the recommendation within guidelines [13].

Lastly, it is possible that in a competitive choice situation, where there are 2 treatments, uncertainty information may matter more (ceteris paribus) because it helps to discriminate drugs from each other. This may be particularly true for the outcome intended treatment decision, where, in our study, participants in all groups tended to opt for taking the drug. A possible explanation for the lack of differences between the groups is that the scenario suggested that there are no other proven treatments for tinnitus.

On an average, 52.4% (832/1588) of all the participants considered the treatment to be possibly helpful, and about 30.4% (483/1588) considered the benefit of the drug to be unclear, including the group without any expressed uncertainty (A). A possible explanation for this result is that the use of numbers to communicate the response rate for the treatment may have diluted the effects of the uncertainty descriptions. This raises the question of whether communicating certainty in the quality

of evidence in parallel with probability information is the actual challenge for communicators rather than communicating uncertainty. The explanation seems likely since the public is still not used to such presentations and probably has difficulties in distinguishing between uncertainty from limitations in the quality of evidence (the certainty in an effect estimate) and uncertainty in terms of dealing with probabilistic outcomes (the magnitude of an effect). This has been shown in a qualitative study conducted to improve Cochrane plain language summaries [14]. Users ignored or reduced the 4 grading levels (very low to high) that were used in the plain language summaries, irrespective of the choice of words used to delineate uncertainty. An important difference between that study and ours is that we looked at uncertainties from specific sources, whereas Glenton and colleagues evaluated modified GRADE summary of findings tables without making the sources of uncertainty explicit to the users. In a follow-up randomized trial, one of the enhanced plain-language summaries appeared to improve understanding of the quality of evidence [15]. However, the format is limited to reviews using the GRADE system or requires health information developers to apply the GRADE framework. Furthermore, the way understanding was measured in this study was problematic. Specifically, the use of catchphrases in the enhanced plain language summary may have led more participants to answer correctly to the questions rather than improved understanding over the conventional format. A focus group study of a decision aid of management options for women with breast cancer and *BRCA*-positive gene mutations looked at ribbons to illustrate levels of evidence (4 grades from bronze to platinum). In this study, the users assumed that if data are presented using numbers and graphs, people will automatically consider them to be of sufficient quality [16]. These explanations aside, people may simply tend to focus on effectiveness when reading health information and often overestimate the benefits of treatments [17].

Possible Solutions

A possible but, in many cases, an inappropriate solution would be to avoid the use of numbers. Presenting quantitative information on treatment effects has become a standard requirement for evidence-based health information and decision aids because probabilities on the benefits and harms support realistic expectations and allow patients to weigh the pros and cons [18]. Furthermore, words have been shown to be unsuitable to express probabilities in several studies [19,20]. Instead, it might be necessary to provide readers with a clear statement on what a specific type of uncertainty means in order to aid interpretation. In a randomized trial on drug fact boxes that studied uncertainty due to the use of surrogate outcomes and unknown safety profiles for newly approved drugs, both directive and nondirective explanations on what these uncertainties mean changed the intention to take the drug [21]. However, the effects were relatively small, with an absolute change of 12% and 19%. While this appears to be a promising approach, the implications of the uncertainties arising from different populations, publication bias, and small sample sizes as tested in our study can be ambiguous and require a high level of judgement. Thus, providing such explanations may be difficult in written health information aimed at a broad range of readers in some cases.

Lastly, providers of evidence-based health information should carefully consider whether the quality of the evidence is sufficient for an effect to be presented at all. While fine-graded levels of evidence such as in the GRADE system are useful to health care professionals, they may be too nuanced for patients and the public. Instead, it may be more sensible to use a higher cut-off for presenting treatment effects (eg, at least moderate quality of evidence within GRADE) while labelling low or very low quality of evidence to be unclear. This seems reasonable as the true effect may substantially differ from the study estimate for evidence of low or very low quality according to the current definition within the GRADE system [22].

While there are many possible solutions to dealing with uncertainty, it is unlikely that one will fit all scenarios in practice. How uncertainty is presented may depend on many factors such as the format and aims of the health information, the target group, the severity of the condition, the risks of the intervention, and the availability and number of treatment options. Thus, and maybe most importantly, the judgements made in this process should be carefully reflected.

Strengths and Limitations

The main strengths of our study are the large sample population based on a sample size calculation for a conservative effect, few missing data, and a robust protocol-driven methodology. Furthermore, we used a research summary that is concordant with the current standards of evidence-based health information, including the use of probabilities to communicate benefits and harms in order to allow weighing the pros and cons. The weaknesses include the use of a hypothetical scenario and restriction to 1 (relatively benign) condition and treatment. Furthermore, the choice of the size of the presented treatment effects is debatable. We decided to use a small absolute effect since large treatment effects are rare in practice [23]. In our experience, they often range from around 0.1% to 1% risk difference for primary preventive or screening interventions (small) to about 5% (medium) to 20% (large) for symptom improvement. Thus, we believe that our information provides a good representation of what developers of evidence-based health information are confronted with in practice. A possible limitation to using simple frequencies is that readers may mistake them for sample size. We cannot exclude that this had an impact on their interpretation of the research summaries. Another limitation is that this study was conducted in Germany and these results may not be generalizable to other countries. Finally, the way the primary outcome was measured may not have been sensitive enough to capture differences and there is always a risk that participants tend to avoid more extreme answers. However, the lack of meaningful effects on secondary outcomes supports the notion that there are indeed few perceived differences between the research summaries. Thus, it seems unlikely that measuring perceived treatment effectiveness using a visual analog or Likert scale, for example, would have made a difference to the results. The downside of such measurements is that the relevance of the results is difficult to interpret. Owing

to these limitations, we cannot completely rule out that participants are insensitive to uncertainty information.

Recommendations for Future Research

In order for uncertainty descriptions to be useful in conveying limitations in the quality of evidence, the public has to be aware that not all research is of equal quality and that the size of an effect and the confidence in the effect are separate concepts. Therefore, it would be helpful to develop or enhance and implement interventions that explain basic research concepts and skills. A notable recent effort in this direction has been the Key Concepts for Informed Health Choices Project [24-26]. It may also be helpful to develop more radical ways of presenting uncertainty together with consumers. Future studies could also assess the effect of uncertainty information in situations wherein patients decide between several treatments. In terms of the methods, future studies could use factorial designs to study different combinations of uncertainty descriptions.

Conclusion

In conclusion, providers of written evidence-based health information or decision aids should not assume that consumers and patients will understand the intentions and meaning of uncertainty descriptions without further explanation or support. The most likely explanation for this is that the public does not distinguish between the concepts of quality of evidence and magnitudes of effect. The challenge in communicating uncertainty should not be used as an argument to withhold uncertainty information from patients and the public. Future research should aim to understand how the public or patients process information regarding limitations of evidences in order to understand why the degree and types of uncertainty seem to have little impact on the perceptions of effectiveness or treatment intentions.

Acknowledgments

The study was funded by the Institute for Quality and Efficiency in Health Care (IQWiG) within the merits of it general commission to select topics for scientific evaluation independently. IQWiG paid an honorarium to CB and CR to provide academic expertise and for questionnaire development. All other authors are employees of IQWiG.

Authors' Contributions

RBB, ME, DF, SK, AW, and RM had the initial idea for the study, conceived the study design, developed the research summaries, and drafted a preliminary version of the questionnaire. CB and CR elaborated, extended, and pretested the questionnaire and commented on the study design. CB and UG performed sample size calculations. UG and RBB developed the statistical analysis plan. UG analyzed the data. RBB drafted the first version of this manuscript. All authors critically reviewed and approved the final version.

Conflicts of Interest

RBB, ME, DF, UG, SK, RM, and AW are employees of IQWiG.

Multimedia Appendix 1
Exemplary research summary.
[DOCX File , 14 KB - jmir_v22i8e15899_app1.docx]

Multimedia Appendix 2
German versions of the research summaries.
[DOCX File , 17 KB - jmir_v22i8e15899_app2.docx]

Multimedia Appendix 3
CONSORT-EHEALTH checklist.
[PDF File (Adobe PDF File), 280 KB - jmir_v22i8e15899_app3.pdf]

Multimedia Appendix 4
Sociodemographic characteristics of the excluded participants.
[DOC File , 48 KB - jmir_v22i8e15899_app4.doc]

References

1. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington (DC): National Academies Press (US); 2001.
2. Gigerenzer G, Gaissmaier W. Heuristic decision making. Annu Rev Psychol 2011;62:451-482. [doi: 10.1146/annurev-psych-120709-145346] [Medline: 21126183]

3. Longman T, Turner RM, King M, McCaffery KJ. The effects of communicating uncertainty in quantitative health risk estimates. Patient Educ Couns 2012 Nov;89(2):252-259. [doi: 10.1016/j.pec.2012.07.010] [Medline: 22858415]
4. Sladakovic J, Jansen J, Hersch J, Turner R, McCaffery K. The differential effects of presenting uncertainty around benefits and harms on treatment decision making. Patient Educ Couns 2016 Jun;99(6):974-980. [doi: 10.1016/j.pec.2016.01.009] [Medline: 26823204]
5. McCormack L, Sheridan S, Lewis M, Boudewyns V, Melvin CL, Kistler C, et al. Communication and dissemination strategies to facilitate the use of health-related evidence. Evid Rep Technol Assess (Full Rep) 2013 Nov(213):1-520. [doi: 10.23970/ahrqepcerta213] [Medline: 24423078]
6. Büchter RB, Betsch C, Ehrlich M, Fechtelpeter D, Grouven U, Keller S, et al. Communicating Uncertainty From Limitations in Quality of Evidence to the Public in Written Health Information: Protocol for a Web-Based Randomized Controlled Trial. JMIR Res Protoc 2019 May 13;8(5):e13425 [FREE Full text] [doi: 10.2196/13425] [Medline: 31094343]
7. What we do at InformedHealth.org. Informed Health Online. URL: https://www.informedhealth.org/our-approach.2059.en.html [accessed 2019-12-12]
8. Institute for Quality and Efficiency in Health Care (IQWiG). General Methods: Version 6.0. Cologne: IQWiG; 2017. URL: https://www.iqwig.de/de/methoden/methodenpapier.3020.html [accessed 2019-12-12]
9. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
10. Brosius H, Birk M. Text-Bild-Korrespondenz und Informationsvermittlung durch Fernsehnachrichten. Rundfunk und Fernsehen. 1994. URL: https://www.ls1.ifkw.uni-muenchen.de/forschung/publikationen/fz_artikel/text-bild-korrespond.html [accessed 2019-10-12]
11. Gehrau V. Fernsehbewertung Und Fernsehhandlung. München: Edition Reinhard Fischer Im Nomos Verlag; 2019.
12. Kumkale GT, Albarracín D. The sleeper effect in persuasion: a meta-analytic review. Psychol Bull 2004 Jan;130(1):143-172 [FREE Full text] [doi: 10.1037/0033-2909.130.1.143] [Medline: 14717653]
13. Akl EA, Guyatt GH, Irani J, Feldstein D, Wasi P, Shaw E, et al. "Might" or "suggest"? No wording approach was clearly superior in conveying the strength of recommendation. J Clin Epidemiol 2012 Mar;65(3):268-275. [doi: 10.1016/j.jclinepi.2011.08.001] [Medline: 22075112]
14. Glenton C, Santesso N, Rosenbaum S, Nilsen ES, Rader T, Ciapponi A, et al. Presenting the results of Cochrane Systematic Reviews to a consumer audience: a qualitative study. Med Decis Making 2010;30(5):566-577. [doi: 10.1177/0272989X10375853] [Medline: 20643912]
15. Santesso N, Rader T, Nilsen ES, Glenton C, Rosenbaum S, Ciapponi A, et al. A summary to communicate evidence from systematic reviews to the public improved understanding and accessibility of information: a randomized controlled trial. J Clin Epidemiol 2015 Mar;68(2):182-190 [FREE Full text] [doi: 10.1016/j.jclinepi.2014.04.009] [Medline: 25034199]
16. Culver JO, MacDonald DJ, Thornton AA, Sand SR, Grant M, Bowen DJ, et al. Development and evaluation of a decision aid for BRCA carriers with breast cancer. J Genet Couns 2011 Jun;20(3):294-307 [FREE Full text] [doi: 10.1007/s10897-011-9350-4] [Medline: 21369831]
17. Hoffmann TC, Del Mar C. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. JAMA Intern Med 2015 Mar;175(2):274-286. [doi: 10.1001/jamainternmed.2014.6016] [Medline: 25531451]
18. Trevena LJ, Zikmund-Fisher BJ, Edwards A, Gaissmaier W, Galesic M, Han PKJ, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. BMC Med Inform Decis Mak 2013;13 Suppl 2:S7 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S7] [Medline: 24625237]
19. Blalock SJ, Sage A, Bitonti M, Patel P, Dickinson R, Knapp P. Communicating information concerning potential medication harms and benefits: What gist do numbers convey? Patient Educ Couns 2016 Dec;99(12):1964-1970. [doi: 10.1016/j.pec.2016.07.022] [Medline: 27444232]
20. Büchter RB, Fechtelpeter D, Knelangen M, Ehrlich M, Waltering A. Words or numbers? Communicating risk of adverse effects in written consumer health information: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2014 Aug 26;14:76 [FREE Full text] [doi: 10.1186/1472-6947-14-76] [Medline: 25155972]
21. Schwartz LM, Woloshin S. Communicating uncertainties about prescription drugs to the public: a national randomized trial. Arch Intern Med 2011 Sep 12;171(16):1463-1468. [doi: 10.1001/archinternmed.2011.396] [Medline: 21911629]
22. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015] [Medline: 21208779]
23. Kent DM, Trikalinos TA. Therapeutic innovations, diminishing returns, and control rate preservation. JAMA 2009 Nov 25;302(20):2254-2256 [FREE Full text] [doi: 10.1001/jama.2009.1679] [Medline: 19934428]
24. Castle JC, Chalmers I, Atkinson P, Badenoch D, Oxman AD, Austvoll-Dahlgren A, et al. Establishing a library of resources to help people understand key concepts in assessing treatment claims-The "Critical thinking and Appraisal Resource Library" (CARL). PLoS One 2017;12(7):e0178666 [FREE Full text] [doi: 10.1371/journal.pone.0178666] [Medline: 28738058]
25. Chalmers I, Oxman AD, Austvoll-Dahlgren A, Ryan-Vig S, Pannell S, Sewankambo N, et al. Key Concepts for Informed Health Choices: a framework for helping people learn how to assess treatment claims and make informed choices. BMJ Evid Based Med 2018 Mar;23(1):29-33. [doi: 10.1136/ebmed-2017-110829] [Medline: 29367324]

26. Cusack L, Del Mar CB, Chalmers I, Gibson E, Hoffmann TC. Educational interventions to improve people's understanding of key concepts in assessing the effects of health interventions: a systematic review. Syst Rev 2018 May 02;7(1):68 [FREE Full text] [doi: 10.1186/s13643-018-0719-4] [Medline: 29716639]

Abbreviations

ANOVA: analysis of variance
GRADE: Grading of Recommendations Assessment, Development and Evaluation
IQWiG: German Institute for Quality and Efficiency in Health Care

Edited by G Eysenbach; submitted 08.11.19; peer-reviewed by N Cui, B Zikmund-Fisher; comments to author 21.02.20; revised version received 14.04.20; accepted 21.06.20; published 10.08.20.

Please cite as:
Büchter RB, Betsch C, Ehrlich M, Fechtelpeter D, Grouven U, Keller S, Meuer R, Rossmann C, Waltering A
Communicating Uncertainty in Written Consumer Health Information to the Public: Parallel-Group, Web-Based Randomized Controlled Trial
J Med Internet Res 2020;22(8):e15899
URL: http://www.jmir.org/2020/8/e15899/
doi:10.2196/15899
PMID:32773375

Original Paper

Googling Musculoskeletal-Related Pain and Ranking of Medical Associations' Patient Information Pages: Google Ads Keyword Planner Analysis

Yoichiro Yamaguchi[1], MD; Deokcheol Lee[1], MD, PhD; Takuya Nagai[1], MD; Taro Funamoto[1], MD, PhD; Takuya Tajima[1], MD, PhD; Etsuo Chosa[1], MD, PhD

Department of Orthopaedic Surgery, Faculty of Medicine, University of Miyazaki, Miyazaki, Japan

Corresponding Author:
Yoichiro Yamaguchi, MD
Department of Orthopaedic Surgery
Faculty of Medicine
University of Miyazaki
5200 Kihara
Kiyotake
Miyazaki, 889-1692
Japan
Phone: 81 0985 85 0986
Email: yoichiroh_yamaguchi@med.miyazaki-u.ac.jp

Abstract

Background: Most people currently use the internet to obtain information about many subjects, including health information. Thus, medical associations need to provide accurate medical information websites. Although medical associations have their own patient education pages, it is not clear if these websites actually show up in search results.

Objective: The aim of this study was to evaluate how well medical associations function as online information providers by searching for information about musculoskeletal-related pain online and determining the ranking of the websites of medical associations.

Methods: We conducted a Google search for frequently searched keywords. Keywords were extracted using Google Ads Keyword Planner associated with "pain" relevant to the musculoskeletal system from June 2016 to December 2019. The top 20 search queries were extracted and searched using the Google search engine in Japan and the United States.

Results: The number of suggested queries for "pain" provided by Google Ads Keyword Planner was 930 in the United States and 2400 in Japan. Among the top 20 musculoskeletal-related pain queries chosen, the probability that the medical associations' websites would appear in the top 10 results was 30% in the United States and 45% in Japan. In five queries each, the associations' websites did not appear among the top 100 results. No significant difference was found in the rank of the associations' website search results (P=.28).

Conclusions: To provide accurate medical information to patients, it is essential to undertake effective measures for search engine optimization. For orthopedic associations, it is necessary that their websites should appear among the top search results.

(J Med Internet Res 2020;22(8):e18684) doi:10.2196/18684

KEYWORDS

Google; ad words; infodemiology; musculoskeletal-related pain; patient education; medical information

Introduction

Justification

Searching for information on the internet is now common practice. Medical information is no exception, and the term "consulting Dr. Google" has been popularized to represent searching for medical information on the internet [1]. Various studies have investigated the reliability and accuracy of medical information acquired on the internet [2-4]. However, there has been no research to specifically examine the extent to which medical information from websites supervised by medical associations is acquired from search results.

Background

As of 2018, more than 100 million people in Japan (79.8% of the total population) used the internet with various devices, including half the population aged 65-79 years [5]. Japan is an increasingly aging society, and many older people suffer from locomotive disorders [6]; thus, it can be assumed that these individuals use the internet to obtain information about their conditions. According to one survey, 21.2% of patients referred to medical information from a medical institution on the internet, and 12.1% of patients referred to the social networking services of a medical institution or the government before going to the hospital [7].

Among all currently available search engines, Google has an overwhelming market share in Japan, accounting for 91.2% of searches on computers and 99.1% of searches on mobile devices [8]. Many people with musculoskeletal pain such as lower back pain and sciatica use Google to obtain information about their conditions. In such situations, it is valuable to know the type of information that people actually obtain from internet searches. Many hospitals, companies, public interest groups, bloggers, and celebrities have posted medical information on websites. Accordingly, a substantial amount of inaccurate information about medical health has been posted globally [9]. We work in the orthopedic surgery department of a hospital and encounter many patients who are misled by inaccurate medical information about musculoskeletal-related pain. The Japan Medical Association states that it will support the health and cultural well-being of the Japanese population. Providing accurate medical information to patients is one of the missions of such medical associations, and orthopedic associations play a key role in the provision of online information about musculoskeletal-related pain.

Orthopedic associations have their own websites with patient education resources that provide information on symptoms and diseases. However, it is not known whether these websites provided by the orthopedic associations appear in the results of searches with search engines. Google Trends and Google Ads are often used to investigate what searches are actually performed on the internet. A relationship between Google's search volume increase and the influenza epidemic was noted based on such analyses [10-12]. In this study, we used the same tools to verify actual search queries and their relationship to orthopedic associations' website ranks. We also investigated seasonally varying varieties of musculoskeletal-related pain queries that were frequently searched. In addition, we investigated whether there is a more appropriate season for providing effective medical information for each symptom related to musculoskeletal-related pain. Based on this information, we investigated whether the associations' websites were able to provide suitable medical information to patients.

Methods

Search Strategy for Data Extraction

We extracted research volumes for the term "pain" (Japanese: "itami") and generated keywords using Google Ads Keywords Planner. The data were generated separately for each region: the United States (in English) and Japan (in Japanese). The data were collected in June 2020 and Google was used for all searches. The planner generated a list of proposed terms associated with "pain" and expressed the search volume for every month. The monthly search volumes for all keywords were extracted from June 2016 to December 2019. Total search volumes for the 4 years were calculated by population ratio and sorted in descending order. Among the generated keywords, two orthopedic doctors chose musculoskeletal pain–related keywords independently. The final list was refined by the senior author (EC). From this list, the top 20 search keywords were selected. Google searches were performed with the top 20 keywords and total hit websites were recorded. We employed Google Chrome without logging in to a Google account; we blocked location information and did not use the autocomplete function. We conducted the search in incognito mode. When searching, a proxy server was used. The search was performed through servers in Japan and the United States for each language. The Japanese Orthopedic Association (JOA) patient information pages [13] and the American Academy of Orthopedic Surgeons (AAOS) patient education pages [14] were selected as the association websites. The results of the searches were checked for the top 100 sites and those not displayed within the top 100 were ranked as over 100.

Statistical Analysis

All pain-related suggested searches in a 4-year period in both Japan and the United States were collected, and the number of searches per 100 people in each country was calculated. The monthly average number of musculoskeletal-related pain searches from June 2016 to December 2019 was also calculated. Among the keywords ranked within the top 20, those common to the United States and Japan were compared with seasonal variations. The Kruskal-Wallis test, Mann-Whitney *U* test, and Spearman rank correlation coefficient were used for statistical analysis of the search results. *P* values of <.05 were considered to indicate a significant difference. Statistical analyses were performed using Statistical Package for the Social Sciences (IBM Corp Released 2012, IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY, USA).

Results

The total number of generated keywords from "pain" was 930 in the United States and was 2400 in Japan. In Japan, the search term was "itami," a word that contains a Chinese character and Hiragana. In all generated queries, the average search volume ranged from 90 to 301,000 per month in the United States and from 140 to 135,000 per month in Japan. The search volume of the top 20 musculoskeletal pain–related queries ranged from 60,500 to 301,000 per month in the United States and from 22,200 to 135,000 per month in Japan. The population ratio was 0.85-4.99 per month per 100 people in the United States and was 0.88-4.94 per month per 100 people in Japan. No significant difference was found between the number of searches per population for the two countries (Mann-Whitney *U* test *P*=.67). Regarding the association websites that appeared within the top 100, the average rank was 16.47 (1st to 35th, mean 17th) in the United States and was 24.47 (2nd to 95th, mean 7th) in Japan. Within the top 20 queries, the association websites ranked within

the top 10 for 30% (6/20) of queries in the United States and for 45% (9/20) of queries in Japan. No significant difference was found between the number of ranked websites within the top 10 (Mann-Whitney *U* test *P*=.28). The total search volume of the top 20 queries was 0.75 per 100 people in the United States and was 0.73 per 100 people in Japan (Table 1 and Table 2). As shown in Figure 1, the total search volume increased year after year. In the top 20 queries, the most common search queries in the United States and Japan were "sciatica," "lower back pain," "back pain," "knee pain," and "leg cramps" (Figure 2). Seasonal variation in the number of searches was found for "leg cramps," which was significantly higher in the summer in both the United States and Japan. No significant difference was found between search volume and the number of total hit websites for the top 20 queries (Figure 3).

Table 1. Top 20 search queries in the United States.

Search queries	Average search volume per month	Four-year total search volume (per 100 people)	Total hit websites	Association website rank
sciatica	301,000	15,491,000 (4.99)	35,400,000	35
lower back pain	246,000	11,862,000 (3.82)	664,000,000	28
herniated disc	165,000	7,957,500 (2.56)	5,010,000	7
bursitis	165,000	7,687,000 (2.48)	5,610,000	8
knee pain	135,000	6,926,000 (2.23)	278,000,000	12
shoulder pain	135,000	5,941,000 (1.91)	302,000,000	7
carpal tunnel syndrome	110,000	5,594,000 (1.80)	13,400,000	1
muscle relaxers	110,000	5,043,500 (1.62)	1,370,000	>100
heel pain	110,000	4,869,000 (1.57)	143,000,000	6
back pain	90,500	4,624,000 (1.49)	1,510,000,000	27
neck pain	90,500	4,044,000 (1.30)	364,000,000	19
muscle spasm	90,500	4,038,000 (1.30)	11,400,000	25
hip pain	74,000	3,918,000 (1.26)	432,000,000	17
sciatic nerve pain	74,000	3,910,400 (1.26)	6,680,000	>100
piriformis syndrome	74,000	3,680,000 (1.19)	1,800,000	>100
heel spur	74,000	3,676,500 (1.18)	15,500,000	7
foot pain	60,500	3,237,000 (1.04)	1,440,000,000	>100
myalgia	60,500	3,062,500 (0.99)	3,240,000	>100
leg cramps	60,500	2,892,500 (0.93)	31,700,000	23
hip bursitis	60,500	2,639,600 (0.85)	5,910,000	25

Table 2. Top 20 search queries in Japan.

Search queries (English translation)	Average search volume per month	Four-year total search volume (per 100 people)	Total hit websites	Association website rank
zakotsushinkeitu (sciatica)	135,000	6,250,000 (4.94)	5,870,000	13
youtuu (lower back pain)	90,500	4,737,000 (3.74)	83,300,000	2
rokkanshinkeitsu (intercostal neuralgia)	74,000	3,724,500 (2.94)	577,000	>100
kensyouen (tendosynovitis)	74,000	3,307,000 (2.61)	8,660,000	3
komuragaeri (leg cramps)	49,500	2,217,300 (1.75)	995,000	>100
senaka no itami (back pain)	40,500	2,120,800 (1.68)	26,600,000	91
koshi ga itai (my lower back aches)	40,500	2,103,600 (1.66)	35,900,000	88
hiza no itami (knee pain)	40,500	2,090,900 (1.65)	25,000,000	7
ashi no ura itai (sole pain)	40,500	2,081,400 (1.64)	26,800,000	95
senaka ga itai (my back aches)	33,100	1,721,300 (1.36)	28,000,000	>100
hiza ga itai (my knee aches)	33,100	1,670,600 (1.32)	19,000,000	23
senaka itai (back, pain)	33,100	1,635,300 (1.29)	30,100,000	>100
kubi ga itai (my neck aches)	33,100	1,580,000 (1.25)	38,800,000	7
gojukata (painful shoulder at age 50s)	33,100	1,476,200 (1.17)	8,650,000	2
henkeiseihizakansetsusho (osteoarthritis of the knee)	33,100	1,413,600 (1.12)	4,860,000	2
kokansetsu itami (hip, pain)	27,100	1,310,700 (1.04)	7,360,000	5
shinkeitsu (neuralgia)	27,100	1,224,600 (0.97)	7,660,000	17
fukurahagi itai (calf, pain)	22,200	1,163,900 (0.92)	1,900,000	8
hiza itami (knee, pain)	22,200	1,139,400 (0.90)	26,600,000	4
shijukata (painful shoulder at age 40s)	22,200	1,115,300 (0.88)	253,000,000	>100

Figure 1. Total number of musculoskeletal-related searches in the United States and Japan (total monthly search volume of the top 20 search queries). Vertical axis: Searches per 100 people.

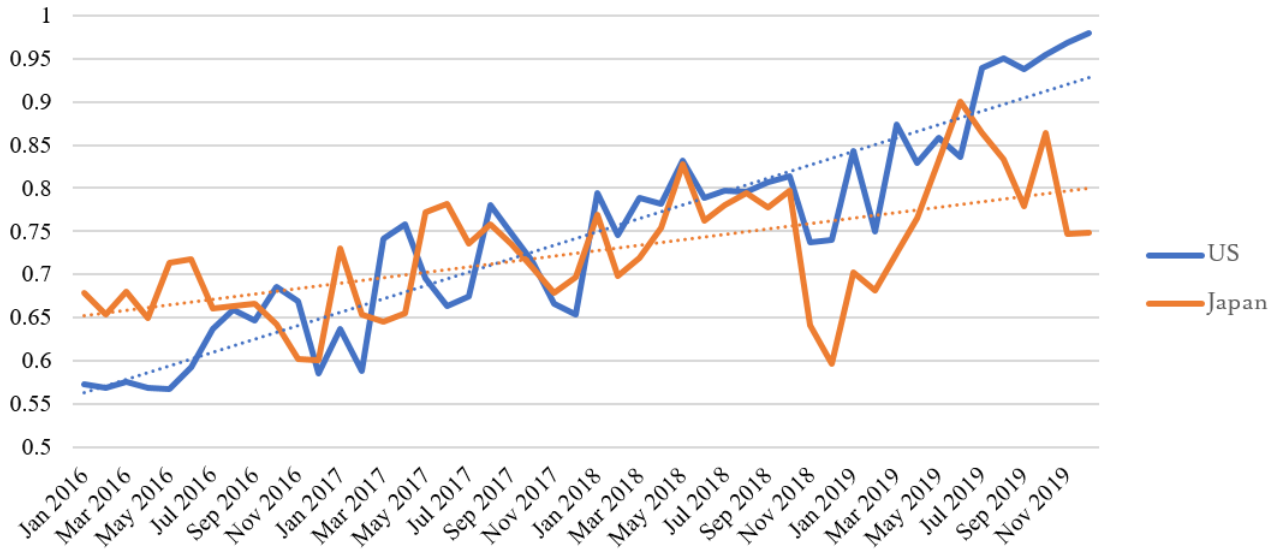


Figure 2. Comparison of musculoskeletal pain–related searches per month for each season (2016-2019) Vertical axis: Monthly searches per 10,000 people. [a]Differences between seasons (Kruskal–Wallis test); [b]Differences between the United States and Japan (Mann–Whitney U test).

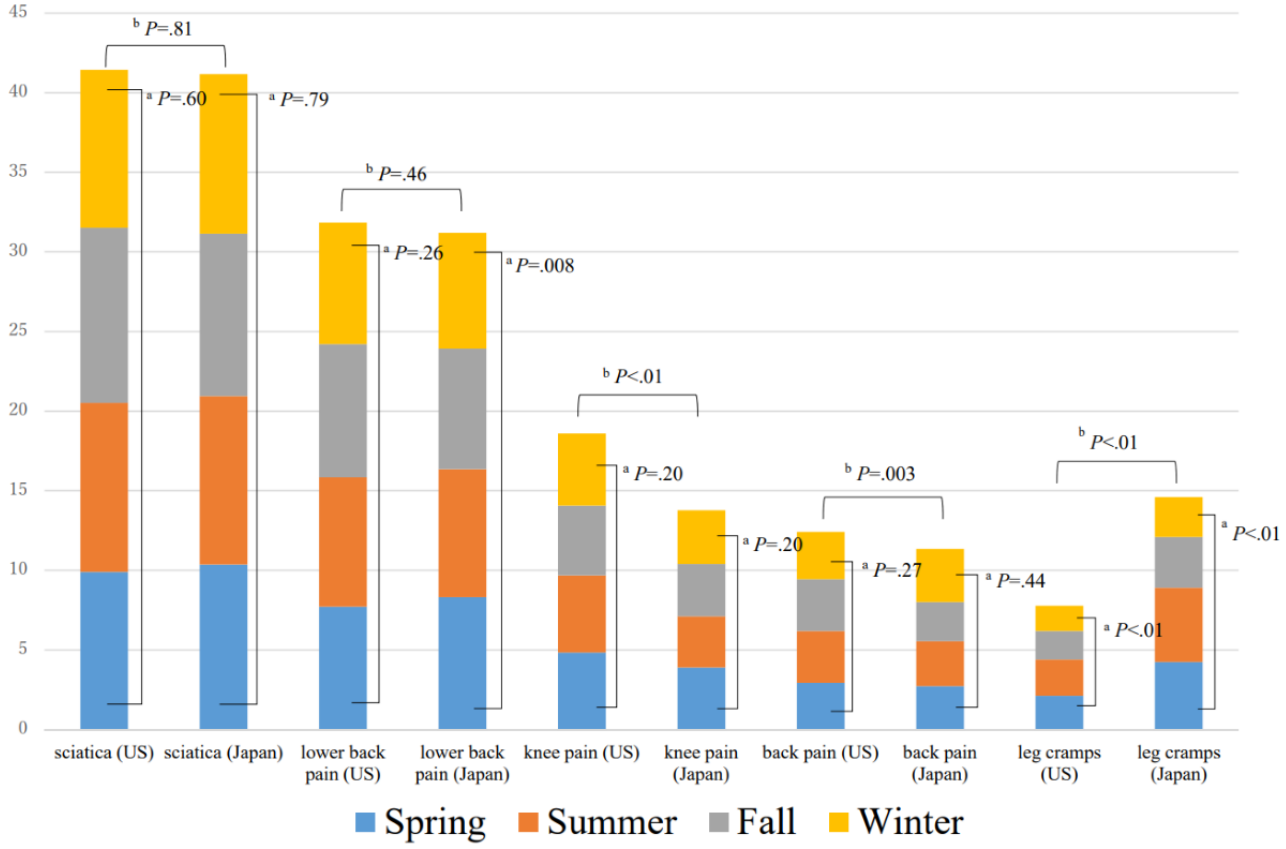
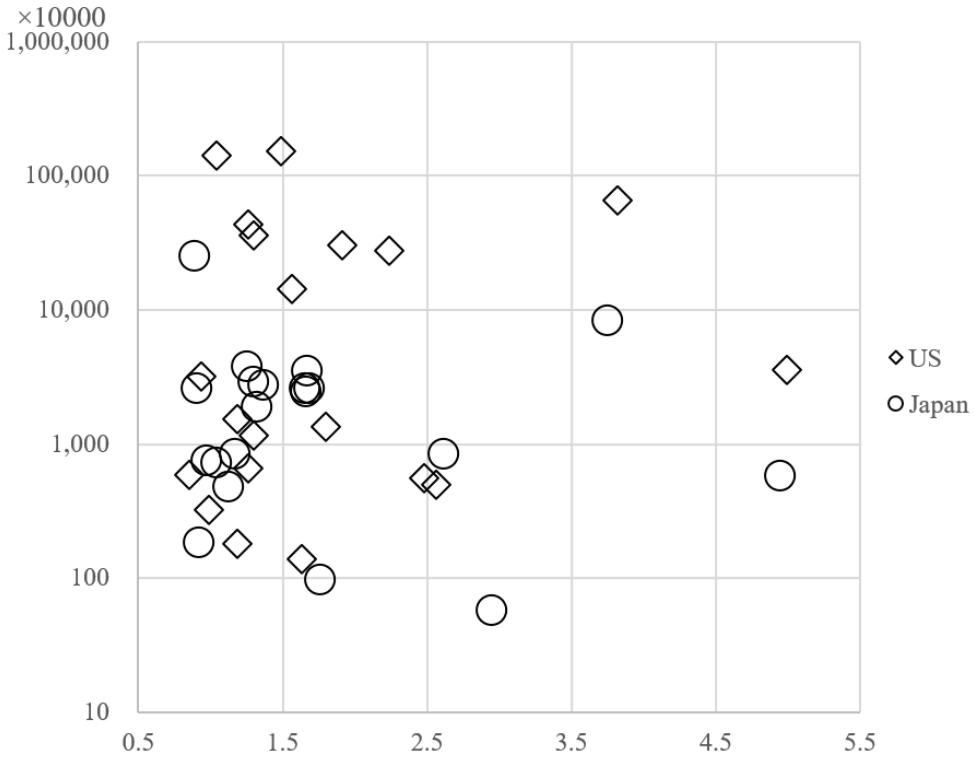


Figure 3. Search volumes and number of total hit websites. Vertical axis: Number of total hit websites (logarithmic scale). Horizontal axis: Four-year search volume per 100 people. No significant difference was observed in the relationship between the search volume and the number of total hit websites (Spearman rank correlation coefficient=0.13).



Discussion

This study is the first to examine how frequently the patient information pages of orthopedic associations appear when search terms related to musculoskeletal pain are entered. Based on the results of this study, we believe that there is a need for orthopedic associations to do more to provide medical information on the internet by optimizing their position in the search results for frequently searched symptom and disease-related terms.

The click-through rate for the first position in Google was 21.12%, followed by 10.65% and 7.57%. If the search result ranked 11th, the click-through rate became only 1.46% [15]. In a German study, when 289 patients searched for medical information on the internet, only 5 clicked a link on the second or subsequent search result pages [16].

In the early 2000s, over 60% of people in the United States and Australia reported that they had used the internet to search for medical information and were satisfied with the resulting health information and services [17-19]. More recently, 96% of people in Australia reported searching for medical information online; of those, 63% searched using smartphones [20]. Over time, it has become increasingly common for individuals to use online search engines when they need medical or health information.

According to many reports, inaccurate websites often appear among the top 10 results and provide erroneous information [21,22]. In the United Kingdom, Wikipedia sometimes appears more often in the search results than the National Health Service's website [23]. If members of the general public encounter websites containing incorrect information, it is difficult for them to assess its accuracy. In a US study, high school students with a science-focused education had difficulty distinguishing between reliable and unreliable medical sites, even though most of the information produced in the searches was inaccurate [24]. To solve health problems, government organizations need to work together to provide more accurate medical information on the web [9].

The algorithm used for Google searches is confidential; thus, many websites try to increase the quantity and quality of their traffic using organic search engine results. This practice is termed search engine optimization (SEO). At present, if items do not appear among search engine results, they are practically nonexistent.

In this study, we surveyed the search rankings of the patient information pages of two orthopedic associations and found that these pages are not likely to be viewed by patients seeking medical information with frequently searched queries. There was no significant difference in the number of searches per 100 people between the United States and Japan in the top 20 pages retrieved for musculoskeletal-related pain. Symptoms and illnesses are searched differently between the United States and Japan. The common query "leg cramps" was searched much more often in Japan per 100 people and was searched more often in the summer in both the United States and Japan (Figure 2). The predominance in summer suggests a link to climate. Indeed, in the United Kingdom and Australia, the volume of searches related to leg cramps increased in summer [25]. Another study found that searches related to ankle swelling also increased in midsummer [26]. Heartburn-related searches increased during winter, as did heart failure hospitalizations [10,27]. These studies indicate that an increase in the search volume for a particular condition reflects an actual increase in the number of patients affected with that condition. Thus, when the search volume increases, it may be effective for medical associations to provide medical information to patients who need accurate information using social networking sites such as Facebook and Twitter. The act of conducting medical interventions via social media (ie, social media intervention) is reported to be an effective tool for enhancing public health [28]. In the future, it will be important to increase interventions in addition to disseminating medical information via social networking sites. The AAOS has a Twitter account, whereas the JOA currently does not (as of June 2020).

On December 6, 2017, Google made an official announcement about improving the quality of search results for medical and health matters [29]. After this statement was made, information from health professionals, specialists, and medical institutions should have become more readily available, and information sources such as blogs and websites dealing with folk medicine should have been excluded from searches. However, the websites of the two associations we focused on rarely appeared among the top 10 results, indicating that no dramatic change has in fact taken place. In that announcement, Google stated that health experts should make their websites appropriate for general users by avoiding the use of technical terms. Fortunately, personal blogs and unofficial websites that may have posted incorrect medical information tended not to appear in the search results owing to Google's corporate efforts.

The findings of our study suggest that better SEO is required to boost the ranking of medical associations' websites. If their websites were to achieve the highest ranking, ordinary patients could more readily obtain appropriate medical information. Of course, it is also necessary to include accurate medical information on the associations' patient information pages. An SEO strategy comprises two categories: on-page and off-page SEO. Details about this may be found in the Google Search Engine Optimization Starter Guide [30] and the Google Webmaster Guidelines [31]. Unfortunately, development of an SEO takes time. Daily updates and accurate articles need to be posted online to raise the rankings of the medical associations. In the future, a dual approach may be necessary, involving passive information transmission through companies such as Google, and active information transmission using social networking services such as Twitter and Facebook.

Acknowledgments

We thank Edanz Group for editing a draft of this manuscript.

Conflicts of Interest

None declared.

References

1. Burke E, Harkins P, Saeed M, Salama M, Ahmed I. "Dr. Google" Will See You Now-Assessing the Quality of Information on Oesophageal Cancer on the Internet. J Gastrointest Surg 2019 Nov 25:Epub ahead of print. [doi: 10.1007/s11605-019-04416-5] [Medline: 31768833]
2. Murphy EP, Fenelon C, Murphy F, Baig MN, Murphy RP, Diack M, et al. Does Google™ Have the Answers? The Internet-based Information on Pelvic and Acetabular Fractures. Cureus 2019 Oct 21;11(10):e5952 [FREE Full text] [doi: 10.7759/cureus.5952] [Medline: 31799093]
3. Okuhara T, Ishikawa H, Kato M, Okada M, Kiuchi T. A qualitative analysis of the beliefs of Japanese anti-influenza vaccination website authors. Heliyon 2018 Apr;4(4):e00609 [FREE Full text] [doi: 10.1016/j.heliyon.2018.e00609] [Medline: 29862370]
4. Pehora C, Gajaria N, Stoute M, Fracassa S, Serebale-O'Sullivan R, Matava CT. Are Parents Getting it Right? A Survey of Parents' Internet Use for Children's Health Care Information. Interact J Med Res 2015 Jun 22;4(2):e12 [FREE Full text] [doi: 10.2196/ijmr.3790] [Medline: 26099207]
5. Ministry of Internal Affairs and Communications Japan. Looking Back at the History of the Digital Economy in the Heisei Era. WHITE PAPER Information and Communications in Japan. Tokyo: The Ministry of Internal Affairs and Communications Japan; 2019. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r01/pdf/01honpen.pdf [accessed 2020-02-28]
6. Chosa E. [Locomotive syndrome and frailty. Locomotive syndrome due to the underlying disease of degenerative arthritis]. Clin Calcium 2012 Apr;22(4):49-57. [Medline: 22460511]
7. Director-General for Statistics and Information Policy HSO. Patient's Behavior Survey 2017 (Final Report). Tokyo: Ministry of Health, Labour and Welfare; 2019 Mar 15. URL: https://www.mhlw.go.jp/toukei/saikin/hw/jyuryo/17/kakutei.html [accessed 2020-07-13]
8. StatCounter. Search Engine Share Worldwide. URL: http://gs.statcounter.com/search-engine-market-share [accessed 2020-03-09]
9. King C, Judge C, Byrne A, Conlon N. Googling Allergy in Ireland: Content Analysis. J Med Internet Res 2020 May 13;22(5):e16763 [FREE Full text] [doi: 10.2196/16763] [Medline: 32401220]
10. Ssendikaddiwa J, Lavergne R. Access to Primary Care and Internet Searches for Walk-In Clinics and Emergency Departments in Canada: Observational Study Using Google Trends and Population Health Survey Data. JMIR Public Health Surveill 2019 Nov 18;5(4):e13130 [FREE Full text] [doi: 10.2196/13130] [Medline: 31738175]
11. Kamiński M, Łoniewski I, Misera A, Marlicz W. Heartburn-Related Internet Searches and Trends of Interest across Six Western Countries: A Four-Year Retrospective Analysis Using Google Ads Keyword Planner. Int J Environ Res Public Health 2019 Nov 20;16(23):4591. [doi: 10.3390/ijerph16234591] [Medline: 31756947]
12. Seo D, Shin S. Methods Using Social Media and Search Queries to Predict Infectious Disease Outbreaks. Healthc Inform Res 2017 Oct;23(4):343-348 [FREE Full text] [doi: 10.4258/hir.2017.23.4.343] [Medline: 29181246]
13. The Japanese Orthopaedic Association. Patient Information. 2019. URL: https://www.joa.or.jp/public/index.html [accessed 2020-03-09]
14. The American Academy of Orthopaedic Surgeons. OrthiInfo. 2020. URL: https://orthoinfo.aaos.org [accessed 2020-03-09]
15. Internet Marketing Ninjas. Google Organic Clickthrough Study: Comparison of Google's CTR by position, industry, and query type. 2017. URL: https://www.internetmarketingninjas.com/additional-resources/google-ctr-white-paper.htm [accessed 2020-03-09]
16. Scullard P, Peacock C, Davies P. Googling children's health: reliability of medical advice on the internet. Arch Dis Child 2010 Aug;95(8):580-582. [doi: 10.1136/adc.2009.168856] [Medline: 20371593]
17. Wainstein BK, Sterling-Levis K, Baker SA, Taitz J, Brydon M. Use of the Internet by parents of paediatric patients. J Paediatr Child Health 2006 Sep;42(9):528-532. [doi: 10.1111/j.1440-1754.2006.00916.x] [Medline: 16925539]
18. Pew Internet & American Life Project. Health Information Online. 2005 May 17. URL: https://www.pewresearch.org/internet/2005/05/17/health-information-online/ [accessed 2020-03-09]
19. Pew Internet & American Life Project. Wired for Health: How Californians Compare to the Rest of the Nation. 2003 Dec 15. URL: https://www.chcf.org/publication/wired-for-health-how-californians-compare-to-the-rest-of-the-nation/ [accessed 2020-07-13]
20. Yardi S, Caldwell PH, Barnes EH, Scott KM. Determining parents' patterns of behaviour when searching for online information on their child's health. J Paediatr Child Health 2018 Nov;54(11):1246-1254. [doi: 10.1111/jpc.14068] [Medline: 29864197]
21. Okuhara T, Ishikawa H, Kato M, Okada M, Kiuchi T. A qualitative analysis of the beliefs of Japanese anti-influenza vaccination website authors. Heliyon 2018 Apr;4(4):e00609 [FREE Full text] [doi: 10.1016/j.heliyon.2018.e00609] [Medline: 29862370]

22. Downs JS, de Bruin WB, Fischhoff B. Parents' vaccination comprehension and decisions. Vaccine 2008 Mar 17;26(12):1595-1607. [doi: 10.1016/j.vaccine.2008.01.011] [Medline: 18295940]
23. Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 2009;16(4):471-479 [FREE Full text] [doi: 10.1197/jamia.M3059] [Medline: 19390105]
24. Kortum P, Edwards C, Richards-Kortum R. The impact of inaccurate Internet health information in a secondary school learning environment. J Med Internet Res 2008 Jun 30;10(2):e17 [FREE Full text] [doi: 10.2196/jmir.986] [Medline: 18653441]
25. O'Keeffe ST. Summertime blues? A re-examination of the seasonality of web searches for restless legs and leg cramps. Sleep Med 2017 Sep;37:119-123. [doi: 10.1016/j.sleep.2017.06.017] [Medline: 28899521]
26. Liu F, Allan GM, Korownyk C, Kolber M, Flook N, Sternberg H, et al. Seasonality of Ankle Swelling: Population Symptom Reporting Using Google Trends. Ann Fam Med 2016 Jul;14(4):356-358 [FREE Full text] [doi: 10.1370/afm.1953] [Medline: 27401424]
27. Patel NJ, Nalluri N, Deshmukh A, Pant S, Shah N, Badheka AO, et al. Seasonal trends of heart failure hospitalizations in the United States: a national perspective from 2000 to 2011. Int J Cardiol 2014 May 15;173(3):562-563. [doi: 10.1016/j.ijcard.2014.03.122] [Medline: 24713458]
28. Bonar EE, Schneeberger DM, Bourque C, Bauermeister JA, Young SD, Blow FC, et al. Social Media Interventions for Risky Drinking Among Adolescents and Emerging Adults: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2020 May 13;9(5):e16688 [FREE Full text] [doi: 10.2196/16688] [Medline: 32401225]
29. Shin N. Improvement of search results quality about medical and health contents Internet: Google. 2017 Dec 6. URL: https://webmaster-ja.googleblog.com/2017/12/for-more-reliable-health-search.html [accessed 2020-03-09]
30. The Google Search Quality team. Search Engine Optimization (SEO) Starter Guide. URL: https://support.google.com/webmasters/answer/7451184?hl=en [accessed 2020-03-09]
31. The Google Search Quality team. Webmaster Guidelines. URL: https://support.google.com/webmasters/answer/35769?hl=en [accessed 2020-03-09]

Abbreviations

AAOS: American Academy of Orthopedic Surgeons
JOA: Japanese Orthopedic Association
SEO: search engine optimization

Edited by G Eysenbach; submitted 18.03.20; peer-reviewed by T Cruvinel, V Gianfredi; comments to author 20.04.20; revised version received 04.06.20; accepted 11.06.20; published 14.08.20.

Please cite as:
Yamaguchi Y, Lee D, Nagai T, Funamoto T, Tajima T, Chosa E
Googling Musculoskeletal-Related Pain and Ranking of Medical Associations' Patient Information Pages: Google Ads Keyword Planner Analysis
J Med Internet Res 2020;22(8):e18684
URL: https://www.jmir.org/2020/8/e18684
doi:10.2196/18684
PMID:32795991

Original Paper

User-Centered Design and Evaluation of a Web-Based Decision Aid for Older Adults Living With Mild Cognitive Impairment and Their Health Care Providers: Mixed Methods Study

Laura-Mihaela Bogza[1,2,3*], BA; Cassandra Patry-Lebeau[2,3*], BA; Elina Farmanova[2,3,4], PhD; Holly O Witteman[1,3], PhD; Jacobi Elliott[5], PhD; Paul Stolee[5], BA, MPA, MSc, PhD; Carol Hudon[6], PhD; Anik M C Giguere[1,2,3], PhD

[1]Laval University Research Centre on Community-Based Primary Health Care, Quebec, QC, Canada
[2]Quebec Centre for Excellence on Aging, Québec, QC, Canada
[3]Department of Family Medicine and Emergency Medicine, Laval University, Québec, QC, Canada
[4]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[5]School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada
[6]School of Psychology, Laval University, Québec, QC, Canada
[*]these authors contributed equally

Corresponding Author:
Anik M C Giguere, PhD
Laval University Research Centre on Community-Based Primary Health Care
Pavillon Landry-Poulin, Door #A-1-2, Office #2416
2525 Chemin de la Canardière
Quebec, QC, G1J 0A4
Canada
Phone: 1 4188214123
Email: anik.giguere@fmed.ulaval.ca

Abstract

Background: Mild cognitive impairment (MCI) is often considered a transitional state between normal and pathologic (eg, dementia) cognitive aging. Although its prognosis varies largely, the diagnosis carries the risk of causing uncertainty and overtreatment of older adults with MCI who may never progress to dementia. Decision aids help people become better informed and more involved in decision making by providing evidence-based information about options and possible outcomes and by assisting them in clarifying their personal values in relation to the decision to be made.

Objective: This study aimed to incorporate features that best support values clarification and adjust the level of detail of a web-based decision aid for individuals with MCI.

Methods: We conducted a rapid review to identify options to maintain or improve cognitive functions in individuals with MCI. The evidence was structured into a novel web-based decision aid designed in collaboration with digital specialists and graphic designers. Qualitative and user-centered evaluations were used to draw on users' knowledge, clarify values, and inform potential adoption in routine clinical practice. We invited clinicians, older adults with MCI, and their caregivers to evaluate the decision aid in 6 consecutive rounds, with new participants in each round. Quantitative data were collected using the Values Clarity and Informed subscales of the Decisional Conflict Scale, the System Usability Scale, the Ottawa Acceptability questionnaire, and a 5-point satisfaction rating scale. We verified their comprehension using a teach-back method and recorded usability issues. We recorded the audio and computer screen during the session. An inductive thematic qualitative analysis approach was used to identify and describe the issues that arose. After each round, an expert panel met to prioritize and find solutions to mitigate the issues. An integrated analysis was conducted to confirm our choices.

Results: A total of 7 clinicians (social workers, nurses, family physicians, psychologists) and 12 older (≥60 years) community-dwelling individuals with MCI, half of them women, with education levels going from none to university diploma, were recruited and completed testing. The thematic analysis revealed 3 major issues. First, the user should be guided through the decision-making process by tailoring the presentation of options to users' priorities using the values clarification exercise. Second, its content should be simple, but not simplistic, notably by using information layering, plain language, and pictograms. Third, the interface should be intuitive and user friendly, utilize pop-up windows and information tips, avoid drop-down menus, and limit the need to scroll down. The quantitative assessments corroborated the qualitative findings.

Conclusions: This project resulted in a promising web-based decision aid that can support decision making for MCI intervention, based on the personal values and preferences of the users. Further ongoing research will allow its implementation to be tested in clinical settings.

(J Med Internet Res 2020;22(8):e17406) doi:10.2196/17406

KEYWORDS

decision aid; mild cognitive impairment; elderly; decision support technique; aging

Introduction

Mild Cognitive Impairment

Mild cognitive impairment (MCI) is characterized by a decline in cognitive functioning that is more pronounced than normal aging but does not significantly compromise activities of daily living. The impairment can affect memory, language, problem solving, or attention, among others [1]. MCI is common in adults aged ≥60 years, with a prevalence of 6.7% among adults aged 60 to 64 years and 25.2% among adults aged 80 to 84 years [2]. MCI is often considered an early manifestation of pathologic cognitive decline, such as dementia [3]. However, as it can have several causes, its prognosis varies considerably and, although there is a clear risk of dementia, it is also possible to see stability and improvement over time [4]. A diagnosis of MCI therefore carries uncertainty and the risk of overtreatment of older adults with MCI who may never progress to dementia [4]. When worried individuals consult their physicians, they need clear information on the options available to prevent cognitive decline and on each option's probable benefits and harms.

Decision Aids

Decision aids are standardized tools that provide this information and support decision making for the best course of action [5,6]. They help people become more informed and engaged in decision making by providing evidence-based information about all available options and their positive and negative outcomes, and by assisting them in clarifying their personal values in relation to the decisions to be made [7]. These tools help improve people's knowledge regarding options and risks, increase decisional comfort, stimulate them to participate in decision making, and improve patient-professional communication when compared with care without the use of decision aids [7]. In recent years, our team has studied the design features of printable decision aids, or *decision boxes*, to help health care professionals (HCPs) and patients understand evidence-based information and promote shared decision making [8-10]. However, little is known about how detailed decision aids need to be to support an evidence-informed and value-based decision [7]. Moreover, decision aids have seldom been tested with older populations [11] or in the context of MCI.

Considering the differences between the decision-making patterns of older and younger adults [12], such as the preference of older adults for fewer options than younger adults [13], lower levels of literacy and numeracy among older adults compared with younger adults [14,15], the presence of caregivers in the decision-making circles of older adults [16,17], and more frequent sensory deficits among older adults compared with younger adults [18], there is a need for a decision aid developed in a user-centered way with input from older adults, family or friend caregivers, and professionals, to empower and support this specific population in health care decision making.

Design of Web-Based Decision Aids for Older Adults With Mild Cognitive Impairment

With the internet becoming a primary source of information for consumers looking for health information and advice [19,20], web-based decision aids need to be investigated further [21]. Although several models of web-based decision aids have been tested, evidence about the best use of interactive features is still limited. A 2016 systematic review of studies conducted with adults found that the use of interactive features in web-based decision aids can help improve decision making in preference-sensitive contexts [22]. The review concluded that features that allowed users to control the order in which they viewed the content, its level of detail, and the type of evidence presented were helpful in supporting decision making. However, it also demonstrated that having too many tailoring options might cause more decisional conflict and decreased knowledge compared with decision aids without these features. The review also concluded that features such as values clarification exercises, feedback, and social support had varying effects on decision making, depending on their design, which points to a need for further studies.

Therefore, this study aimed to identify and apply features that best support values clarification and understanding of evidence in older adults with MCI, and to adjust the level of detail of a web-based decision aid inspired by the decision box template, for individuals with MCI.

Methods

Study Approach and Design

We used a user-centered approach [23] to design a web-based decision aid that improves the users' knowledge, values clarification, and its adoption in routine clinical practice. We asked individuals with MCI, as well as HCPs, to evaluate the web-based decision aid in 6 consecutive rounds using mixed methods comprising semistructured interviews and questionnaires [24-26]. New participants were recruited for each round. We also verified the comprehension of the information using a teach-back method and recorded any usability issues. We recorded the audio and computer screen during the session. After each evaluation round, we analyzed the data and, based on the findings, tailored the web-based decision aid to improve user experience.

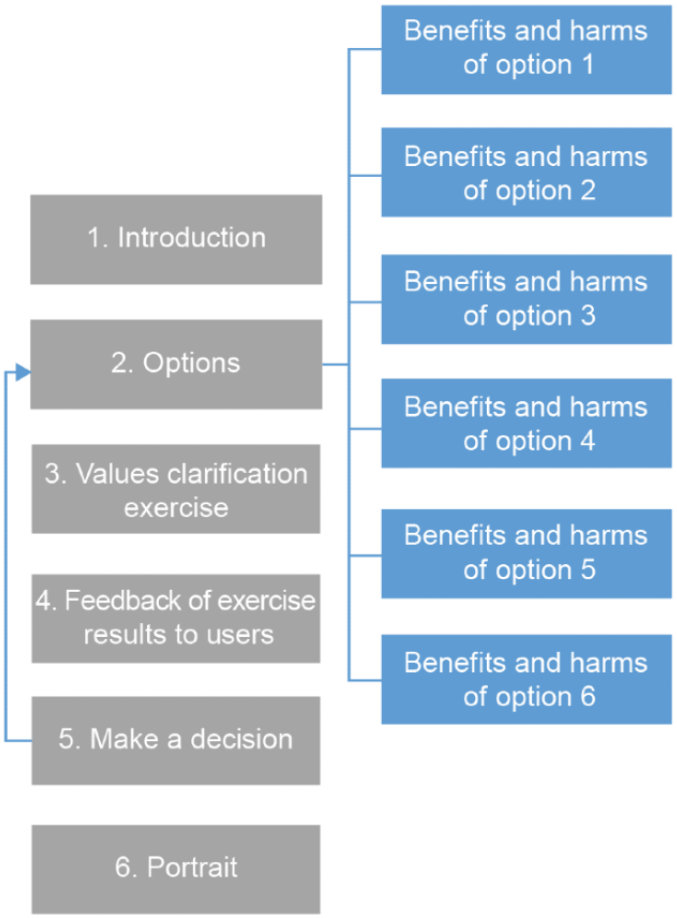
Prototype Conception

On the basis of a published rapid-review approach [27], we synthesized evidence on the benefits and harms of the available options to maintain or improve cognitive functions in older adults with MCI. The evidence gathered was structured into a decision box template, which is in accordance with the international patient decision aid standards and has been described in earlier publications [8,9]. To ensure that the content addressed the users' needs and was scientifically valid, we then formed an expert advisory panel consisting of 8 clinicians (2 nurses, 2 family physicians, a neuropsychologist, a geriatrician, an occupational therapist, and a pharmacist). We invited the members of the panel to review the content, which was then adjusted to take into account their recommendations before testing with potential users.

The informational content was then inserted into a mock-up designed to support older adults with MCI in making informed and value-based decisions with their health care team. More specifically, we designed the web-based decision aid to (1) allow people to weigh the probabilities of experiencing benefits against the probabilities of experiencing harms, for all the available options; (2) allow people to clarify what is most important to them; and (3) support people in voicing their priorities to their health care team. The design process was inspired by steps 3 to 5 of the Center for eHealth and Wellbeing Research guidelines [28,29], which describe the design, operationalization, and evaluation of a tool. Development was initially informed by a review of the literature in the process of developing web-based decision aids and their features [22,30-32] and select examples of web-based evidence summaries and decision aids (eg, drug facts boxes developed by Steven Woloshin and Lisa Schwartz and used by the US Food and Drug Administration [no longer available in a web-based format], MAGICapp developed by Making Grading of Recommendations Assessment, Development, and Evaluation [GRADE] the Irresistible Choice, and option grids developed by Peter Scalia, Glyn Elwyn and Marie-Anne Durand and commercialized by EBSCO Health). Then, in a collaborative and iterative effort, we conducted several brainstorming sessions to discuss the functionality, user experience, and purpose of each of the web-based decision aid's components. The team comprised a computer analyst, 2 graphic designers specialized in website interfaces (one of whom was EB), a researcher specialized in dementia and health literacy (EF), and the principal investigator (AG), who specializes in shared decision making and is the developer of the decision box template.

In this prototype, potential users were guided through a linear decision-making process going from a general introduction to a description of the available options to a values clarification exercise before making a decision (Figure 1). Users could then go back to change their choices as many times as they liked. Once a decision was made, they could view their profile, which summarized their priorities, decisions, and decisional comfort.

Figure 1. Navigation through the various sections of the web-based decision box as planned in the first prototype, before user testing.



Population and Sampling Strategy

Adults aged ≥60 years with MCI, their informal caregivers if they had one, and HCPs of any profession who practiced in family medicine clinics were eligible to participate. We initially invited HCPs who worked in 7 primary care clinics in the areas surrounding Quebec City. We sent them an email to explain the study and invited them to confirm their interest in participating in the study using a link to a web survey. We then asked those who agreed to participate to identify eligible individuals among their patients, who scored between 18 and 26 on the Montreal Cognitive Assessment [33] and to contact them to ask their permission for the research team to contact them. Only the contact details of people who expressed an interest in participating in the study were shared with the research team. As we experienced challenges recruiting participants with this strategy, we also invited people with MCI who were registered in a Quebec database managed by one of us (CH). A diagnostic procedure described elsewhere [34] was used to verify that these participants with MCI met the Petersen criteria for single- or multiple-domain MCI [35], which include (1) a subjective memory complaint, corroborated by an informant; (2) an objective memory impairment, based on a cut-off score of ≥1.5 SD below age-adjusted norms on at least one standardized memory test in the neuropsychological battery; (3) preserved general cognitive functioning; (4) largely intact functional activities; and (5) not demented.

Evaluations

We invited the participants to evaluate the web-based decision aid during a 60-min session, conducted either in participating clinics, the research center, or in their own homes. The duration of the session was adapted to the capacities of each participant.

At the beginning of the session (pretest), all participants completed a self-report questionnaire to record their sociodemographic characteristics, and older adults completed the *Values Clarity* and *Informed* subscales of the Decisional Conflict Scale (DCS) to rate, respectively, their feeling of being clear about personal values for benefits and risks/side effects, and their feelings of being informed [36]. We reversed the DCS subscale scores to facilitate interpretation; increased scores represented increases in feelings of being clear about personal values and being informed. The research assistant was available to support the participants in completing the questionnaire as needed.

Then, they reviewed the web-based decision aid during a semistructured interview that used a think-aloud approach [37]. A trained interviewer followed an interview guide designed to evaluate the participants' information needs and their perception of the strengths and limitations of the web-based decision aid. The guide used to interview older adults also contained questions inspired by the *teach-back* method to assess the users' understanding of the information. Teach-back is used by HCPs and consists of asking patients to explain the information back to them [24,38-41]. The discussions were audio recorded, and we captured the participants' navigation on their computer screens using the video feature.

At the end of the think-aloud evaluation of the web-based decision aid (posttest), all participants completed 5 items of the Ottawa Acceptability questionnaire [42] to assess the quality of decision aids as well as the 10-item System Usability Scale (SUS) [43,44]. They also rated their satisfaction with the web-based decision aid using a single item, 5-point smiley face rating scale. In addition, the HCPs completed a number of questions to explore their willingness to adopt the decision aid in clinical settings [25]. At posttest, the participating older adults once again completed the *Values Clarity* and *Informed* subscales of the DCS.

Tailoring the Web-Based Decision Aid to Users' Needs

After each round, a thematic qualitative analysis of the discussions highlighted the strengths and limitations of the web-based decision aid. One researcher (LB, EF, or CP) conducted the analysis, and another reviewed it (AG). The expert panel then met to prioritize the issues and find solutions to mitigate them. To this end, members of the same interdisciplinary expert panel who created the prototype met to review the qualitative findings. The experts started by prioritizing each issue using the following severity scale [45]: (1) critical flaw: if unresolved, then users will not be able to complete the task; (2) serious flaw: several users may be frustrated and give up; (3) minor flaws: users may be annoyed, but they will complete the task. They then chose solutions by considering the magnitude, frequency, and severity of these problems, and how they prevented the web-based decision aid from achieving its primary goal (supporting evidence-informed and value-based decision making).

The same evaluation and tailoring process was then used again in 5 more rounds, with new participant subsamples each time.

Integrated Analyses

In the final analysis, we integrated the findings from the 6 rounds of qualitative and quantitative evaluations into the web-based decision aid design. This triangulation process served to validate our findings [26,46,47] and to form a comprehensive picture of the web-based decision aid's acceptability and usability and of the level of detail required to support evidence-informed and value-based decision making by older adults with MCI, their caregivers, and HCPs. We determined the level of detail based on the participants' knowledge and action. An appropriate level of detail was interpreted from the decision aid's potential to inform decision making and increase decisional comfort, thereby inciting the user to take action. More specifically, we considered the level of detail as appropriate if the decision aid increased decisional comfort (ie, achieved *100* or a value closer to *100* on the DCS informed subscale at posttest) and if the participants demonstrated high levels of intention to use the decision aid to make a decision following the diagnosis of MCI.

We assessed the suitability of the values clarification exercise using the participants' responses to the DCS *Values Clarity* subscale and to open-ended questions about their experience with the values clarification exercise. We considered the values clarification exercise as suitable if it increased the participants' feelings of being clear about personal values (ie, achieved *100*

on the DCS *Values Clarity* subscale) and if participants (especially those in the fifth and sixth evaluation rounds) reported a positive experience using the decision aid.

Ethical Considerations

Ethical approval to conduct this pilot study was granted by the Integrated University Health and Social Services Center of Québec City (*CIUSSS-Capitale-Nationale*, no. 2016-2017-08).

Results

Participant Characteristics

We recruited 7 HCPs (3 social workers, 2 nurses, 1 physician, and 1 psychologist; Table 1) and 12 community-based older adults with MCI (Table 2). Out of the 12 older adults, 3 did the interview in the presence of a friend or family member, whom they did not consider to be a caregiver, and 9 did it alone with the research assistant. Of which, 2 of the persons with MCI were aged between 60 and 64 years, 6 between 65 and 74 years, 3 between 75 and 84 years, and 1 was >85 years. The participants' education levels varied from none to university degree.

Table 1. Description of the characteristics of the participating health care professionals (N=7).

Characteristic	Values, n (%)
Age (years)	
<30	1 (14)
30-39	4 (57)
50-59	2 (29)
Sex	
Female	6 (86)
Male	1 (14)
Profession	
Physician	1 (14)
Nurse	2 (29)
Social worker	3 (43)
Psychologist	1 (14)
Practice experience (years)	
0-4	2 (29)
5-9	0 (0)
10-14	1 (14)
15-19	2 (29)
20-24	1 (14)
25-29	0 (0)
30-34	0 (0)
35-39	1 (14)

Table 2. Description of the characteristics of the participating older adults with mild cognitive impairment (N=12).

Characteristic	Values, n (%)
Age (years)	
60-64	2 (17)
65-74	6 (50)
75-84	3 (25)
≥85	1 (8)
Sex	
Female	6 (50)
Male	6 (50)
Ethnicity	
White	12 (100)
Primary language	
French	12 (100)
Marital status	
Single	1 (8)
Married or common-law relationship	9 (75)
Widow	2 (17)
Education [a]	
No formal education	1 (8)
High school diploma	4 (33)
College diploma	3 (25)
University diploma	3 (25)
Place of residence	
Home	12 (100)
Has a caregiver	
Yes	2 (17)
If so, who?	
Family member	1 (8)
Friend	1 (8)
Lives with their caregiver	
Yes	1 (8)
No	1 (8)

[a]One of the participants had missing data for education.

We conducted 6 evaluation rounds, starting with the prototype version (V1), and then produced 5 different versions (V2-V6) before creating the final version (VF) that is currently available on the Decision Box website [48]. Of which 6 HCPs evaluated the prototype (V1) and 1 evaluated V2. Additionally, 3, 3, 2, 3, and 1 older adult with MCI evaluated versions 2, 3, 4, 5, and 6, respectively. The participating HCPs reviewed only the first and second mock-up.

Qualitative Findings

A total of 3 major issues emerged from the qualitative analysis: (1) users should be guided through the decision-making process; (2) the content should be simple, but not simplistic; and (3) the interface should be intuitive and user friendly.

Guided Decision-Making Process

As mentioned, the first prototype (V1) proposed a linear directional navigation, going from general to specific information based on users' preferences (Figure 1). It led users to read all available options before clarifying their values and making a decision. As a result, we observed that participants often paused in the *Benefits and Harms* section and did not complete the next steps because they grew tired or lost focus:

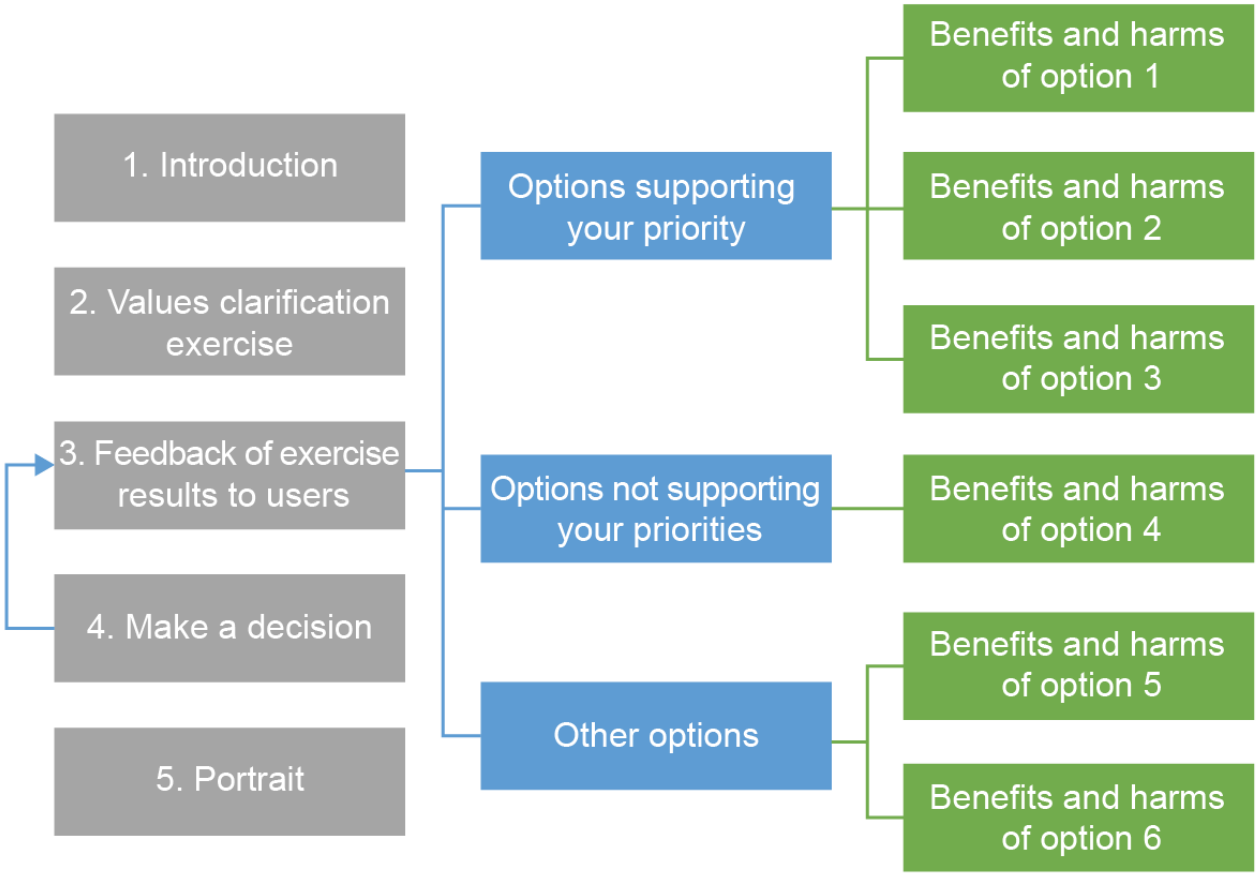
> *Participant: Geez, this is quite something, isn't it? You're making me work hard, you know. Where are we at now? [interviewer explains the next steps]. I don't know whether I'm coming or going.* [Older adult #1001]
>
> *Interviewer: What do you mean?*
>
> *Participant: Well, I...I'm tired, now.*

Hence, to ensure participants would complete the values clarification exercise, we used a nondirectional navigation strategy in V2, allowing users to start either with a review of the options or with the values clarification exercise. This also helped meet the needs of HCPs who visit the web-based decision aid without their patients, to review the evidence about the available options. However, user testing revealed that this nondirectional design confused several users who wondered about the *correct* path:

> *[After the interviewer explains that they can choose whichever path they prefer] So, are we going to explore all the options or uhh…?* [Nurse #3009, V2]

The experts therefore decided to revert to a directional navigation in the next version (V3), leading users to complete the values clarification exercise before reviewing the options (Figure 2), which is in the opposite direction from V1. Once people had clarified their values, the web-based decision aid presented them with the options that matched their values. The other options were still available for review in case they wished to learn more about them.

Figure 2. Navigation through various sections of the web-based decision box as planned after user testing.



Moreover, V3 included clear indications of the next steps to improve usability. Despite directional navigation, the prototype still allowed users to access any page at any time, using a progression bar that we added above the menu.

After V3, the users did not report feeling fatigued, overworked, or lost as they went through the web-based decision aid. This was likely because they could read the options most relevant to them and then move on without feeling that they had missed out. This meant that they could better manage their navigation to match their energy levels and improve their likelihood of making it through all the decision-making steps. Furthermore, it helped simplify the web-based decision aid, as discussed in detail in the following section because it limited the number of options to be read before making a decision. This modification also made the web-based decision aid more patient centered, as opposed to professional centered.

Simple but Not Simplistic

Throughout the various versions, participants consistently asked for a simpler design. Therefore, the challenge was to simplify the web-based decision aid while retaining its informative value to ensure that users would have enough information to make informed and value-based decisions. The amount of information to read (content density) and its complexity (content clarity) were of major concern to the users.

Content Density

Participants made frequent comments about the overwhelming nature of the information, especially in the *Introduction* and *Options* sections of V1 as shown in Multimedia Appendix 1 (Warning: this is a preliminary version and the data is not final. For a final version, please see [48].). In V1, we observed that many participants did not read all the available information in these sections:

> *Oh my! It's going to take so long to read all that.* [Social worker #3004, V1]

To address these concerns, we modified the information layout in the *Introduction* section. The information that was initially structured into columns was put into bulleted lists, in separate accordion tabs that could be expanded by clicking on a + button. In V3, we also removed the references from the text and moved them to a separate page, as shown in Multimedia Appendix 2.

To improve the *Options* section, we tried integrating the options to the values clarification exercise results page. This was meant to allow people to read the information most important to them first when they had the most focus. This strategy offered a more straightforward reading path, but some participants still did not read the information regarding all the options in V4:

Wow, there are a lot of things [talking about all the options]! I won't read all that. [Older adult #1008, V4]

After 5 unsuccessful versions trying to limit the users' burden as they read the options, we classified the options into 3 subcategories: (1) options that matched their priority, (2) options that did not match their priority, and (3) other options to consider. People could only view the subcategories related to a single priority at a time. This prevented us from having to remove information that we deemed essential to support informed, value-based decision making (which would have led to a simplistic design), while still allowing us to highlight the information most relevant to the user.

Content Clarity

Scientific Terms

Several of the HCPs who evaluated the first version found the content too scientific and potentially difficult to understand for laypeople, as pointed out by this participant:

> *I mean, maybe a doctor or a nurse will be able to understand, but me, I don't have… I don't have any medical training. I'm a social worker, so when you show me terms like that,… you lose me.* [Social worker #3004, V1]

Users did not like the style either, as pointed out by this participant:

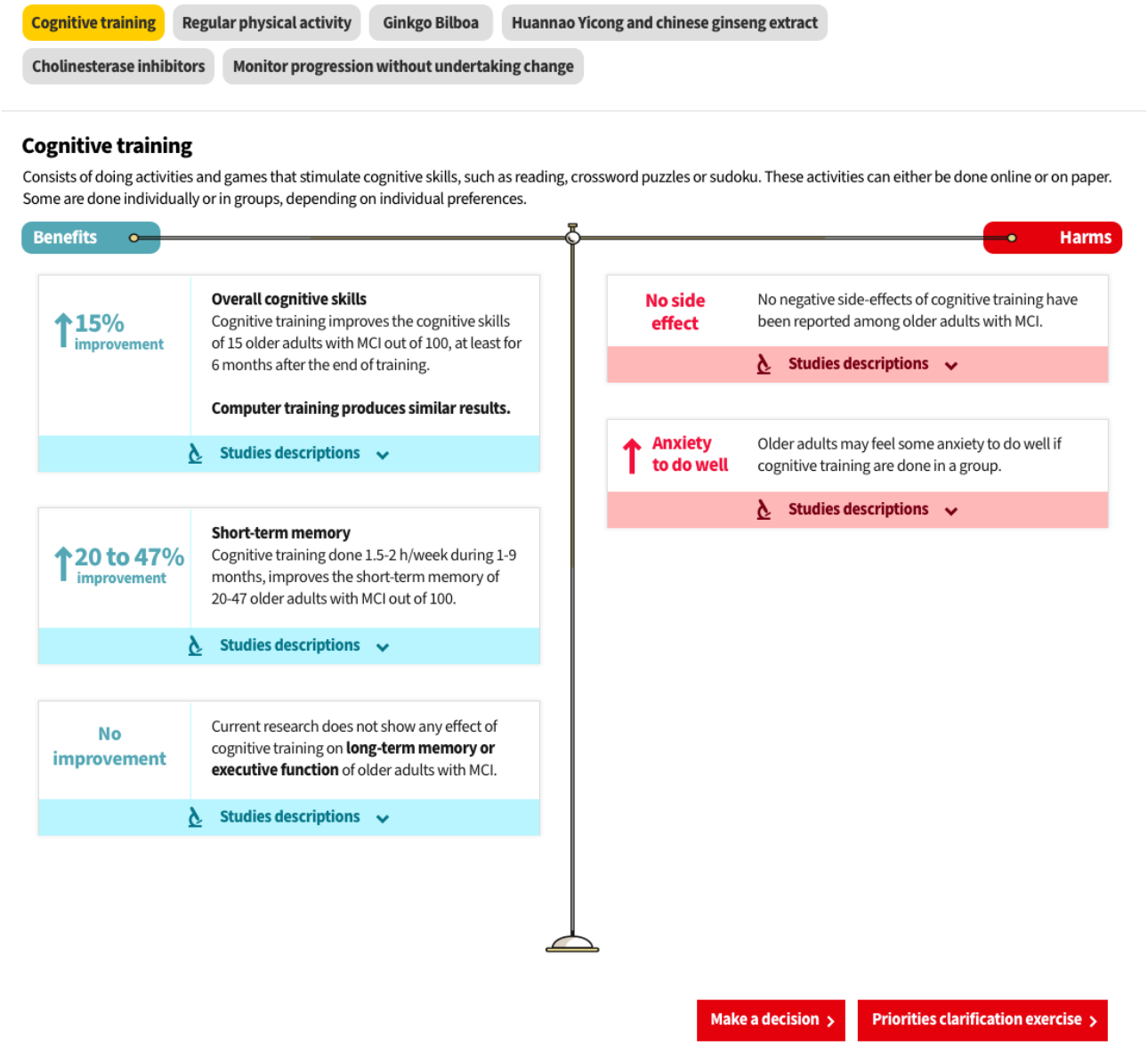
> *It's very scientific. I don't identify with the text.* [Older adult #1008, V4]

The content was progressively modified to use more lay terms, as demonstrated by the change of the decision aid title from Mild Cognitive Impairment to Mild Problems with Thinking and Memory in the final version.

Probabilities

Information clarity was also jeopardized by the users' misunderstanding of the probabilities presented in the *Options* section (V2, V3, V4, and V6). In fact, most participants misunderstood the absolute effect sizes presented in larger font sizes, in percentages, in V1 (Figure 3; warning: this is a preliminary version and the data is not final. For a final version, please see [48]). They interpreted them as mean effect sizes (eg, a 15% increase in mental capacities) while in reality they represent the proportion of people who might benefit from the intervention (15 persons out of 100 improve their mental capacities):

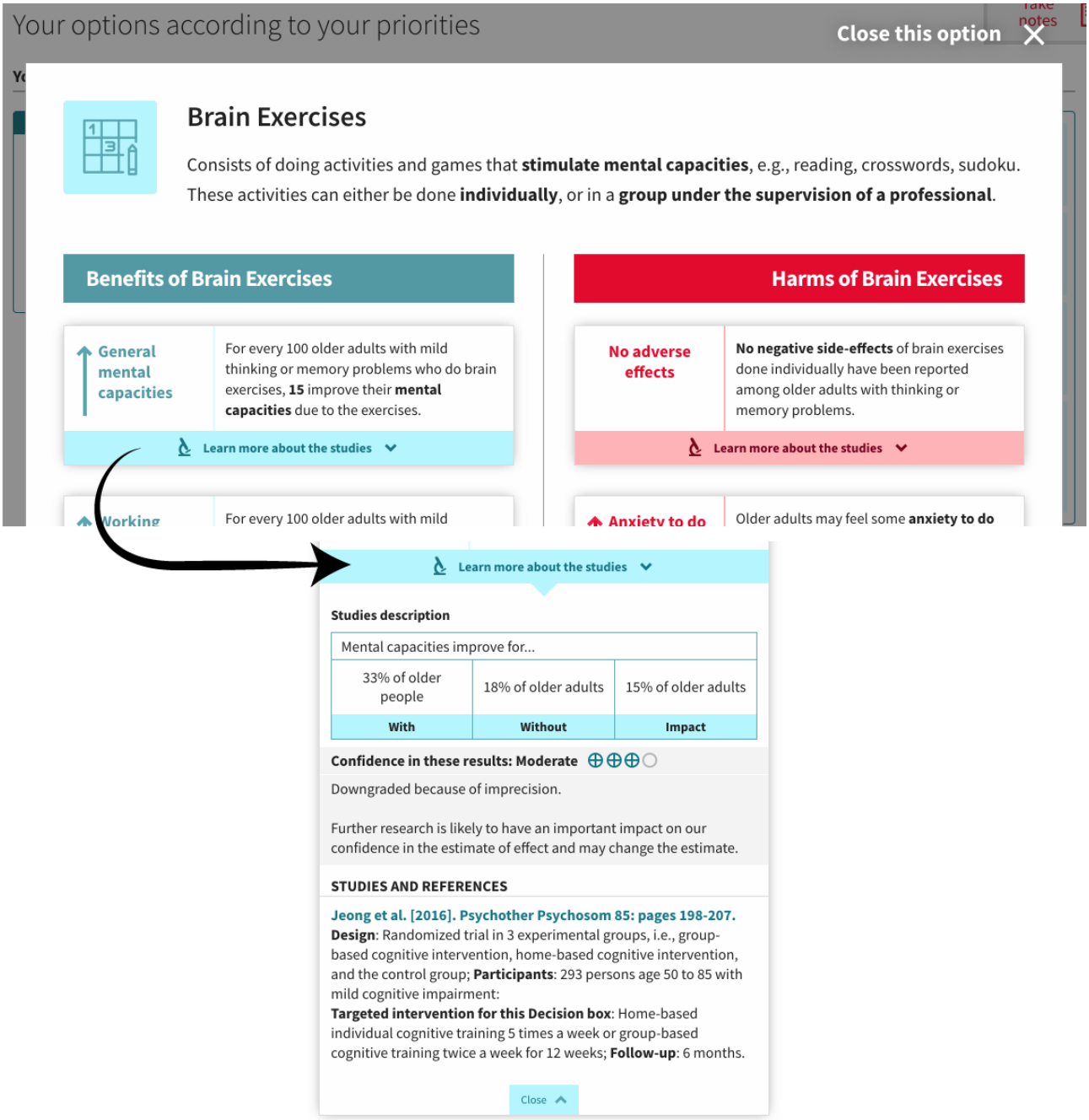
Figure 3. Presentation of one of the options in prototype V2. Warning: This is a preliminary version, and the data are not final.



> *It can improve by 11% compared to my… compared to now.* [Older adult #1004, V3]

To minimize influence on decision making, we replaced numbers and percentages with text in the final version. For example, we replaced "↑15% improvement" with "↑ General mental capacities" in a slightly larger font size (Figure 4), and we presented the detailed proportions (*for every 100 older adults, 15 improve their mental capacities*) in smaller font on the right-hand side.

Figure 4. Presentation of one of the options in the final version showing the pop-up format and with an enlargement showing an example of the expanded accordion tabs to learn more about the grading of recommendations assessment, development, and evaluation ratings.



Grading of Recommendations Assessment, Development, and Evaluation Ratings on Confidence in the Results

Participants also often misunderstood the GRADE ratings associated with each effect size in the earlier versions of the web-based decision aid (V1 and V4):

> *Oh, this is high: 4 out of 4 stars! It's like when you choose a holiday spot down south!* [Older adult #1007, V4]

Therefore, we layered the GRADE ratings, allowing easy access by the most curious users under a *Learn more about the studies* accordion tab placed under each option in V5 of the web-based decision aid (Figure 4).

Feedback After the Values Clarification Exercise

As mentioned earlier, the feedback provided to users after they had completed the values clarification exercise (Figure 5) required several modification rounds before it was understood. The goal was to present the options that matched the person's values. The options were initially shown in a matrix table, which most users misunderstood (Figure 6; warning: this is a preliminary version and the data is not final. For a final version, please see [48]), as demonstrated by this participant's comments:

Figure 5. First prototype version of the web-based decision aid, showing the values clarification exercise.

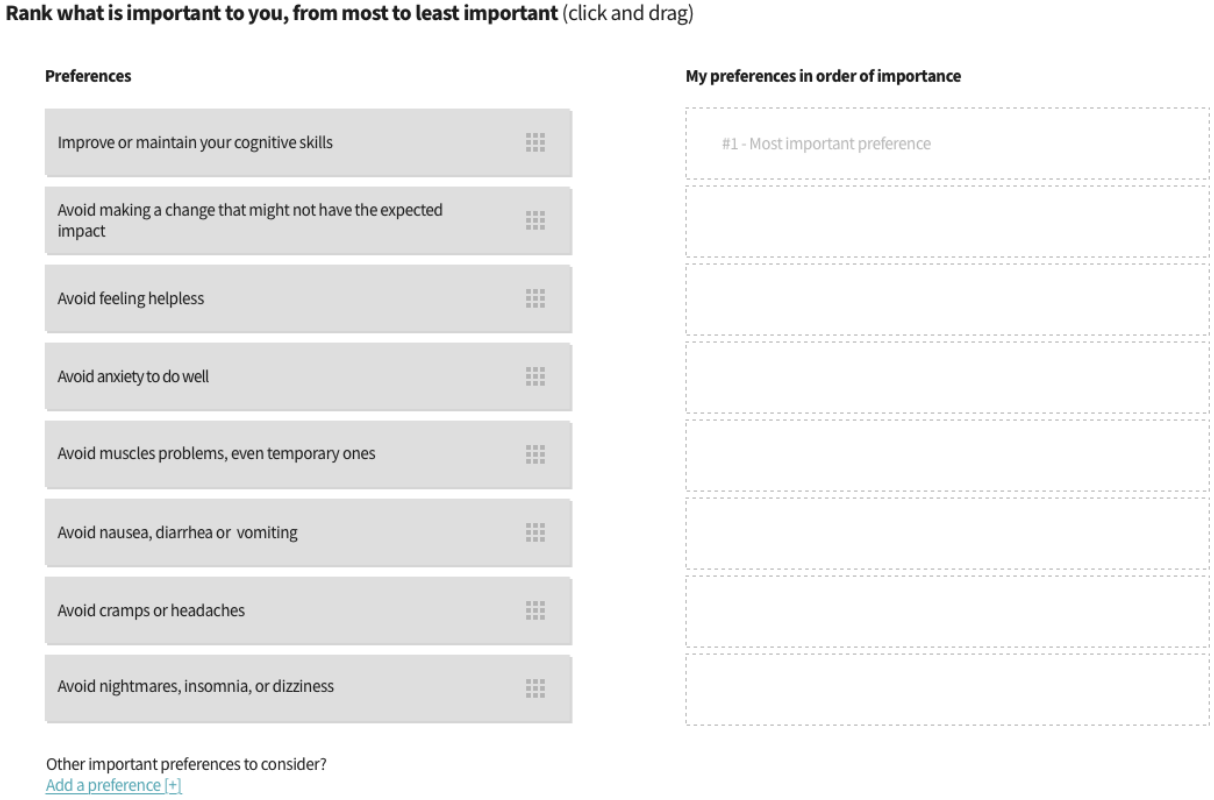
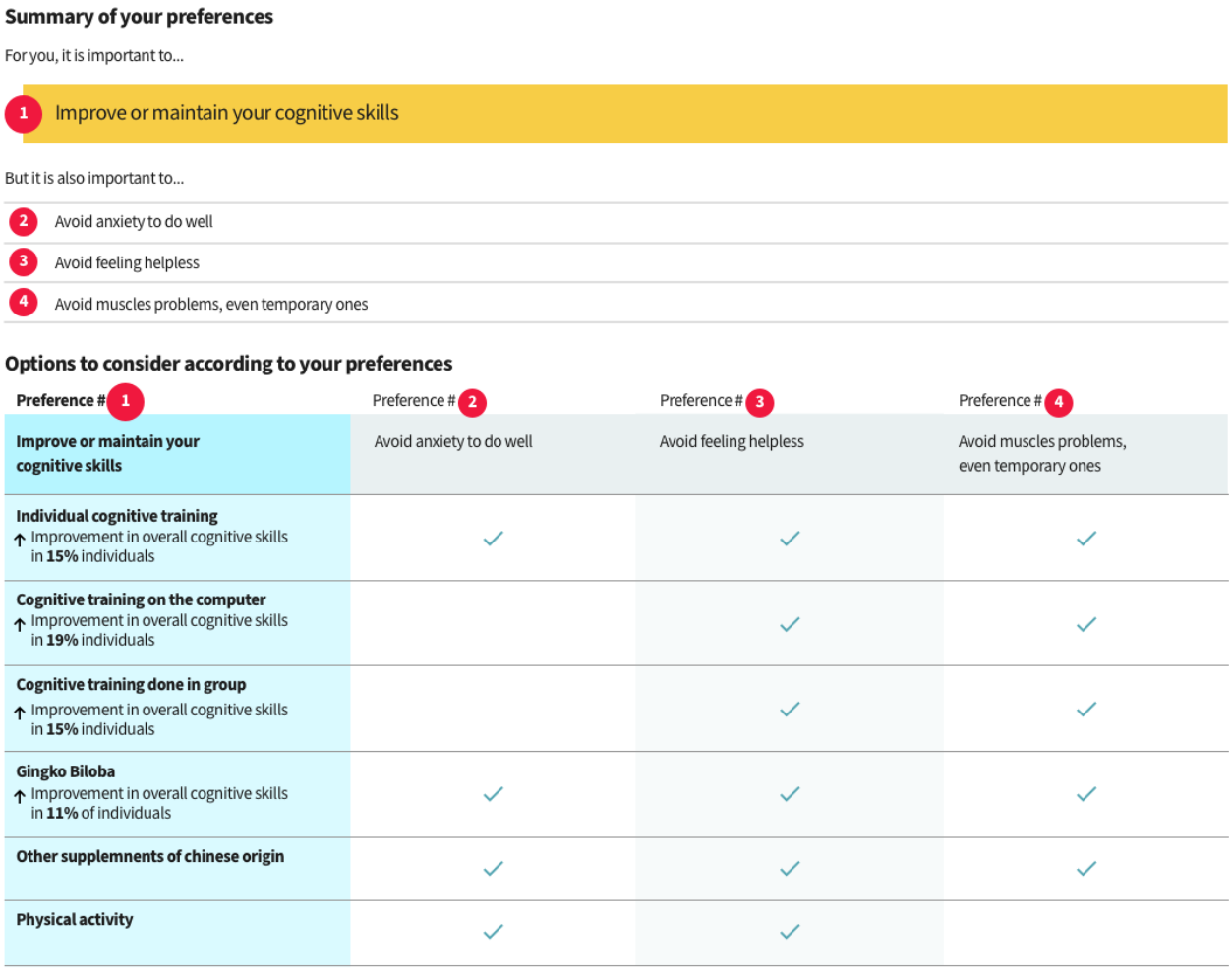


Figure 6. First prototype version of the web-based decision aid, showing the feedback page after the exercise. Warning: This is a preliminary version, and the data are not final.



Interviewer: Did you think the check marks were choices you had to make?

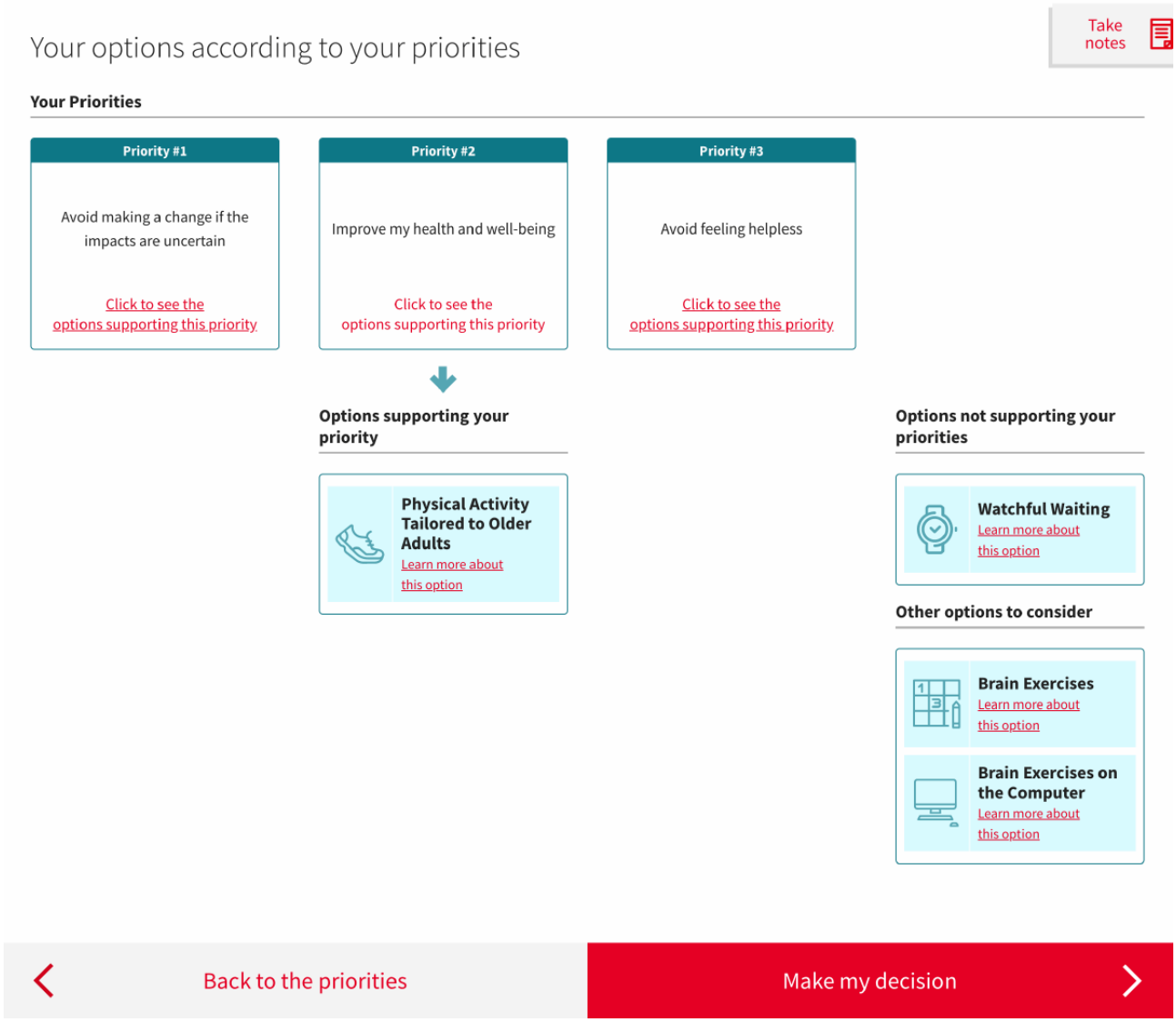
Participant: Yeah, yeah...I thought I was the one placing the check marks. [Older adult #1006, V3]

The modifications made to the matrix table in the next round did not improve the participants' understanding, as demonstrated by this quote:

Here, what am I supposed to do? [...] I have to make a check mark...Should I make a check mark? What is it? [...] I see there's an X. If I click, will it give me a check mark? [Older adult #1007, V4]

Therefore, we decided to limit the feedback to its simplest form. We discarded the matrix table in V5, and instead listed the options as a series of boxes underneath each of the values (described hereafter as *priorities*) selected by the person, with a maximum of 3 values (Figure 7). We kept a visual connection between the values clarification exercise (Figure 8) and the feedback by using the same format in both pages for the values. We also reduced the number of values presented from 8 to 4, a decision that will be explained later in the *Values Clarification Exercise* section.

Figure 7. Final version of the web-based decision aid, showing the feedback page after the exercise.



We initially included the option of taking cholinesterase inhibitors, as health care providers sometimes consider this medication. However, the scientific evidence described in the decision aid underscored their ineffectiveness in improving cognitive functioning, and this was more confusing than helpful for the participants. Therefore, we removed it from the options available in the subsequent versions. All options lacking evidence of benefits have been removed in the same manner.

Pictograms

Once they reached the *Decision* section, several participants did not recall the information related to each of the available options. We, therefore, added a pictogram representing each option every time it was discussed (eg, Figures 4 and 7). This feature was meant to improve recall, and we did indeed observe some improvements in this regard after it was added:

> *Participant: The definition of...umm of...the sneakers, right?* [Older adult #1012, V6]
>
> *Interviewer: Yes, the physical activity.*
>
> *Participant: Yes, physical activity.*

Intuitive and User-Friendly Interface

Interface Features

We incorporated several modifications, at all stages, to design an intuitive and user-friendly web-based decision aid. For example, we defined more complex words in a glossary. We initially placed the glossary on a separate page at the end of the web-based decision aid, which was accessible when users clicked on a complex word. However, some users became confused when redirected to a different page; they forgot why they were there and did not know how to get back to the main activity. We therefore arranged for definitions to appear as information tips when the cursor hovers over a word, as shown in Multimedia Appendix 3. This way, users could see the definition of a word while remaining on the *Introduction* page.

We also had to add instructions on how to expand the accordion tabs, as participants did not understand at first (Multimedia Appendix 1):

> *Participant: [reading out loud] Take into account the patient's preferences. [to the interviewer] Do I...What do I do now...?* [Older adult #1005, V3]

Interviewer: Yes, here you could click, it's a link to open a section with a bit more information about that.

To this end, we added a Learn more link next to the + sign to indicate the possibility of expanding the accordion (Multimedia Appendix 2).

Several users could not find a way to return to the previous page after clicking on an option to read more information about it (Figure 3). Thus, we showed the benefits and harms of the options in a pop-up instead of on a standalone page (Figure 4). This facilitated navigation, as users simply had to close the option to get back to the main page.

We also modified another feature related to the *Decision* section, where users were invited to select an option. We used a drop-down menu in V1 (Multimedia Appendix 4); however, the task was not understood by all participants. We therefore replaced the drop-down menu with checkboxes from V3 onward (Multimedia Appendix 5), which resolved the issue:

Participant: Do we do all that? [Older adult #1001, V2]

Interviewer: Have you seen the menu? There is a menu. It says you have to select the option you prefer.

Participant: Oh. My goodness!

Interviewer: If you click on… [interviewer shows the participant how to use the drop-down menu].

Values Clarification Exercise

In the first version of the values clarification exercise tested by older adults, most participants needed explanations from the moderator to complete the exercise:

Participant: [Participant reads the priorities out loud] Now you want me to look at that, what do you want to know about what I'm looking at? [Older adult #1002, V2]

[Interviewer explains the exercise]

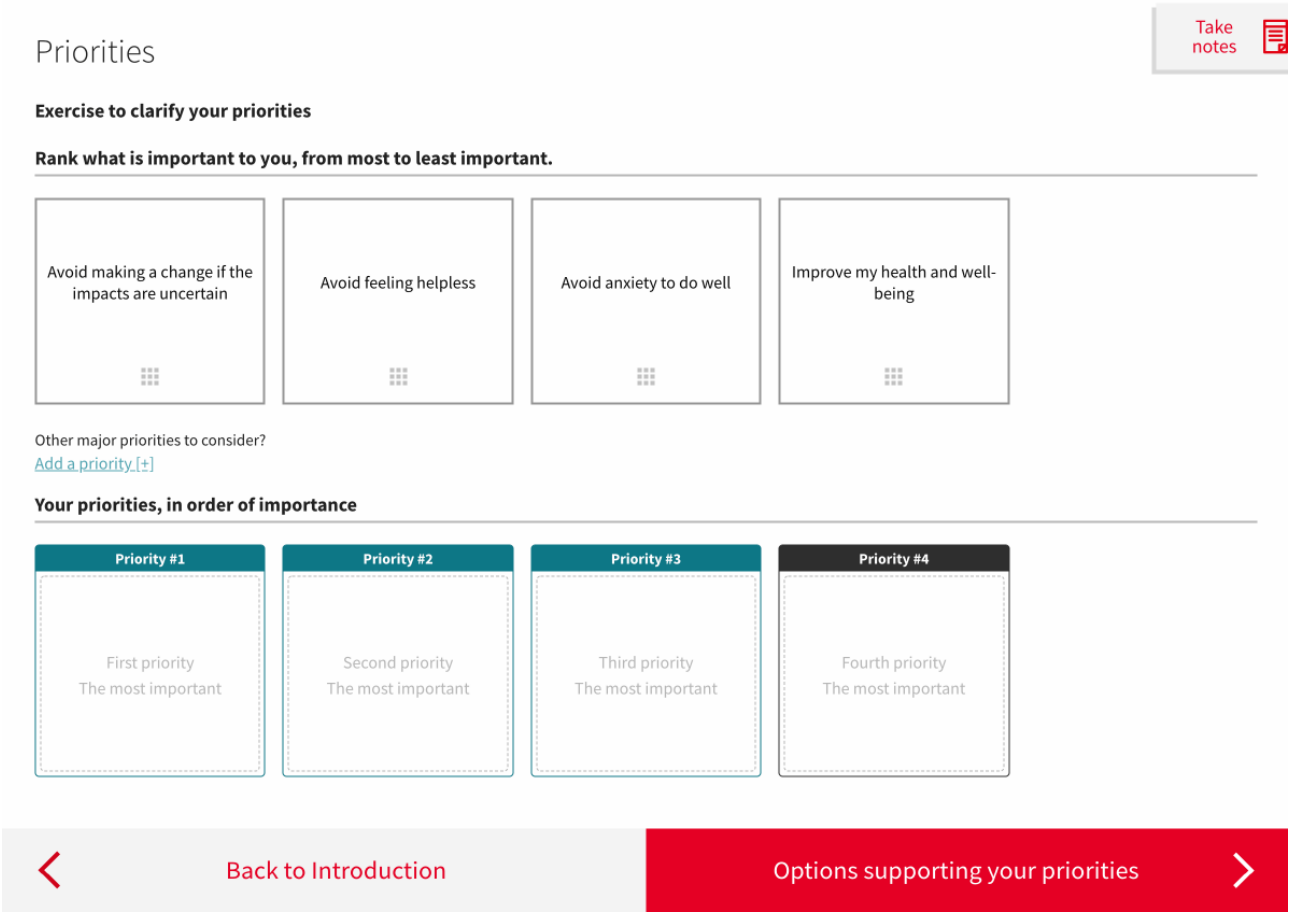
Participant: Uhh, let's say I put one here, then two...

Caregiver: Actually, aren't you supposed to move this here?

Interviewer: Exactly.

Several of the participating HCPs felt that there were too many priorities and that the wording was too complex, partly because they were worded negatively. Therefore, we changed the wording and reduced the number of priorities from 8 (Figure 5) to 4 (Figure 8). We also changed the format of the exercise from vertical (Figure 5) to horizontal (Figure 8), primarily to allow the content to fit most computer screens without scrolling down, but also to improve continuity with the feedback, which also displayed the selected priorities horizontally.

Figure 8. Final version of the web-based decision aid, showing the values clarification exercise.



Quantitative Results

Assessments of the Web-Based Decision Aid

The limited number of participants did not allow for any statistical testing to compare the users' assessment of the web-based decision aid, or impacts of the decision aid, between rounds, or between types of users (Table 3). There is no consistent trend, as the rounds progressed, for any of the studied variables.

Impacts of the Web-Based Decision Aid

We did, however, have sufficient data to compare the mean of all users' ratings on the DCS before and after using the decision aid. We observed statistically significant improvements in older adults' feelings of being clear about personal values and being informed before and after their use of the web-based decision aid (one-tailed t-test for values clarity: t_{11}=4.16; P<.001 and informed: t_{11}=1.92; P=.04).

Integrated Analyses

Several changes made to V1 addressed comments that the decision aid was too complex and too dense. These comments were translated to one of the lowest usability scores across all versions (Table 3).

Table 3. Assessments of the successive versions of a web-based decision aid by mixed samples of older adults and health care professionals and impact of the decision aid on the perception of older adults of the clarity of their values and of being well-informed before (pre) and after (post) using the decision aid.

Outcome	Version 1 (n=6 HCP[a])	Version 2 (n=3 OA[b] and 1 HCP)	Version 3 (n=3 OA)	Version 4 (n=2 OA)	Version 5 (n=3 OA)	Version 6 (n=1 OA)	Total (n=12 OA and 7 HCP)
Willingness to use the decision aid—post[c], mean (SD)	3 (1)	4 (0)	N/A[d]	N/A	N/A	N/A	N/A
System usability scale—post[e], mean (SD)	56 (23)	65 (20)	77 (10)	56 (16)	75 (16)	55 (0)	N/A
Acceptability—post[f], mean (SD)	76 (9)	84 (6)	92 (8)	82 (9)	86 (4)	73 (0)	N/A
Satisfaction—post[g], mean (SD)	4 (1)	4 (1)	5 (0)	3 (1)	4 (1)	5 (0)	N/A
Values clarity (DCS[h])—pre[i], mean (SD)	N/A	64 (43)	64 (10)	38 (41)	75 (17)	50 (0)	61 (27)
Values clarity (DCS)—post[i], mean (SD)	N/A	83 (8)	83 (14)	63 (30)	92 (8)	92 (0)	83 (16)
Pre-post mean values clarity (DCS) score increase	N/A	19	19	25	17	42	22
Informed (DCS)—pre[j], mean (SD)	N/A	61 (54)	61 (24)	54 (30)	72 (27)	67 (0)	63 (30)
Informed (DCS)—post[j], mean (SD)	N/A	69 (10)	92 (8)	79 (6)	92 (14)	67 (0)	82 (14)
Pre-post mean informed (DCS) score increase	N/A	8	31	25	20	0	19

[a]HCP: health care practitioner.

[b]OA: older adult.

[c]Scale from 1 to 5, with 5 indicating higher willingness. Answered only by HCP.

[d]N/A: not applicable, certain measures were used only with HCP or OA.

[e]Scale from 0 to 100, with 100 indicating higher usability. Answered by both OA and HCP.

[f]Proportion of people who find the web-based decision aid acceptable. Answered by both OA and HCP.

[g]Scale from 1 to 5, with 5 indicating higher satisfaction. Answered by both OA and HCP.

[h]DCS: decisional conflict scale.

[i]Scale ranging from 0 to 100, with 100 indicating higher clarity. Answered only by OA.

[j]Scale ranging from 0 to 100, with 100 indicating. Answered only by OA.

Comments on V2 came from 3 older adults and 1 HCP. Issues with content density were still prominent, as evidenced by one of the lowest increases in participants' feelings of being informed before and after using the decision aid.

Changes in V3 increased the postdecisional comfort score on the *Informed* subscale from 61.1 to 91.7, which reflects the effort put into facilitating navigation. Version 3 had the highest acceptability (92%), usability (mean 76.7, SD 10.1), and satisfaction (mean 5.0, SD 0.0) scores, which confirmed the

team's efforts to layer the web-based decision aid to make it more manageable. The older adults still expressed misunderstanding of the values clarification exercise feedback page, which was the focus of the alterations in the next versions.

We made a final attempt at fixing the matrix table used on the VCE feedback page in V4. The older adults who evaluated this version presented relatively lower values clarity and perceptions of being informed before using the decision aid, and the qualitative comments indicated some frustration with the matrix table (the crux of the web-based decision aid). This seems to have influenced their assessments of the decision aid, as it scored the lowest in satisfaction and garnered one of the lowest usability scores.

In V5, we introduced a simpler values clarification exercise feedback format, which coincided with the highest posttest values for the *Informed* and *Values Clarity* measures, although the participants also had the highest scores in both of these subscales before using the decision aid. Nonetheless, these scores correspond to relatively high usability, acceptability, and satisfaction, and the comments appear to indicate that users understood this version of the feedback the best.

Version 6 introduced pictograms and a better classification of options to improve information recall, but it was tested with a single user only. The older adult that tested version 6 appreciated the usefulness of the pictograms. The highest increase in decisional comfort on the *Values Clarity* subscale was also measured in this round ($mean_{pre}$ 50, $mean_{post}$ 91.7), which suggests that the new feedback design achieved its goal. Usability (mean 55.0) and acceptability (mean 71.4) scores are, however, the lowest across all versions, but several comments made during data collection led us to believe that some of the SUS items may have been interpreted incorrectly. For example, to the SUS question "I think that I would need the support of a technical person to be able to use this system," the sole participant that tested version 6 mentioned appreciating having access to technical resources in general, and thus gave a low usability score. These relatively low SUS scores also contrast with the user's high level of satisfaction (mean 5.0) with the decision aid.

Users consistently demonstrated enthusiasm with regard to decision aid across all versions. HCPs sometimes expressed concerns about being able to integrate it into their practice, and older adults were not always sure that clinicians would care or have time to go through the whole process with them.

Research Process

Throughout the data collection process, several observations were made. First, asking users to review a mock-up instead of a finished product requires much more involvement from the interviewer, which prevented us from observing real user interaction with the web-based decision aid. Second, having the users communicate their difficulties was not always easy because they generally felt that they understood the content even when it was clear to the interviewer that they had not. This highlights the importance of the evaluation approaches used in this project because observations of users interacting with the decision aid and teach-back allowed more issues with the web-based decision aid to be raised, compared with direct comments from users regarding their appreciation of it. Finally, spending time with community-based users and listening to them talk about their experiences and share their stories while navigating the web-based decision aid emphasized the wealth of information available from direct contact with users. Many modifications made to the web-based decision aid were first suggested by users, such as the addition of pictograms, the use of arrow buttons at the bottom of each page, and enlargement of the font size.

Discussion

Principal Findings

This study aimed to adjust the level of detail of a web-based decision aid and to apply features that best support evidence-informed and value-based decision making for older adults with MCI. Our main findings indicate that, to achieve this goal, the web-based decision aid needs to provide a guided decision-making process, the content needs to be simple but not simplistic, and the interface and navigation should be intuitive and user friendly.

Guided Decision-Making Process

First, we found that a patient-centered web-based decision aid that helps adults >60 years with MCI to make a value-based decision performs better when it helps users clarify their priorities early on in the process. This allows for a tailored presentation of content that guides the user toward the most relevant options for their specific needs. Presenting fewer options at a time helps to avoid overwhelming the user with reading material about all available options, especially when there are several. Moreover, a study that was conducted with adults of various ages found that the first option presented in a decision aid was more likely to be chosen as a treatment option, especially when there were several options [49]. This suggests that it is important to present options in a way that reflects the user's priorities, especially if their cognitive abilities may be impaired. It also coincides with the findings from a 2016 study that showed that a linear navigation had a higher success rate, lower performance time, better satisfaction ratings, and greater user preference than hypertextual navigation for elderly users [50]. In addition, it converged with studies showing that older people prefer having fewer options [13].

However, tailoring must be performed carefully to avoid removing essential information, as this could lead to bias in the decision-making process, increased decision conflict, and decreased knowledge [22]. More specifically, if a values clarification exercises is proposed to tailor information about options in a decision aid, then the aid should still allow access to information on all options as patients need sufficient knowledge of the options before being able to clarify their values [51].

Considering that web-based decision aids can also be used by HCPs and caregivers, in addition to older adults with MCI, they should remain flexible despite their general linearity, so people who prefer to either skip or read some sections exhaustively have the option to do so. In brief, the web-based decision aid

should be directive rather than restrictive. This implies that, in the case of cognitively impaired people, the decision relies more heavily on the accuracy of the values clarification exercise, and therefore on the accurate pairing of each option with the priorities they are designed to meet.

Simple, Not Simplistic

Next, our team resorted to information layering to simplify the design while retaining the level of detail to support informed, shared decision making. This helped provide a straightforward layout that allowed users to understand the decision aid's structure. The text was also modified to use clear and comprehensible wording to retain only meaningful content and remove overwhelming details. These results converge with the findings of Peters [12] that *less is more* when it comes to making informed and high-quality decisions, especially in people with lower numeracy skills. Considering that people with MCI often have episodic memory impairments [52], it is important that the information be concise, meaningful, and easy to remember. Our results indicate that the pictograms are a useful strategy to help users remember the options available and, in the process, summarize the information into a visual reminder. These findings converge with those found in literature reviews, indicating that pictograms are good for, among other things, increasing visual attention and recall [53,54].

In our streamlining efforts, we had to be careful about oversimplifying the decision aid's content. Although users might find it easier for someone else to make the decision for them, studies show that informing them of the benefits and harms supports better outcomes [55]. Therefore, even if probabilities are hard to understand for most users, they are the very core of the decision aid, and we tried several approaches to make them clearer. The scientific literature reviews several ways of presenting risk probabilities (benefits and harms). Most agree that percentages and natural frequencies using a uniform denominator, as was done in the final version of the web-based decision aid, are the most effective way to present risks [56,57]. It has also been found that when visual aids such as icon arrays and bar graphs are added, comprehension—and retention—of the numbers increases significantly [57,58]. Nevertheless, other findings suggest that people with low graph literacy understand risks presented in numbers better, whereas people with high graphical literacy skills better understand those presented in graph form more readily [59]. It would be interesting to test a web-based decision aid that offers a graphical display of the risks in a layer for people seeking to learn more. Having said that, when an analysis was made of the participants' perceptions of being informed and clear about their personal values before and after using the web-based decision aid, it showed that their comfort with decision making generally improved with the use of the tool.

Intuitive and User Friendly

Finally, we conclude from our findings that users require a seamless experience when navigating the web-based decision aid to avoid any distraction and frustration with the decision-making process. The SUS scores that we measured throughout the rounds suggest that we achieved *good* usability at best, according to Sauro's reported average score of 68 across studies [44], despite several rounds of user testing. Some comments also suggest that the SUS items were not clearly understood, which might warrant validation in this population. These findings converge with those of Malinowsky et al [60], who found that older adults with MCI experience more difficulty using technology than those without any cognitive problems. Although further testing might help us reach greater usability, this emphasizes the importance of testing tools with the intended demographic during the design phase, to avoid creating decision aids that hinder the decision-making process instead of helping it. Our results show improved usability when the interface displays clear components that require little interaction (eg, checkboxes instead of drop-down menus) when page jumping is minimized (eg, pop-ups and information tips) and when all page content fits on a single screen, thereby eliminating the need to scroll down. This last observation converges with the results of another study on the implementation of a mobile health app, showing the value of content that is *at one's fingertips* [45]. This is also consistent with the guidelines of the United States Department of Health and Human Services [61], which are based on Morville and Rosenfeld's book [62] and which state that the interface should stay simple, with clear instructions and a page structure that is clear in its purpose [61].

Limitations of the Study

One of the limitations of the study is that the web-based decision aid shown at each round to the participants was a mock-up. This approach allowed us to minimize costs, to avoid programming several test versions of the decision aid, and is a standard approach used in website design. However, the mock-up functionality was limited and caused some unintended glitches. It is difficult to estimate the extent to which this affected the participants' understanding of the web-based decision aid in general; however, the mock-up clearly impaired understanding of the values clarification exercise. In the mock-up, we had to predetermine values clarification exercise priorities to avoid unnecessary design work at an early stage. Knowing that older adults with MCI present deficits in their executive functions and working memory [63], it is possible that the task of planning for a fictitious case was even more difficult for them, perhaps even impossible for some. Many of the participants with MCI were confused or even frustrated at not being able to put their real priorities in the values clarification exercise. This type of reaction was also found in another study to build a web-based shared decision-making tool with older adults living with neurocognitive disorders [64]. We recommend using a fully functional values clarification exercise when testing prototypes for this type of participant.

This work is also limited by the small sample sizes of HCPs and older adults with MCI. The vulnerability of the population made it hard to recruit subjects with the help of HCPs, as they wanted to protect them from the hassles of a research project. Consequently, we had to rely on alternative strategies to recruit older adults with MCI. We could not, therefore, diversify the participants who helped at each evaluation round, for example, for levels of MCI, educational background, and technology skills. Further research projects could lead to qualitative interviews with a smaller number of participants while simultaneously asking a larger group to test the web-based

decision aid. This would allow for the analysis of a larger amount of quantitative data while retaining the richness of the qualitative data gathered during the interviews.

Potential limitations of the quantitative data collected should also be clarified. The DCS that we used to help us tailor the decision aid measures only feelings about the decision, not the degree to which the values presented coincided with the individuals' actual values. In addition, given that participants were face to face with their interviewers, some may have answered in a way they felt was expected of them (social desirability bias).

Conclusions

Overall, we can say that we have successfully identified and applied features that best support values clarification and understanding of evidence in older adults with MCI and have adjusted the level of detail of a web-based decision aid for individuals with MCI, although there is still room for improvement. For example, we designed this interactive decision aid for tablets and computers, whereas internet access is increasingly used via mobile devices.

In conclusion, this study resulted in a promising web-based decision aid. Further ongoing research will allow its implementation to be tested in clinical settings. This decision aid will be useful in supporting health workers who regularly interact with older adults, such as nurses, physicians, geriatricians, social workers, psychologists, community pharmacists, or volunteers from community organizations. This decision aid will allow HCPs to meaningfully engage older adults in the decision-making process, so that patient values are prioritized. It will also be useful in informing older adults with MCI of the options available to improve their cognitive abilities, allowing them to feel a greater sense of control over their condition.

Our findings support three recommendations to create web-based decision aids for older adults with MCI:

1. Although we recognize that some evidence supports understanding options before completing a values clarification exercise [5,51], our research suggests that the website should guide users in completing a values clarification exercise before presenting any options, allowing for tailoring of the presentation of options to the users' priorities, while also allowing flexibility for those users who prefer jumping from one section to the other in no specific order.
2. Content should be simple, but not simplistic, notably by using information layering, lay language, and pictograms to represent each option.
3. The interface should be intuitive and user friendly, for example, by preferring check marks to drop-down menus, using pop-up windows and information tips, and by limiting the height of the window to avoid having to scroll down.

We also recommend that decision-aid developers consult users early on in the design process and combine observations of user interactions with the decision aid with a teach-back approach, to collect rich information from the get-go, and to ensure that they design an understandable and usable web-based decision aid for this more vulnerable population.

Acknowledgments

The authors would like to thank the expert panel comprising Michèle Morin, Émilie Fortier-Brochu, Edeltraut Kroger, Marie-Josée Sirois, Gabriel Bilodeau, Julie Fortin, Michel Cauchon, and Philippe Voyer who helped review the content of the decision aid before user testing. The authors would also like to thank Émilie Beaulieu and Philippe Fekete of the Toumoro agency for their support in creating the various versions of the mock-up of the web-based decision aid. They would also like to acknowledge the invaluable support of Sabrina Crevier, Abdelhamid Kadhi, Alexandre Drolet, and Jean-François Duguay from the Information Technology Branch of the Faculty of Medicine at Laval University, for helping design the mock-up for later integration into the university's website. The authors would further like to acknowledge the help from several clinician experts, notably Émilie Fortier-Brochu, Yves Turgeon, and Gabriel Bilodeau, who provided feedback on the mock-ups. Finally, they would like to acknowledge the contributions of David August in data collection and analysis, Andréanne Simard for helping with participant recruitment, and Katherine Hastings for writing assistance. This research is funded by the Canadian Frailty Network (Technology Evaluation in the Elderly Network), which is supported by the Government of Canada through the Networks of Centres of Excellence program. This study was also supported by an unrestricted grant to AG by the Research Chair in Aging at Laval University. LB is funded by a fellowship from the Canadian Institutes of Health Research. AG and HW are funded by Research Scholar Junior 2 Career Development Awards by the Fonds de Recherche du Québec—Santé.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Introduction page of the web-based decision aid, as tested in prototype version 1. Warning: this is a preliminary version, and the data are not final.
[PNG File , 120 KB - jmir_v22i8e17406_app1.png]

Multimedia Appendix 2

Introduction page of the web-based decision aid, as seen in the final version.
[PNG File , 83 KB - jmir_v22i8e17406_app2.png]

Multimedia Appendix 3
Definition appearing as an info-tip when the user hovers the cursor over a word, in the final version of the web-based decision aid.
[PNG File , 88 KB - jmir_v22i8e17406_app3.png]

Multimedia Appendix 4
Decision page in the initial prototype.
[PNG File , 66 KB - jmir_v22i8e17406_app4.png]

Multimedia Appendix 5
Decision page in the final version.
[PNG File , 72 KB - jmir_v22i8e17406_app5.png]

References

1. Petersen RC. Clinical practice. Mild cognitive impairment. N Engl J Med 2011 Jun 9;364(23):2227-2234. [doi: 10.1056/NEJMcp0910237] [Medline: 21651394]
2. Petersen RC, Lopez O, Armstrong MJ, Getchius TS, Ganguli M, Gloss D, et al. Practice guideline update summary: mild cognitive impairment: report of the guideline development, dissemination, and implementation subcommittee of the American academy of neurology. Neurology 2018 Jan 16;90(3):126-135 [FREE Full text] [doi: 10.1212/WNL.0000000000004826] [Medline: 29282327]
3. Nagamatsu LS, Handy TC, Hsu CL, Voss M, Liu-Ambrose T. Resistance training promotes cognitive and functional brain plasticity in seniors with probable mild cognitive impairment. Arch Intern Med 2012 Apr 23;172(8):666-668 [FREE Full text] [doi: 10.1001/archinternmed.2012.379] [Medline: 22529236]
4. Kaduszkiewicz H, Eisele M, Wiese B, Prokein J, Luppa M, Luck T, Study on Aging, Cognition,Dementia in Primary Care Patients (AgeCoDe) Study Group. Prognosis of mild cognitive impairment in general practice: results of the German AgeCoDe study. Ann Fam Med 2014;12(2):158-165 [FREE Full text] [doi: 10.1370/afm.1596] [Medline: 24615312]
5. Joseph-Williams N, Newcombe R, Politi M, Durand M, Sivell S, Stacey D, et al. Toward minimum standards for certifying patient decision AIDS: a modified delphi consensus process. Med Decis Making 2014 Aug;34(6):699-710. [doi: 10.1177/0272989X13501721] [Medline: 23963501]
6. O'Connor AM, Llewellyn-Thomas HA, Flood AB. Modifying unwarranted variations in health care: shared decision making using patient decision aids. Health Aff (Millwood) 2004;Suppl Variation:VAR63-VAR72. [doi: 10.1377/hlthaff.var.63] [Medline: 15471770]
7. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions (Review). Cochrane Database Syst Rev 2014;1 [FREE Full text] [Medline: 21975733]
8. Giguere A, Légaré F, Grad R, Pluye P, Haynes RB, Cauchon M, et al. Decision boxes for clinicians to support evidence-based practice and shared decision making: the user experience. Implement Sci 2012 Aug 3;7:72 [FREE Full text] [doi: 10.1186/1748-5908-7-72] [Medline: 22862935]
9. Giguere AM, Labrecque M, Haynes RB, Grad R, Pluye P, Légaré F, et al. Evidence summaries (decision boxes) to prepare clinicians for shared decision-making with patients: a mixed methods implementation study. Implement Sci 2014 Oct 5;9:144 [FREE Full text] [doi: 10.1186/s13012-014-0144-6] [Medline: 25280742]
10. Bilodeau G, Witteman H, Légaré F, Lafontaine-Bruneau J, Voyer P, Kröger E, et al. Reducing complexity of patient decision aids for community-based older adults with dementia and their caregivers: multiple case study of Decision Boxes. BMJ Open 2019 May 9;9(5):e027727 [FREE Full text] [doi: 10.1136/bmjopen-2018-027727] [Medline: 31072861]
11. van Weert JC, van Munster BC, Sanders R, Spijker R, Hooft L, Jansen J. Decision aids to help older people make health decisions: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2016 Apr 21;16:45 [FREE Full text] [doi: 10.1186/s12911-016-0281-8] [Medline: 27098100]
12. Peters E, Hess TM, Västfjäll D, Auman C. Adult age differences in dual information processes: implications for the role of affective and deliberative processes in older adults' decision making. Perspect Psychol Sci 2007 Mar;2(1):1-23. [doi: 10.1111/j.1745-6916.2007.00025.x] [Medline: 26151915]
13. Reed AE, Mikels JA, Simon KI. Older adults prefer less choice than young adults. Psychol Aging 2008 Sep;23(3):671-675 [FREE Full text] [doi: 10.1037/a0012772] [Medline: 18808256]
14. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. Psychol Bull 2009 Nov;135(6):943-973 [FREE Full text] [doi: 10.1037/a0017327] [Medline: 19883143]

15. Gazmararian JA, Baker DW, Williams MV, Parker RM, Scott TL, Green DC, et al. Health literacy among medicare enrollees in a managed care organization. J Am Med Assoc 1999 Feb 10;281(6):545-551. [doi: 10.1001/jama.281.6.545] [Medline: 10022111]
16. Clayman ML, Roter D, Wissow LS, Bandeen-Roche K. Autonomy-related behaviors of patient companions and their effect on decision-making activity in geriatric primary care visits. Soc Sci Med 2005 Apr;60(7):1583-1591. [doi: 10.1016/j.socscimed.2004.08.004] [Medline: 15652689]
17. Laidsaar-Powell RC, Butow PN, Bu S, Charles C, Gafni A, Lam WW, et al. Physician-patient-companion communication and decision-making: a systematic review of triadic medical consultations. Patient Educ Couns 2013 Apr;91(1):3-13. [doi: 10.1016/j.pec.2012.11.007] [Medline: 23332193]
18. Frozard J. Vision and hearing in aging. In: Birren JE, Schaie KW, editors. Handbook of the Psychology of Aging. San Diego, CA: Academic Press; 1990:150-170.
19. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10603.x] [Medline: 11929503]
20. Gordon NP, Crouch E. Digital information technology use and patient preferences for internet-based health education modalities: cross-sectional survey study of middle-aged and older adults with chronic health conditions. JMIR Aging 2019 Apr 4;2(1):e12243 [FREE Full text] [doi: 10.2196/12243] [Medline: 31518291]
21. Hoffman AS, Volk RJ, Saarimaki A, Stirling C, Li LC, Härter M, et al. Delivering patient decision aids on the internet: definitions, theories, current evidence, and emerging research areas. BMC Med Inform Decis Mak 2013;13(Suppl 2):S13 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S13] [Medline: 24625064]
22. Syrowatka A, Krömker D, Meguerditchian AN, Tamblyn R. Features of computer-based decision aids: systematic review, thematic synthesis, and meta-analyses. J Med Internet Res 2016 Jan 26;18(1):e20 [FREE Full text] [doi: 10.2196/jmir.4982] [Medline: 26813512]
23. Sanders EB. From user-centered to participatory design approaches. In: Frascara J, editor. Design and the Social Sciences: Making Connections. Abingdon, UK: Taylor & Francis Books; 2002:1-8.
24. White M, Garbez R, Carroll M, Brinker E, Howie-Esquivel J. Is 'teach-back' associated with knowledge retention and hospital readmission in hospitalized heart failure patients? J Cardiovasc Nurs 2013;28(2):137-146. [doi: 10.1097/JCN.0b013e31824987bd] [Medline: 22580624]
25. Graham ID, Logan J, Bennett CL, Presseau J, O'Connor AM, Mitchell SL, et al. Physicians' intentions and use of three patient decision aids. BMC Med Inform Decis Mak 2007 Jul 6;7:20 [FREE Full text] [doi: 10.1186/1472-6947-7-20] [Medline: 17617908]
26. Jick TD. Mixing qualitative and quantitative methods: triangulation in action. Adm Sci Q 1979 Dec;24(4):602-671. [doi: 10.2307/2392366]
27. Lawani MA, Valéra B, Fortier-Brochu E, Légaré F, Carmichael P, Côté L, et al. Five shared decision-making tools in 5 months: use of rapid reviews to develop decision boxes for seniors living with dementia and their caregivers. Syst Rev 2017 Mar 15;6(1):56 [FREE Full text] [doi: 10.1186/s13643-017-0446-2] [Medline: 28298241]
28. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of ehealth technologies. J Med Internet Res 2011 Dec 5;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
29. van Gemert-Pijnen L, Span M. CeHRes roadmap to improve dementia care. In: van Hoof J, Demiris G, Wouters EJ, editors. Handbook of Smart Homes, Health Care and Well-Being. New York, USA: Springer; 2016:133-146.
30. Hoffman AS, Llewellyn-Thomas HA, Tosteson AN, O'Connor AM, Volk RJ, Tomek IM, et al. Launching a virtual decision lab: development and field-testing of a web-based patient decision support research platform. BMC Med Inform Decis Mak 2014 Dec 12;14:112 [FREE Full text] [doi: 10.1186/s12911-014-0112-8] [Medline: 25495552]
31. Skjøth MM, Hansen HP, Draborg E, Pedersen CD, Lamont RF, Jørgensen JS. Informed choice for participation in down syndrome screening: development and content of a web-based decision aid. JMIR Res Protoc 2015 Sep 21;4(3):e113 [FREE Full text] [doi: 10.2196/resprot.4291] [Medline: 26392319]
32. Ng CJ, Mathers N, Bradley A, Colwell B. A 'combined framework' approach to developing a patient decision aid: the PANDAs model. BMC Health Serv Res 2014 Oct 24;14:503 [FREE Full text] [doi: 10.1186/s12913-014-0503-7] [Medline: 25341370]
33. Hoops S, Nazem S, Siderowf AD, Duda JE, Xie SX, Stern MB, et al. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. Neurology 2009 Nov 24;73(21):1738-1745 [FREE Full text] [doi: 10.1212/WNL.0b013e3181c34b47] [Medline: 19933974]
34. Larouche E, Hudon C, Goulet S. Mindfulness mechanisms and psychological effects for aMCI patients: a comparison with psychoeducation. Complement Ther Clin Pract 2019 Feb;34:93-104. [doi: 10.1016/j.ctcp.2018.11.008] [Medline: 30712752]
35. Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004 Sep;256(3):183-194 [FREE Full text] [doi: 10.1111/j.1365-2796.2004.01388.x] [Medline: 15324362]
36. O'Connor A. User Manual - Decisional Conflict Scale. Patient Decision Aids - Ottawa Hospital Research. 2010. URL: https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decisional_Conflict.pdf [accessed 2020-05-07]

37. Jaspers MW, Steen T, van den Bos C, Geenen M. The think aloud method: a guide to user interface design. Int J Med Inform 2004 Nov;73(11-12):781-795. [doi: 10.1016/j.ijmedinf.2004.08.003] [Medline: 15491929]
38. Protheroe J, Nutbeam D, Rowlands G. Health literacy: a necessity for increasing participation in health care. Br J Gen Pract 2009 Oct;59(567):721-723 [FREE Full text] [doi: 10.3399/bjgp09X472584] [Medline: 19843420]
39. Low Health Literacy and Chronic Disease Prevention and Control's Perspectives From the Health and Public Sectors. Canadian Public Health Association. 2006. URL: http://www.cpha.ca/uploads/portals/h-l/kl_summary_e.pdf [accessed 2020-04-26]
40. Price KA. Teach-Back Effect on Self-Reported Understanding of Health Management After Discharge. Walden University. 2014. URL: http://search.proquest.com/openview/7ec355a097a68c9dee09966fd3d7f37f/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2020-05-07]
41. Brega A, Barnard J, Mabachi N, Weiss B, DeWalt D, Brach C, et al. AHRQ Health Literacy Universal Precautions Toolkit. Agency for Healthcare Research Quality. 2015. URL: https://www.ahrq.gov/sites/default/files/publications/files/healthlittoolkit2_3.pdf [accessed 2020-04-29]
42. O'Connor A, Cranney A. Users Manual: Accessibility. Patient Decision Aids - Ottawa Hospital Research Institute. 2002. URL: http://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Acceptability.pdf [accessed 2020-04-29]
43. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. London, UK: Taylor and Francis; 1996:189-194.
44. Sauro J. Measuring Usability With the System Usability Scale (SUS). MeasuringU. 2011. URL: https://measuringu.com/sus/ [accessed 2019-05-15]
45. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User preferences and design recommendations for an mhealth app to promote cystic fibrosis self-management. JMIR Mhealth Uhealth 2014 Oct 24;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
46. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncol Nurs Forum 2014 Sep;41(5):545-547. [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]
47. Moran-Ellis J, Alexander VD, Cronin A, Dickinson M, Fielding J, Sleney J, et al. Triangulation and integration: processes, claims and implications. Qual Res 2016 Aug 15;6(1):45-59. [doi: 10.1177/1468794106058870]
48. Mild Problems With Thinking or Memory. Decision Box. 2018. URL: https://www.decisionbox.ulaval.ca/en/box-details/?tx_tmboites_tmboitesmain%5Bclear%5D=1&tx_tmboites_tmboitesmain%5Bboite%5D=2 [accessed 2019-07-23]
49. Bansback N, Li LC, Lynd L, Bryan S. Exploiting order effects to improve the quality of decisions. Patient Educ Couns 2014 Aug;96(2):197-203 [FREE Full text] [doi: 10.1016/j.pec.2014.05.021] [Medline: 24961445]
50. Castilla D, Garcia-Palacios A, Miralles I, Breton-Lopez J, Parra E, Rodriguez-Berges S, et al. Effect of web navigation style in elderly users. Comput Human Behav 2016 Feb;55:909-920. [doi: 10.1016/j.chb.2015.10.034]
51. Reyna VF. A theory of medical decision making and health: fuzzy trace theory. Med Decis Making 2008;28(6):850-865 [FREE Full text] [doi: 10.1177/0272989X08327066] [Medline: 19015287]
52. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):270-279 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.008] [Medline: 21514249]
53. Barros IM, Alcântara TS, Mesquita AR, Santos AC, Paixão FP, Lyra DP. The use of pictograms in the health care: a literature review. Res Social Adm Pharm 2014;10(5):704-719. [doi: 10.1016/j.sapharm.2013.11.002] [Medline: 24332470]
54. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns 2006 May;61(2):173-190. [doi: 10.1016/j.pec.2005.05.004] [Medline: 16122896]
55. Durand M, Carpenter L, Dolan H, Bravo P, Mann M, Bunn F, et al. Do interventions designed to support shared decision-making reduce health inequalities? A systematic review and meta-analysis. PLoS One 2014;9(4):e94670 [FREE Full text] [doi: 10.1371/journal.pone.0094670] [Medline: 24736389]
56. McDowell M, Rebitschek FG, Gigerenzer G, Wegwarth O. A simple tool for communicating the benefits and harms of health interventions: a guide for creating a fact box. MDM Policy Pract 2016;1(1):2381468316665365 [FREE Full text] [doi: 10.1177/2381468316665365] [Medline: 30288405]
57. Garcia-Retamero R, Galesic M. Who profits from visual aids: overcoming challenges in people's understanding of risks [corrected]. Soc Sci Med 2010 Apr;70(7):1019-1025. [doi: 10.1016/j.socscimed.2009.11.031] [Medline: 20116159]
58. Kreuzmair C, Siegrist M, Keller C. Does iconicity in pictographs matter? The influence of iconicity and numeracy on information processing, decision making, and liking in an eye-tracking study. Risk Anal 2017 Mar;37(3):546-556. [doi: 10.1111/risa.12623] [Medline: 27089046]
59. Gaissmaier W, Wegwarth O, Skopec D, Müller AS, Broschinski S, Politi MC. Numbers can be worth a thousand pictures: individual differences in understanding graphical and numerical representations of health-related information. Health Psychol 2012 May;31(3):286-296. [doi: 10.1037/a0024850] [Medline: 21842998]

60. Malinowsky C, Nygård L, Kottorp A. Using a screening tool to evaluate potential use of e-health services for older people with and without cognitive impairment. Aging Ment Health 2014;18(3):340-345. [doi: 10.1080/13607863.2013.832731] [Medline: 24548108]
61. User Interface Design Basics. Usability. URL: https://www.usability.gov/what-and-why/user-interface-design.html [accessed 2019-05-16]
62. Morville P, Rosenfeld L. Information Architecture for the World Wide Web. Sebastopol, CA, USA: O'Reilly Media Inc; 2006.
63. Kirova A, Bays RB, Lagalwar S. Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease. Biomed Res Int 2015;2015:748212 [FREE Full text] [doi: 10.1155/2015/748212] [Medline: 26550575]
64. Span M, Hettinga M, Groen-van de Ven L, Jukema J, Janssen R, Vernooij-Dassen M, et al. Involving people with dementia in developing an interactive web tool for shared decision-making: experiences with a participatory design approach. Disabil Rehabil 2018 Jun;40(12):1410-1420. [doi: 10.1080/09638288.2017.1298162] [Medline: 28286969]

Abbreviations

DCS: decisional conflict scale
GRADE: grading of recommendations assessment, development, and evaluation
HCP: health care professionals
MCI: mild cognitive impairment
SUS: system usability scale

Edited by G Eysenbach; submitted 29.12.19; peer-reviewed by A Syrowatka, B Jugdutt; comments to author 14.02.20; revised version received 07.04.20; accepted 15.04.20; published 19.08.20.

Please cite as:
Bogza LM, Patry-Lebeau C, Farmanova E, Witteman HO, Elliott J, Stolee P, Hudon C, Giguere AMC
User-Centered Design and Evaluation of a Web-Based Decision Aid for Older Adults Living With Mild Cognitive Impairment and Their Health Care Providers: Mixed Methods Study
J Med Internet Res 2020;22(8):e17406
URL: https://www.jmir.org/2020/8/e17406
doi:10.2196/17406
PMID:32442151

Original Paper

Advance Care Planning Among Users of a Patient Portal During the COVID-19 Pandemic: Retrospective Observational Study

Jennifer D Portz[1], PhD; Adreanne Brungardt[2], MM, MT-BC; Prajakta Shanbhag[2], MPH; Elizabeth W Staton[3], MSTC; Seuli Bose-Brill[4], MD; Chen-Tan Lin[1], MD; Jean S Kutner[1], MD, MSPH; Hillary D Lum[2], MD, PhD

[1]Division of General Internal Medicine, Department of Medicine, University of Colorado School of Medicine, Aurora, CO, United States

[2]Division of Geriatric Medicine, Department of Medicine, University of Colorado School of Medicine, Aurora, CO, United States

[3]Department of Family Medicine, University of Colorado School of Medicine, Aurora, CO, United States

[4]Department of Medicine, The Ohio State University College of Medicine, Columbus, OH, United States

Corresponding Author:
Jennifer D Portz, PhD
Division of General Internal Medicine
Department of Medicine
University of Colorado School of Medicine
12631 East 17th Avenue
Aurora, CO, 80045
United States
Phone: 1 303 724 4438
Email: jennifer.portz@cuanschutz.edu

Abstract

Background: Advance care planning is the process of discussing health care treatment preferences based on patients' personal values, and it often involves the completion of advance directives. In the first months of 2020, a novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), began circulating widely in the American state of Colorado, leading to widespread diagnosis of coronavirus disease (COVID-19), hospitalizations, and deaths. In this context, the importance of technology-based, non–face-to-face methods to conduct advance care planning via patient portals has increased.

Objective: The aim of this study was to determine the rates of use of a web-based advance care planning tool through a health system–based electronic patient portal both before and in the early months of the COVID-19 pandemic.

Methods: In 2017, we implemented web-based tools through the patient portal of UCHealth's electronic health record (EHR) for patients to learn about advance care planning and complete an electronically signed medical durable power of attorney (MDPOA) to legally appoint a medical decision maker. Patients accessing the portal can complete and submit a legally valid MDPOA, which becomes part of their medical record. We collected data on the patients' date of MDPOA completion, use of advance care planning messaging, age, sex, and geographic location during the early phase of the COVID-19 pandemic (December 29, 2019, to May 30, 2020).

Results: Over a 5-month period that includes the early phase of the COVID-19 pandemic in Colorado, total monthly use of the advance care planning portal tool increased from 418 users in January to 1037 users in April and then decreased slightly to 815 users in May. The number of MDPOA forms submitted per week increased 2.4-fold after the stay-at-home order was issued in Colorado on March 26, 2020 (P<.001). The mean age of the advance care planning portal users was 47.7 years (SD 16.1), and 2206/3292 (67.0%) were female. Women were more likely than men to complete an MDPOA, particularly in younger age groups (P<.001). The primary use of the advance care planning portal tools was the completion of an MDPOA (3138/3292, 95.3%), compared to sending an electronic message (148/3292, 4.5%). Over 50% of patients who completed an MDPOA did not have a prior agent in the EHR.

Conclusions: Use of a web-based patient portal to complete an MDPOA increased substantially during the first months of the COVID-19 pandemic in Colorado. There was an increase in advance care planning that corresponded with state government shelter-in-place orders as well as public health reports of increased numbers of COVID-19 cases and deaths. Patient portals are an important tool for providing advance care planning resources and documenting medical decision makers during the pandemic to ensure that medical treatment aligns with patient goals and values.

(J Med Internet Res 2020;22(8):e21385) doi:10.2196/21385

KEYWORDS

advance care planning; electronic health records; pandemic; COVID-19; advance directives; patient portal; planning; web-based tool; health system

Introduction

On March 11, 2020, the World Health Organization characterized coronavirus disease (COVID-19) as a pandemic. At that time, approximately 118,000 cases had been reported in 114 countries, and 4291 people had died worldwide due to COVID-19. As of June 2020, the fatality rates for COVID-19 have been estimated to be between 0.6% and 5%; these rates are highest among older adults and people with chronic conditions [1,2]. In the United States, the Centers for Disease Control and Prevention reports that 14% of patients with COVID-19 are hospitalized and 2% are admitted to an intensive care unit (ICU) [2]. In these cases, invasive procedures such as mechanical ventilation and extracorporeal membrane oxygenation may be used to treat seriously ill patients with COVID-19. As such, there is a need to discuss goals of care and desired treatments with patients, ideally before they become critically ill.

Advance care planning is the process of discussing medical treatment preferences based on personal values, and it often involves the completion of advance directives [3,4]. In the United States, advance directives are state-specific documents that include medical durable powers of attorney (MDPOAs) and living wills [5]. Advance care planning is associated with increased advance directive documentation, completion of medical orders for life-sustaining treatment preferences, and positive health outcomes, including reduced end-of-life hospitalizations, enhanced patient-provider communication, prevention of unwanted treatment, and improved family experience [6,7].

COVID-19 has stimulated advance care planning in emergency departments and ICUs, indicating a need to provide advance care planning interventions and resources prior to hospital admission [8]. Telehealth is the use of electronic information or telecommunication technologies to promote health and health services [9]. With increased social distancing and patient concerns about in-person clinic visits [8,10,11], telehealth use has rapidly increased. Centers for Medicare & Medicaid Services reimbursement has rapidly expanded to cover telehealth services such as nonemergent clinical visits, screening for COVID-19, and advance care planning counseling [12]. Telehealth has potential to provide health care services and applicable advance care planning tools during the pandemic and beyond [13,14]. Telehealth-based advance care planning programs are feasible to deliver and effective in improving advance care planning knowledge, communication, and advance directive documentation [15].

Patient portals provide patients with secure access to personal health information contained in their electronic health record [16]. In addition to the ability to view health information, patient portals support patient-initiated entry of health information and electronic interactions between patients and providers; this promotes engagement in care, including advance care planning [17-19]. Portals thus offer a particularly accessible platform for patients and care partners to learn about COVID-19 treatment options and communicate goals of care with their providers. Over the last 5 years, our research team has partnered with health care system information technology experts, patients, caregivers, and providers to develop, test, and implement patient portal advance care planning strategies [20-24]. Specifically, the UCHealth patient portal offers patients the ability to learn more about advance care planning through web-based resources and provides the ability to create an electronically signed MDPOA form to choose a medical decision maker. The objective of this study is to examine advance care planning patient portal usage during the early stages of the COVID-19 pandemic.

Methods

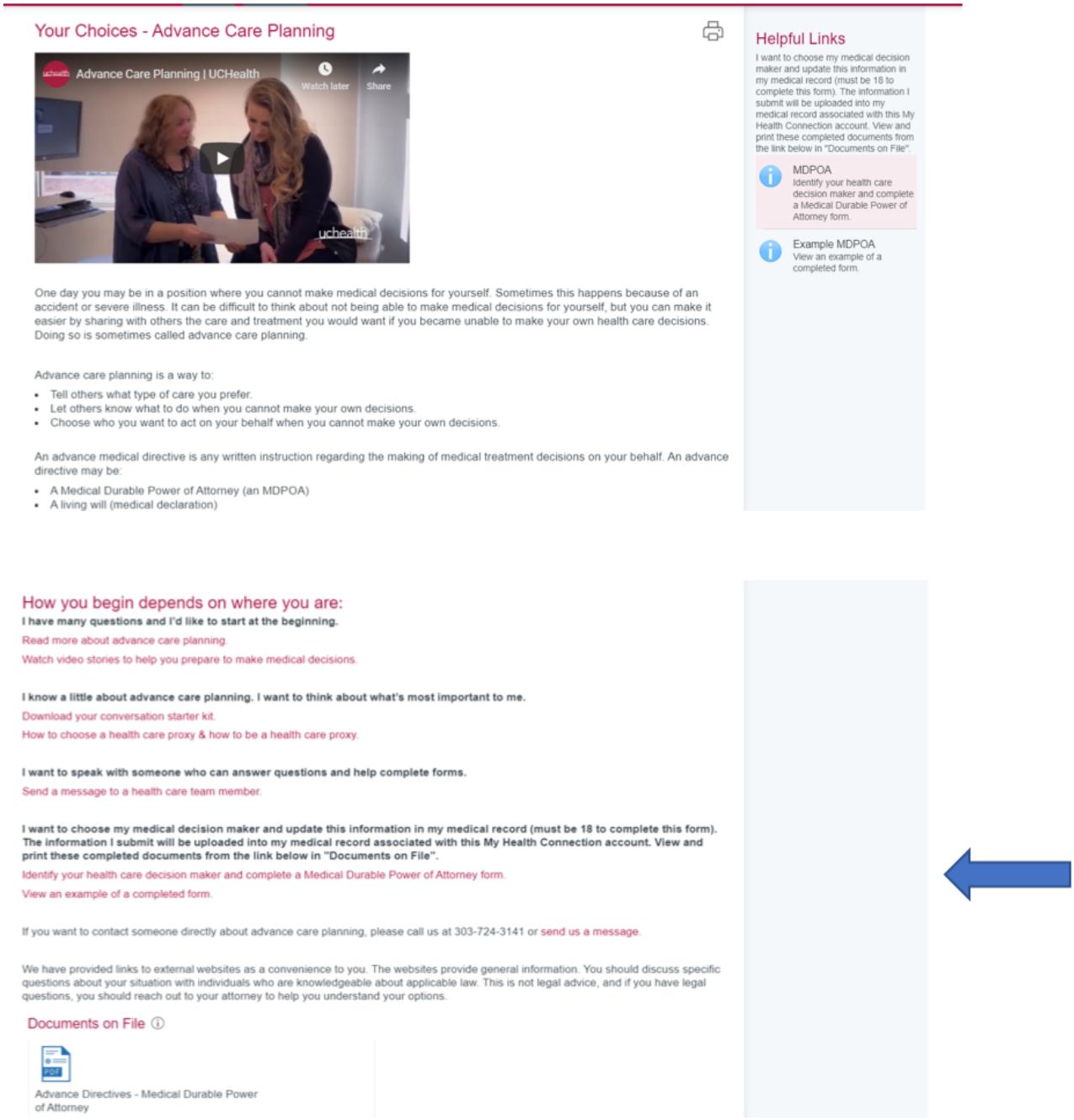
This retrospective observational study examined advance care planning by patients using web-based advance care planning tools available through the UCHealth patient portal that enable the completion of legally valid MDPOA forms. UCHealth is a regional health care system serving Colorado (including metropolitan Denver, Colorado Springs, and Northern Colorado), southern Wyoming, and western Nebraska. In 2019, 1.9 million patients received care in 12 UCHealth hospitals and 800 outpatient clinics. My Health Connection is the UCHealth patient portal integrated within the Epic electronic health record (Epic Systems, version 2017); as of May 2020, approximately 735,000 patients had a My Health Connection account. The Colorado Multiple Institutional Review Board approved this project as an evaluation of a clinical initiative.

Patient Portal–Based Advance Care Planning Tools

We previously developed and tested tools to conduct advance care planning using the My Health Connection patient portal [22,23]. Briefly, since July 2017, the advance care planning portal tools has provided patients with access to evidence-based resources, the ability to send web-based messages with questions about advance care planning, and the ability to complete a legally valid state-specific electronic MDPOA form to choose a medical decision maker. Colorado law allows patients to create an electronic MDPOA with a valid patient signature (including electronic signatures) and does not require witnesses or notary signatures. To complete the MDPOA form (Figure 1), a patient-initiated questionnaire provides information about appointing a medical decision maker, shows the exact language of the legal MDPOA form, automatically provides the name of any previously charted medical decision maker from the EHR (“orally appointed” or from a pre-existing MDPOA), and then allows the patient to name a decision maker and add up to two alternative decision makers. In this process, a printable MDPOA form is created with an electronic signature and date/time stamp.

The decision maker information is displayed in a specific area of the EHR that is accessed by clinical teams. The advance care planning support team is notified when an electronic MDPOA form is submitted, briefly reviews the patient's problem list and relevant clinical documentation for possible decision-making incapacity, then sends a message to the patient to confirm receipt of the MDPOA form. For out-of-state patients, this message notes that the MDPOA is valid for medical care received in Colorado. Patients can also use the portal message feature to contact the health system's centralized advance care planning support team for questions and follow-up. As additional background, in Colorado, an orally appointed decision maker is a surrogate decision maker verbally chosen by an individual; however, that surrogate decision maker does not have as much legal authority as an MDPOA.

Figure 1. Screenshot of the advance care planning tools in the My Health Connection portal.



Data Collection

Data were collected from December 29, 2019, to May 30, 2020, an approximately 5-month period that includes the 2 months prior to community spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in Colorado. These specific dates were chosen to enable comparison of 7-day weekly periods. The first COVID-19 diagnosis in Colorado was announced on March 5, 2020. Thus, January and February were effectively "pre–COVID-19" months in Colorado. Colorado implemented a statewide stay-at-home order on March 26, 2020. We collected data through chart review for UCHealth patients

who interacted with the My Health Connection patient portal advance care planning tools, specifically by completing an electronic MDPOA form or by sending an electronic message to the advance care planning support team. Demographic information included age, sex, and geographic location (ie, metro Denver, northern Colorado, southern Colorado, or out-of-state address). To explore how patients who submitted an electronic MDPOA were choosing their health care agents, we categorized patients into the following five groups: 1) first-time designation of an MDPOA, no prior agent (orally appointed or MDPOA) in the EHR; 2) prior orally appointed agent in the EHR who is retained in the electronic MDPOA; 3) prior MDPOA on file and the same agent is retained in the electronic MDPOA; 4) prior MDPOA on file and the electronic MDPOA names a new agent; and 5) prior orally appointed agent in the EHR and the electronic MDPOA names a new agent. A prior MDPOA could include any MDPOA form in the patient's EHR (either a paper document that was scanned into the patient's record or a previously submitted electronic MDPOA). We also identified invalid MDPOA submissions based on advance care planning support team review and discussion with the patient or their clinical team, as previously described [23]. Invalid MDPOA submissions often included MDPOA forms submitted by someone other than the patient when the patient did not have decision-making capacity to choose a health care agent.

Data Analysis

We used descriptive analyses to describe the characteristics of the patient portal users and expressed them as frequencies with percentages. Age categories were chosen to align with state-based surveys of community-based advance care planning rates [25]. We used chi-square tests to test gender differences across the age group distribution. Time series analyses were conducted on all data. To explore the changes in the use of the advance care planning tools in the patient portal, we used 9 weeks of data before and after March 26, 2020 (the start date of Colorado's stay-at-home order). All tests for statistical significance were two-tailed, and a P value <.05 was considered statistically significant. All statistical analyses were performed using SAS version 9.4 (SAS Institute).

Results

Over a 5-month timeframe that includes the early phase of the COVID-19 pandemic in Colorado, the number of user clicks on the UCHealth advance care planning patient portal page increased from 3511 in January to 6819 in April and 10,077 in May. The total number of monthly advance care planning portal tool users increased from 418 in January to 1037 in April, then slightly decreased to 815 users in May (Figure 2). Week-to-week MDPOA completions varied within each month; however, a positive trend line and greater use were observed in April and May (Figure 3). In an interrupted time-series analysis, the number of MDPOA forms submitted per week increased by 92 after the Colorado stay-at-home order was issued on March 26, 2020 (P<.001). The weekly rate was 2.4-fold higher in the 9 weeks after March 26 compared to the nine weeks prior to March 26. Figure 3 shows the number of submissions of MDPOA forms though the UCHealth patient portal in the cultural context of health orders from the State of Colorado. On March 26, 2020, the governor of Colorado issued a statewide stay-at-home policy [26]. On April 5, the government of Colorado approved and publicized Crisis Standards of Care documents as guidelines for how the medical community should allocate scarce resources such as ventilators and intensive care unit beds in the extreme case when patient needs exceed the resources available [27]. The Colorado Department of Public Health and Environment also updated the statewide numbers of cases and deaths of individuals with COVID-19 throughout this time period [28].

From December 29, 2019, to May 30, 2020, 3292 patients used the advance care planning portal tools. These patients were mostly female (2206/3292, 67.0%) with a mean age of 47.7 years (SD 16.1). The ages of patients who used the advance care planning portal tools were relatively evenly distributed between 25 years and ≥65 years (Table 1). The largest group of users was people aged 25-34 years (721/3292, 21.9%). The age distribution in our total population differed significantly between women and men (P<.001), where the proportion of women who used the advance care planning portal tools was greater than that of men for all age groups and significantly greater in the younger age groups (Table 2). Regionally, all three health care regions of UCHealth were represented, with 1412/3292 (42.9%) from the metro Denver region.

Figure 2. Monthly use of the UCHealth advance care planning portal during the coronavirus disease pandemic by type of portal use. ACP: advance care planning; eMDPOA: electronic medical durable power of attorney.

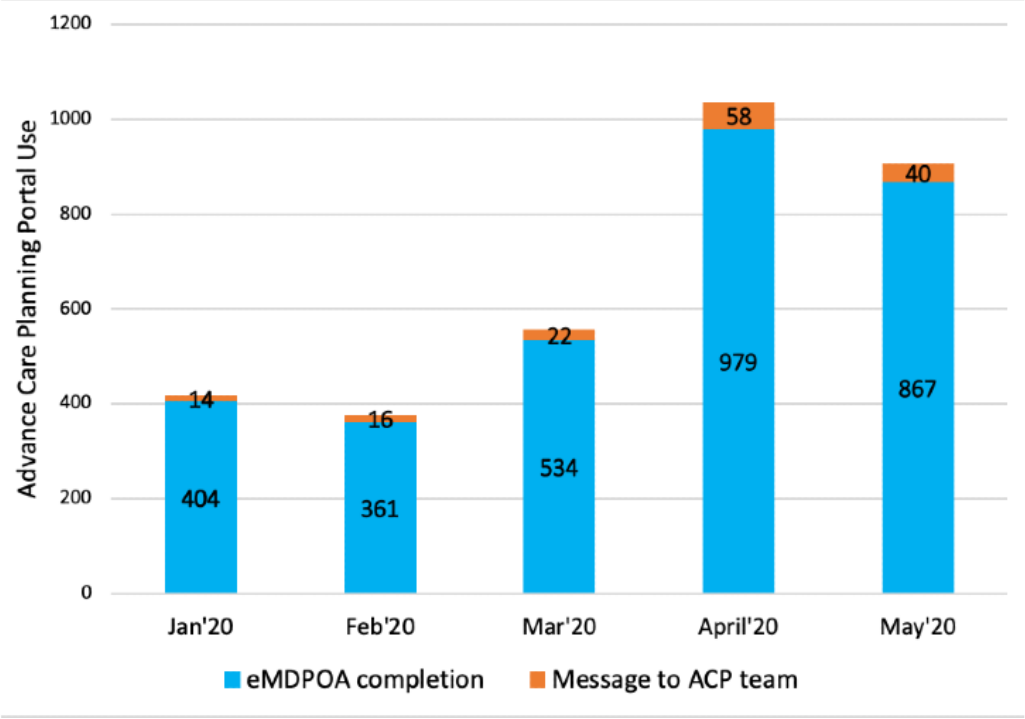


Figure 3. Weekly use of advance care planning portal for electronic MDPOA completion, statewide COVID-19 deaths, and contemporaneous events during the COVID-19 pandemic. COVID-19: coronavirus disease. MDPOA: medical durable power of attorney.

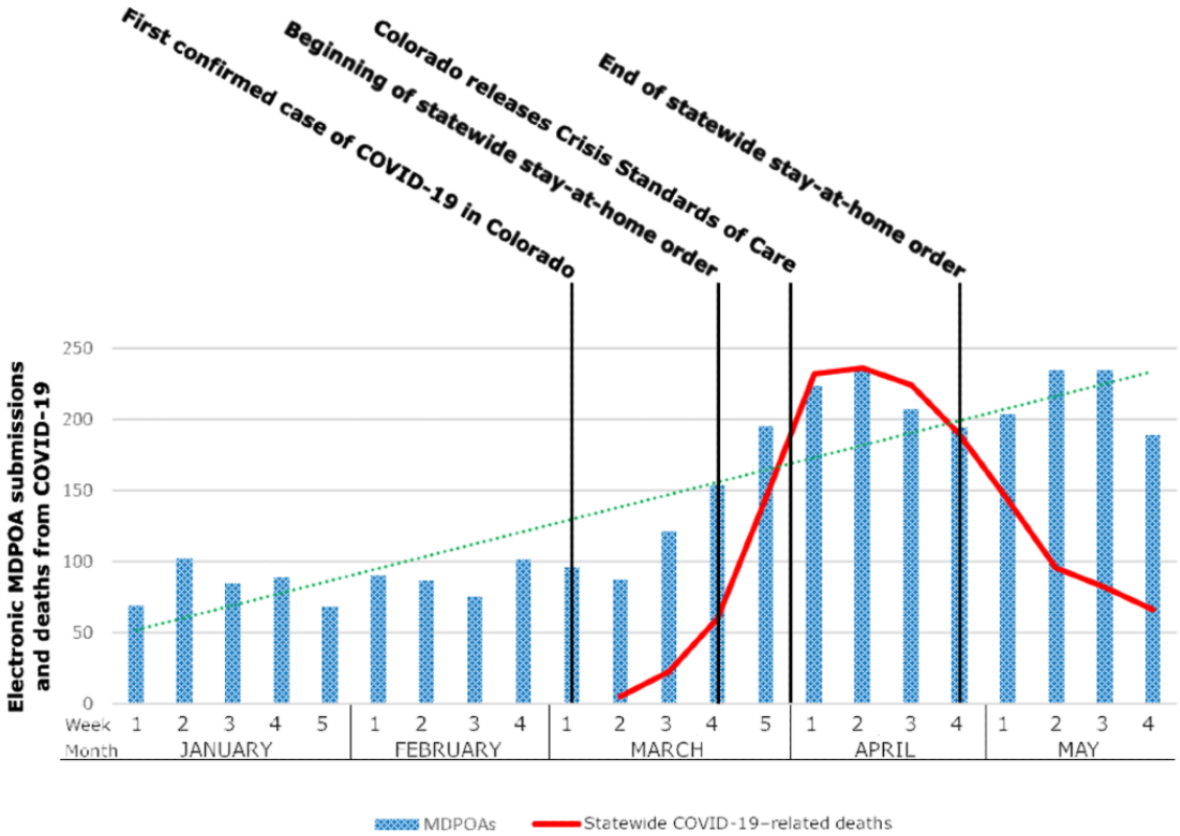


Table 1. Characteristics of patient portal users from December 29, 2019 to May 30, 2020 (N=3292).

Characteristic	n (%)
Age (years)	
18-24	148 (4.5)
25-34	721 (21.9)
35-44	684 (20.8)
45-54	578 (17.6)
55-64	543 (16.5)
≥65	618 (18.8)
Sex	
Female	2206 (67.0)
Male	1082 (33.2)
Nonbinary	4 (0.1)
Region	
Metro Denver	1412 (42.9)
North Colorado	1021 (31.0)
South Colorado	719 (21.8)
Out of state	140 (4.3)
Type of interaction	
eMDPOA[a] completion	3138 (95.3)
Message to advance care planning support team	148 (4.5)
Invalid MDPOA[b] submission	6 (0.2)
eMDPOA subgroups[c]	
No prior agent (oral or MDPOA) and eMDPOA selects a new agent	1665 (50.6)
Prior MDPOA form and eMDPOA does not change the agent	644 (19.6)
Prior MDPOA from and eMDPOA selects a new agent	78 (2.4)
Prior orally appointed agent and eMDPOA does not change the agent	686 (20.8)
Prior orally appointed agent and eMDPOA selects a new agent	65 (1.3)

[a]eMDPOA: electronic medical durable power of attorney.

[b]MDPOA: medical durable power of attorney.

[c]Percentages total 95.3% (proportion of total eMDPOAs).

Table 2. Use of the advance care planning patient portal tools by age (P<.001) and gender (male/female).

Age group (years)	Female, n (%)	Male, n (%)
18-24 (n=148)	124 (83.8)	24 (16.2)
25-34 (n=720)	569 (79.0)	151 (21.0)
35-44 (n=682)	465 (68.2)	217 (31.8)
45-54 (n=578)	381 (65.9)	197 (34.1)
55-64 (n=542)	334 (61.6)	208 (38.4)
≥65 years (n=618)	333 (53.9)	285 (46.1)

The primary use of the advance care planning portal tools was the completion of a MDPOA (Table 1). Of the 3292 patients who completed an MDPOA, over 50% (1665/3292, 50.6%) did not have a prior agent in the EHR; thus, these 1665 patients named a medical decision maker during their completion of the MDPOA, ensuring that this information would be available to their health care providers. An additional 686/3292 patients (20.8%) officially appointed a previously orally appointed

person to be their legal medical power of attorney by submitting the MDPOA form. The third largest group of patient users (644/3292, 19.6%) were patients who already had an MDPOA on file and did not change their primary medical power of attorney in the new MDPOA, although in some cases they changed an alternate decision maker or added specific instructions as part of the form. Six patients in the 5-month period submitted an invalid MDPOA by selecting themselves as the decision maker or lacked decision-making capacity as determined through quality assurance review and discussion with the primary care provider or emergency contact.

Fewer than 5% of patient portal users who used the advance care planning tools sent an electronic message to a centralized advance care planning support team (148/3292, 4.5%). The majority of these messages were questions related to how users could submit existing paper advance directives so that these directives would be available in the EHR (this feature is planned but not currently available).

Discussion

Principal Findings

To our knowledge, this is the first study to identify a significant increase in patient-initiated completion of electronic MDPOAs through a patient portal during the COVID-19 pandemic. In the context of COVID-19, conducting advance care planning is increasingly important to ensure that patients have identified medical decision makers whom they trust and that they receive treatment that aligns with their values and preferences. Patient portals are a particularly useful tool for advance care planning during a pandemic because they are accessible 24 hours per day, do not require face-to-face contact, and are directly linked to health care providers and the EHR.

A significant increase in advance care planning portal tool use was seen in April 2020, aligned with increased awareness of the surge in COVID-19 cases, hospitalizations, and deaths in Colorado, the State of Colorado's stay-at-home order (March 26 to April 26, 2020), and the widely publicized authorization of the state's Crisis Standards of Care on April 5. The trend in increased MDPOA completion began in March, which is closely aligned with the announcement of the first COVID-19 case in Colorado (March 5, 2020). Although advance care planning is currently taking place in emergency departments and ICUs [8], these portal users did not submit the MDPOA during an inpatient encounter (data not shown). These users are likely accessing their patient portal account and documenting their preferences at home, away from the clinic and prior to possible emergency care. An existing MDPOA can streamline and enhance the quality of communication during a time of potential health system strain.

Approximately half of the MDPOA forms submitted during this timeframe were new submissions, suggesting that in the setting of the COVID-19 pandemic, people are interested in advance care planning and documenting their health decision makers. COVID-19 has heightened patients' awareness of the potential of becoming seriously ill. These findings show that patients are willing to act on this knowledge without marked investment in health system resources. While health care providers are calling for improved advance care planning during the pandemic, little is known about patient needs and preferences regarding COVID-19–specific advance care planning processes. The technology of the portal may be an important facilitator. An electronic MDPOA can overcome barriers related to lack of documentation access, lack of ability to share completed forms with the health care system, and perceived time, financial, or legal barriers. These users may welcome the ability to complete the MDPOA through the patient portal instead of with an attorney or health care practitioner because of the need for dedicated appointments, which are likely difficult to access due to stay-at-home orders and clinic preference for telehealth appointments.

During a recent qualitative advance care planning patient portal study, older patients and their caregivers indicated a need for easy access to their current active MDPOA via the portal [29]. Access to the MDPOA enables ongoing review and updating of patient preferences, which are important aspects of advance care planning as an ongoing process [4,30,31]. Patients did use the advance care planning message tool, albeit less frequently; over half of the messages during the study period were sent in the month of April. This corresponds with the MDPOA completion spike in April and indicates that people who are interested in using the portal for advance care planning want to use the portal to share existing advance directives or may need technical assistance to use the portal. Patients have previously reported the need for technical assistance in using advance care planning portal tools [29].

Patients across the age spectrum used the advance care planning portal tools. Surprisingly, the largest age group of the users was 25 to 34 years, demonstrating that users in this age group are interested in engaging in advance care planning by choosing a medical decision maker. Advance care planning typically increases with age [32,33]. In Colorado, only 17% of adults aged 25 to 34 years have an advance directive, compared to 66% of adults over the age of 65 [25]. Although it is recommended to engage young adults in advance care planning, the majority of efforts target older adults or focus on young adults with life-limiting illness. Lack of advance care planning awareness and perception among healthy young adults that advance care planning is unnecessary are two major barriers to their involvement in advance care planning. However, young adults who have previous experiences with a seriously ill loved one have demonstrated increased advance care planning awareness and preference for completing an advance directive [30,31]. Contextually, at the end of May 2020, approximately 33% of Coloradans diagnosed with COVID-19 were aged 20 to 39 years (approximately 8728 individuals, as reported on June 6, 2020), and 6.6% of these cases resulted in hospitalization [28]. Because COVID-19 does not only affect older adults, young people may see the value in advance care planning. The patient portal mechanism for advance care planning documentation aligned with health communication and health information access values patients in the younger age demographic; therefore, patient portals are a sustaining strategy for promoting advance care planning in this population [34].

In addition to advance care planning, age is associated with patient portal adoption and use [35,36]. Healthy young adults and older adults, particularly those over the age of 75, are less likely to register for a patient portal account and to regularly use portal features. While this lack of portal use among younger adults is typically associated with their health status, lack of portal use by older adults is attributed to a lack of technology access or technical barriers [37]. Our studies indicate that patients chose to use the portal for advance care planning regardless of age and technical barriers. However, efforts should be made to increase advance care planning awareness among younger adults and address technical concerns among older adults.

In this cohort, women were more likely to complete an MDPOA across all age groups; however, this difference was more pronounced in younger age groups. Gender differences have been found in both patient portal use and advance care planning, with women being more likely to engage with a portal [36] and with advance care planning [25,32,33]. Even among young adults, a national survey found that male respondents reported less familiarity with the concepts of advance care planning, advance directives, and health care proxies [30]. In Colorado, approximately 48% of patients with COVID-19 are male; however, male patients represent 56% of COVID-19 deaths [28].

Limitations

There are limitations to our study. Our data lacked information on race, ethnicity, education status, and specific geographic location (ie, zip code) to determine rurality, which are all known to impact access to health care, health technology use, and advance care planning engagement [32]. We also did not capture the health status, medical diagnoses, prior engagement with advance care planning, or overall preferences related to consumer health technology, which can also be associated with patient portal use. While our study identified a significant increase in engagement with the portal over time, due to the retrospective cohort design of our study, we are unable to conclude that this increase is solely caused by the COVID-19 pandemic rather than other secular trends. Also, it is difficult to predict if the observed growth will be sustained throughout the pandemic period or if there was an immediate effect during the stay-at-home order. As seen in Figure 2, there was a slight decrease in portal use in May; therefore, ongoing investigation of the rate of engagement via the portal is warranted, as attention to advance care planning is beneficial as COVID-19 continues to spread across the United States. However, the use of electronic MDPOAs is not feasible in all states due to state-level requirements; this limits generalizability across states, where state laws may preclude use of the portal for completion of a legal MDPOA. Future research is needed to assess these potential covariates and capture user experiences regarding advance care planning portal use. Research should explore possible associations between patient MDPOA completion and the care provided in the treatment of COVID-19.

Conclusion

The use of a patient portal to complete a MDPOA form to choose a legal medical decision maker increased substantially during the first months of the spread of COVID-19 in the state of Colorado. With the currently increased interest in advance care planning among patients and health care providers, patient portals may be an important tool for advance care planning during the COVID-19 pandemic to enhance and supplement face-to-face advance care planning efforts in health systems.

Acknowledgments

This project would not have been possible without the organizational support of Dr Bennett Parnes through the UCHealth Anschutz and the Lone Tree Seniors Clinics Comprehensive Primary Care Plus programs or the work of multiple UCHealth Epic analysts, including Kate Sanfilippo and Leah Rusinek, building and integrating advance care planning tools into My Health Connection. Additionally, we thank Elisabeth Montgomery for her work through the Advance Care Planning Support Team and Angela Moss for her guidance for the interrupted time series analysis. This initiative is supported in part by the Comprehensive Primary Care Plus Initiative of the Centers for Medicare & Medicaid Innovation, with additional support from the National Institute on Aging of the National Institutes of Health (K76AG054782 [Lum] and K76AG059934 [Portz]) and the Colorado Clinical & Translational Sciences Institute (CTSI) with the Development and Informatics Service Center (DISC) grant support (NIH/NCRR Colorado CTSI Grant Number UL1 RR025780) for use of REDCap-based data management. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The sponsors had no role in the design, methods, or preparation of the manuscript. The views in this article are those of the authors and do not represent the views of the Department of Veterans Affairs.

Conflicts of Interest

None declared.

References

1. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). World Health Organization. 2020 Feb 16. URL: https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf [accessed 2020-07-28]

2. Stokes EK, Zambrano LD, Anderson KN, Marder EP, Raz KM, El Burai Felix S, et al. Coronavirus Disease 2019 Case Surveillance - United States, January 22-May 30, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 19;69(24):759-765 [FREE Full text] [doi: 10.15585/mmwr.mm6924e2] [Medline: 32555134]
3. Advance Care Planning. National Hospice and Palliative Care Organization. URL: https://www.nhpco.org/patients-and-caregivers/advance-care-planning/ [accessed 2020-04-24]
4. Sudore RL, Lum HD, You JJ, Hanson LC, Meier DE, Pantilat SZ, et al. Defining Advance Care Planning for Adults: A Consensus Definition From a Multidisciplinary Delphi Panel. J Pain Symptom Manage 2017 May;53(5):821-832.e1 [FREE Full text] [doi: 10.1016/j.jpainsymman.2016.12.331] [Medline: 28062339]
5. Advance Care Planning: Healthcare Directives. National Institute on Aging. URL: https://www.nia.nih.gov/health/advance-care-planning-healthcare-directives [accessed 2020-04-24]
6. Brinkman-Stoppelenburg A, Rietjens JAC, van der Heide A. The effects of advance care planning on end-of-life care: a systematic review. Palliat Med 2014 Sep;28(8):1000-1025. [doi: 10.1177/0269216314526272] [Medline: 24651708]
7. Houben CHM, Spruit MA, Groenen MTJ, Wouters EFM, Janssen DJA. Efficacy of advance care planning: a systematic review and meta-analysis. J Am Med Dir Assoc 2014 Jul;15(7):477-489. [doi: 10.1016/j.jamda.2014.01.008] [Medline: 24598477]
8. Curtis JR, Kross EK, Stapleton RD. The Importance of Addressing Advance Care Planning and Decisions About Do-Not-Resuscitate Orders During Novel Coronavirus 2019 (COVID-19). JAMA 2020 Mar 27:1771-1772. [doi: 10.1001/jama.2020.4894] [Medline: 32219360]
9. Telemedicine and Telehealth. Office of the National Coordinator for Health Information Technology (ONC). 2017. URL: https://www.healthit.gov/topic/health-it-initiatives/telemedicine-and-telehealth [accessed 2020-05-19]
10. Padala PR, Jendro AM, Gauss CH, Orr LC, Dean KT, Wilson KB, et al. Participant and Caregiver Perspectives on Clinical Research During Covid-19 Pandemic. J Am Geriatr Soc 2020 Jun;68(6):E14-E18 [FREE Full text] [doi: 10.1111/jgs.16500] [Medline: 32315076]
11. Dewar S, Lee PG, Suh TT, Min L. Uptake of Virtual Visits in A Geriatric Primary Care Clinic During the COVID-19 Pandemic. J Am Geriatr Soc 2020 Jul;68(7):1392-1394 [FREE Full text] [doi: 10.1111/jgs.16534] [Medline: 32383773]
12. Hong Y, Lawrence J, Williams D, Mainous I. Population-Level Interest and Telehealth Capacity of US Hospitals in Response to COVID-19: Cross-Sectional Analysis of Google Search and National Hospital Survey Data. JMIR Public Health Surveill 2020 Apr 07;6(2):e18961 [FREE Full text] [doi: 10.2196/18961] [Medline: 32250963]
13. Mahmood S, Hasan K, Colder Carras M, Labrique A. Global Preparedness Against COVID-19: We Must Leverage the Power of Digital Health. JMIR Public Health Surveill 2020 Apr 16;6(2):e18980 [FREE Full text] [doi: 10.2196/18980] [Medline: 32297868]
14. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
15. van der Smissen D, Overbeek A, van Dulmen S, van Gemert-Pijnen L, van der Heide A, Rietjens JA, et al. The Feasibility and Effectiveness of Web-Based Advance Care Planning Programs: Scoping Review. J Med Internet Res 2020 Mar 17;22(3):e15578 [FREE Full text] [doi: 10.2196/15578] [Medline: 32181750]
16. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
17. Bose-Brill S, Pressler TR. Commentary: opportunities for innovation and improvement in advance care planning using a tethered patient portal in the electronic health record. J Prim Care Community Health 2012 Oct 01;3(4):285-288. [doi: 10.1177/2150131911435525] [Medline: 23804174]
18. Vydra TP, Cuaresma E, Kretovics M, Bose-Brill S. Diffusion and Use of Tethered Personal Health Records in Primary Care. Perspect Health Inf Manag 2015;12:1c [FREE Full text] [Medline: 26755897]
19. Lum HD, Sudore RL, Matlock DD, Juarez-Colunga E, Jones J, Nowels M, et al. A Group Visit Initiative Improves Advance Care Planning Documentation among Older Adults in Primary Care. J Am Board Fam Med 2017;30(4):480-490 [FREE Full text] [doi: 10.3122/jabfm.2017.04.170036] [Medline: 28720629]
20. Bose-Brill S, Feeney M, Prater L, Miles L, Corbett A, Koesters S. Validation of a Novel Electronic Health Record Patient Portal Advance Care Planning Delivery System. J Med Internet Res 2018 Jun 26;20(6):e208 [FREE Full text] [doi: 10.2196/jmir.9203] [Medline: 29945860]
21. Brungardt A, Daddato AE, Parnes B, Lum HD. Use of an Ambulatory Patient Portal for Advance Care Planning Engagement. J Am Board Fam Med 2019;32(6):925-930 [FREE Full text] [doi: 10.3122/jabfm.2019.06.190016] [Medline: 31704762]
22. Jordan SR, Brungardt A, Phimphasone-Brady P, Lum HD. Patient Perspectives on Advance Care Planning via a Patient Portal. Am J Hosp Palliat Care 2019 Aug;36(8):682-687 [FREE Full text] [doi: 10.1177/1049909119832820] [Medline: 30803245]
23. Lum HD, Brungardt A, Jordan SR, Phimphasone-Brady P, Schilling LM, Lin C, et al. Design and Implementation of Patient Portal-Based Advance Care Planning Tools. J Pain Symptom Manage 2019 Jan;57(1):112-117.e2 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.10.500] [Medline: 30595147]

24. Walling AM, Sudore RL, Bell D, Tseng C, Ritchie C, Hays RD, et al. Population-Based Pragmatic Trial of Advance Care Planning in Primary Care in the University of California Health System. J Palliat Med 2019 Sep;22(S1):72-81. [doi: 10.1089/jpm.2019.0142] [Medline: 31486723]
25. Advance Care Planning in Colorado. Colorado Health Institute. 2018. URL: https://www.coloradohealthinstitute.org/sites/default/files/file_attachments/CHAS%20advance%20directives%20report%201.pdf [accessed 2020-07-28]
26. State at Home Order. Colorado Department of Public Health and Environment. URL: https://covid19.colorado.gov/stay-home-except-essential-needs [accessed 2020-06-06]
27. Colorado Crisis Standards of Care. Colorado Department of Public Health & Environment. URL: https://www.colorado.gov/pacific/cdphe/colorado-crisis-standards-care [accessed 2020-06-06]
28. COVID-19 Case Data. Colorado Department of Public Health & Environment. URL: https://covid19.colorado.gov/data/case-data [accessed 2020-06-06]
29. Portz JD, Lum HD, Bull S, Boxer RS, Bekelman DB, Ford KL, et al. Perceptions of Patient Portal Use for Advance Directive Documentation among Older Adults with Multiple Chronic Conditions. J Soc Work End Life Palliat Care 2020 Jun 01:1-12. [doi: 10.1080/15524256.2020.1771806] [Medline: 32482160]
30. Tripken JL, Elrod CS. Young Adults' Perspectives on Advance Care Planning. Am J Hosp Palliat Care 2018 Apr;35(4):627-634. [doi: 10.1177/1049909117727456] [Medline: 28845683]
31. Kavalieratos D, Ernecoff NC, Keim-Malpass J, Degenholtz HB. Knowledge, attitudes, and preferences of healthy young adults regarding advance care planning: a focus group study of university students in Pittsburgh, USA. BMC Public Health 2015 Feb 27;15:197 [FREE Full text] [doi: 10.1186/s12889-015-1575-y] [Medline: 25885778]
32. Bischoff KE, Sudore R, Miao Y, Boscardin WJ, Smith AK. Advance care planning and the quality of end-of-life care in older adults. J Am Geriatr Soc 2013 Feb;61(2):209-214 [FREE Full text] [doi: 10.1111/jgs.12105] [Medline: 23350921]
33. Sable-Smith A, Arnett KR, Nowels MA, Colborn K, Lum HD, Nowels D. Interactions with the healthcare system influence advance care planning activities: results from a representative survey in 11 developed countries. Fam Pract 2018 May 23;35(3):307-311 [FREE Full text] [doi: 10.1093/fampra/cmx113] [Medline: 29140508]
34. Peacock S, Reddy A, Leveille SG, Walker J, Payne TH, Oster NV, et al. Patient portals and personal health information online: perception, access, and use by US adults. J Am Med Inform Assoc 2017 Apr 01;24(e1):e173-e177. [doi: 10.1093/jamia/ocw095] [Medline: 27413120]
35. Ancker JS, Osorio SN, Cheriff A, Cole CL, Silver M, Kaushal R. Patient activation and use of an electronic patient portal. Inform Health Soc Care 2015;40(3):254-266. [doi: 10.3109/17538157.2014.908200] [Medline: 24786648]
36. Irizarry T, DeVito Dabbs A, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
37. Gordon NP, Hornbrook MC. Differences in Access to and Preferences for Using Patient Portals and Other eHealth Technologies Based on Race, Ethnicity, and Age: A Database and Survey Study of Seniors in a Large Health Plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]

Abbreviations

COVID-19: coronavirus disease
EHR: electronic health record
ICU: intensive care unit
MDPOA: medical durable power of attorney
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 12.06.20; peer-reviewed by E Belanger, P Padala; comments to author 08.07.20; revised version received 13.07.20; accepted 24.07.20; published 11.08.20.

Please cite as:
Portz JD, Brungardt A, Shanbhag P, Staton EW, Bose-Brill S, Lin CT, Kutner JS, Lum HD
Advance Care Planning Among Users of a Patient Portal During the COVID-19 Pandemic: Retrospective Observational Study
J Med Internet Res 2020;22(8):e21385
URL: http://www.jmir.org/2020/8/e21385/
doi:10.2196/21385
PMID:32716900

Original Paper

Comparing Web-Based and Lab-Based Cognitive Assessment Using the Cambridge Neuropsychological Test Automated Battery: A Within-Subjects Counterbalanced Study

Rosa Backx[1], MSc; Caroline Skirrow[1,2], PhD; Pasquale Dente[1], MSc; Jennifer H Barnett[1,3], PhD; Francesca K Cormack[1], PhD

[1]Cambridge Cognition Ltd, Cambridge, United Kingdom

[2]School of Psychological Science, University of Bristol, Bristol, United Kingdom

[3]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

Corresponding Author:
Caroline Skirrow, PhD
Cambridge Cognition Ltd
Tunbridge Court, Tunbridge Ln
Bottisham
Cambridge, CB25 9TU
United Kingdom
Phone: 44 (0)1223 810 700
Email: caroline.skirrow@camcog.com

Abstract

Background: Computerized assessments are already used to derive accurate and reliable measures of cognitive function. Web-based cognitive assessment could improve the accessibility and flexibility of research and clinical assessment, widen participation, and promote research recruitment while simultaneously reducing costs. However, differences in context may influence task performance.

Objective: This study aims to determine the comparability of an unsupervised, web-based administration of the Cambridge Neuropsychological Test Automated Battery (CANTAB) against a typical in-person lab-based assessment, using a within-subjects counterbalanced design. The study aims to test (1) reliability, quantifying the relationship between measurements across settings using correlational approaches; (2) equivalence, the extent to which test results in different settings produce similar overall results; and (3) agreement, by quantifying acceptable limits to bias and differences between measurement environments.

Methods: A total of 51 healthy adults (32 women and 19 men; mean age 36.8, SD 15.6 years) completed 2 testing sessions, which were completed on average 1 week apart (SD 4.5 days). Assessments included equivalent tests of emotion recognition (emotion recognition task [ERT]), visual recognition (pattern recognition memory [PRM]), episodic memory (paired associate learning [PAL]), working memory and spatial planning (spatial working memory [SWM] and one touch stockings of Cambridge), and sustained attention (rapid visual information processing [RVP]). Participants were randomly allocated to one of the two groups, either assessed in-person in the laboratory first (n=33) or with unsupervised web-based assessments on their personal computing systems first (n=18). Performance indices (errors, correct trials, and response sensitivity) and median reaction times were extracted. Intraclass and bivariate correlations examined intersetting reliability, linear mixed models and Bayesian paired sample t tests tested for equivalence, and Bland-Altman plots examined agreement.

Results: Intraclass correlation (ICC) coefficients ranged from ρ=0.23-0.67, with high correlations in 3 performance indices (from PAL, SWM, and RVP tasks; ρ≥0.60). High ICC values were also seen for reaction time measures from 2 tasks (PRM and ERT tasks; ρ≥0.60). However, reaction times were slower during web-based assessments, which undermined both equivalence and agreement for reaction time measures. Performance indices did not differ between assessment settings and generally showed satisfactory agreement.

Conclusions: Our findings support the comparability of CANTAB performance indices (errors, correct trials, and response sensitivity) in unsupervised, web-based assessments with in-person and laboratory tests. Reaction times are not as easily translatable from in-person to web-based testing, likely due to variations in computer hardware. The results underline the importance of examining more than one index to ascertain comparability, as high correlations can present in the context of systematic differences, which are a product of differences between measurement environments. Further work is now needed to examine web-based

assessments in clinical populations and in larger samples to improve sensitivity for detecting subtler differences between test settings.

(J Med Internet Res 2020;22(8):e16792) doi:10.2196/16792

KEYWORDS

reliability; mobile health; neuropsychological tests; CANTAB; cognition

Introduction

Cognitive function is typically assessed during one-to-one administration of a neuropsychological test in a clinic or lab setting by a trained psychometrician [1]. However, in-person assessments entail significant costs, requiring employed and trained staff, as well as time and travel costs for personnel and participants [2]. These costs may limit their application and reduce resources for clinical and research activities, including patient care, optimizing power for research, and screening for clinical trials [3]. The requirement for one-to-one test administration may also limit participation to people who are willing and able to travel, making some communities underrepresented in clinical research (eg, individuals who are geographically isolated, nondrivers, physically disabled, and those suffering from agoraphobia or social phobias).

Computerized testing platforms and widespread access to fast and affordable internet has the potential to bring neuropsychological assessment into people's homes [2-4]. Web-based neuropsychological assessments could help to meet increasing demands in clinical and cohort studies [3,5]: providing access to large samples, allowing fine-grained phenotyping of complex clinical conditions, facilitating access to patients and participants in remote areas or those with mobility problems, enhancing coordination of data collection across multiple sites, assisting in monitoring of patients with chronic or progressive neurological diseases, and enabling cost-effective screening for clinical trials.

Web-based automated assessments are inexpensive, are quick to conduct, and provide fewer restrictions on timing and location [2,5-7]. Evidence suggests that broadly targeted web-based assessments allow the recruitment of samples that are reasonably representative in terms of personality and adjustment characteristics and are more diverse than traditionally recruited samples in terms of geographical location, gender, and socioeconomic status [7]. Moreover, web-based assessments can reduce the cost of recruiting specialized samples or special interest groups [4,7].

However, the joint position paper for the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology [8] highlights the necessity of viewing unsupervised computer-based tests as new and different from those that are examiner administered, with adaptations of existing tests requiring equivalency or new normative data. Key differences between examiner-led and unsupervised computerized testing relate to 3 primary factors, which are likely to interact with task-specific characteristics (such as simplicity of the user interface, audibility and clarity of stimuli and instructions, type of response required, and how engaging and how difficult a task is) to influence task performance:

1. Examiner contact: Social demands created by the presence of an examiner may affect performance [9]; examiner contact allows for behavioral observations to assess comprehension, mental state and competency, motivation, and task engagement [8,10]; the examiner can also provide additional explanation regarding tasks where needed [11], and structured encouragement to support participant motivation.
2. Testing environment: While the testing environment can be kept constant in the laboratory, it is uncontrolled elsewhere [8,10]. There is little control over the location, timing, and likelihood of participant distraction in unsupervised testing.
3. Workstation: Differences in the performance of computer hardware, software, processing speed, and internet speed, as well as response input method (touch screen versus key stroke or mouse click), are likely to impact test measures, particularly those relating to response timing [12].

Despite the key differences outlined earlier, web-based assessments have proven to be powerful for identifying age-related changes in cognitive processes [13], thus providing reliable data for a longitudinal and quantitative genetic analysis [2,14]. Previous reports have usually shown moderate correlations between web-based cognitive assessments and paper-and-pencil test variants [1,15], and moderate-to-high correlations between parallel computerized test versions assessing a broad range of cognitive domains administered in the lab and at home, or in supervised and nonsupervised settings [16-19]. This suggests that web-based cognitive assessment may be considered a viable alternative to in-person assessment.

Here, we examine the comparability of unsupervised web-based tests completed at home against in-person lab-based assessment in selected tests from the Cambridge Neuropsychological Test Automated Battery (CANTAB). CANTAB is a widely used computerized assessment battery [20], published in over 2000 peer review papers [21], and is widely used in academic, clinical, and pharmacological research [22]. CANTAB tests include a suite of 19 cognitive assessments measuring aspects of cognitive functioning in different therapeutic areas, including attention and psychomotor speed, executive function, memory, and emotion and social cognition. Tasks can be used individually or as a battery to measure different aspects of cognitive function. CANTAB is usually administered under controlled settings in the presence of a trained researcher or clinician.

This study aimed to determine the comparability of unsupervised web-based assessment on CANTAB against a standard in-person assessment in a healthy adult population. The aim was to

examine the consistency of assessment outcomes across these 2 settings, and by extension to inform whether web-based testing could be used as an alternative or as a complementary assessment method producing similar results. We selected 7 tests from CANTAB, which correspond to those most frequently used in academic and clinical research in the cognitive domain of interest.

For web-based testing to show acceptable comparability, we required assessments to (1) show high levels of intersetting reliability, that is, the reproducibility of measures across settings [23], (2) show equivalence with in-person tests, and (3) meet established thresholds for agreement. Given the results from previous research comparing online and in-person tests reviewed earlier, we expected test performance indices to show acceptable comparability. However, we expected reaction time measures to perform more poorly due to the variance introduced by computing software, hardware, and response method.

Methods

Power Analysis

This study was powered to detect moderate-to-high intraclass correlations (ICCs) and moderate-to-large differences in test performance between test settings.

Power calculations to detect ICCs indicating adequate reliability were completed using the R package ICC.Sample.Size [24,25], a statistical package based on the work of Zou et al [26]. Using thresholds for clinical significance developed by Cicchetti [27], the following interpretations were adopted for ICC coefficients (ρ): <0.4, poor reliability; 0.40-0.59, fair; 0.60-0.74, good; 0.75-1.00, excellent. This indicated that a sample of 18 was required to detect an ICC that is indicative of good reliability (ρ=0.60) at 80% power, with a two-tailed α of .05. A sample of 45 would provide adequate power to detect an ICC that is indicative of fair reliability (ρ=0.40).

The power to detect differences between testing platforms was examined using the program G*power 3 [28]. This indicated that detecting an effect size of 0.4, at 80% power (two-tailed α at .05), would require a sample of 52 in a paired sample test with normal distribution, and between 35 and 47 for the nonparametric equivalent, depending on the underlying distribution of data (laplace and logistic, respectively). An effect size of 0.4 has been reported as relatively typical within psychological sciences [29,30]. This study utilizes the Bayesian approach as an adjunct to our frequentist analysis to consider the strength of evidence in favor of both the alternative and null hypotheses and compare their probabilities [31].

Participants

Participants were approached via fliers and advertisements posted on Facebook, targeting Cambridge, United Kingdom, and the immediate surrounding areas. These directed potential participants to a web-based screening questionnaire, administered via SurveyMonkey [32], through which participants provided basic demographic data (sex, age, and education level) and responses to questions probing eligibility for the study (exclusion criteria: history of dyslexia, concussion, head injury, neurological or psychiatric conditions, and nonfluent in English).

A total of 51 healthy adults were recruited into this study (32 women and 19 men), aged between 20 and 77 years, with a mean age of 36.8 (SD 15.6) years. Participants were highly educated, with 17.6% with school-level qualifications and 82.4% with university-level education, reflecting the demography of this region. All participants provided informed written consent to participate.

Procedure

Participants were allocated to one of the two groups (in-person first or web-based first), through randomization at the time of recruitment. However, where necessary, allocation from randomization was overridden, where participant availability or laboratory space constricted the timing of assessments. The allocation of test sessions was as follows: in-person testing first for 33 participants and web-based assessment first for 18 participants. Test sessions were completed on average 1 week apart (mean 7.24, SD 4.5 days, range 1-25 days, with the majority [82% of tests] between 3 and 9 days), again with variation due to participant and laboratory availability.

In-person assessments were completed at Cambridge University, Cambridge, United Kingdom. Participants were seated in a quiet room and presented with CANTAB loaded onto an iPad (iPad 9.7, IOS operating system, [33]). The CANTAB test administration is fully automated, with on-screen text instructions and additional voiceover guidance for each task, explaining task goals and response requirements. For tests requiring training in addition to instruction (see *Measures*), training trials are incorporated within the automatic test administration. The transition from training to tests proceeds automatically, as do transitions between tests. Responses were logged via the touch screen. A trained psychometrician was present, whose role was to provide technical support where needed or additional instructions where required as well as to log observations (distraction or problems) during task performance.

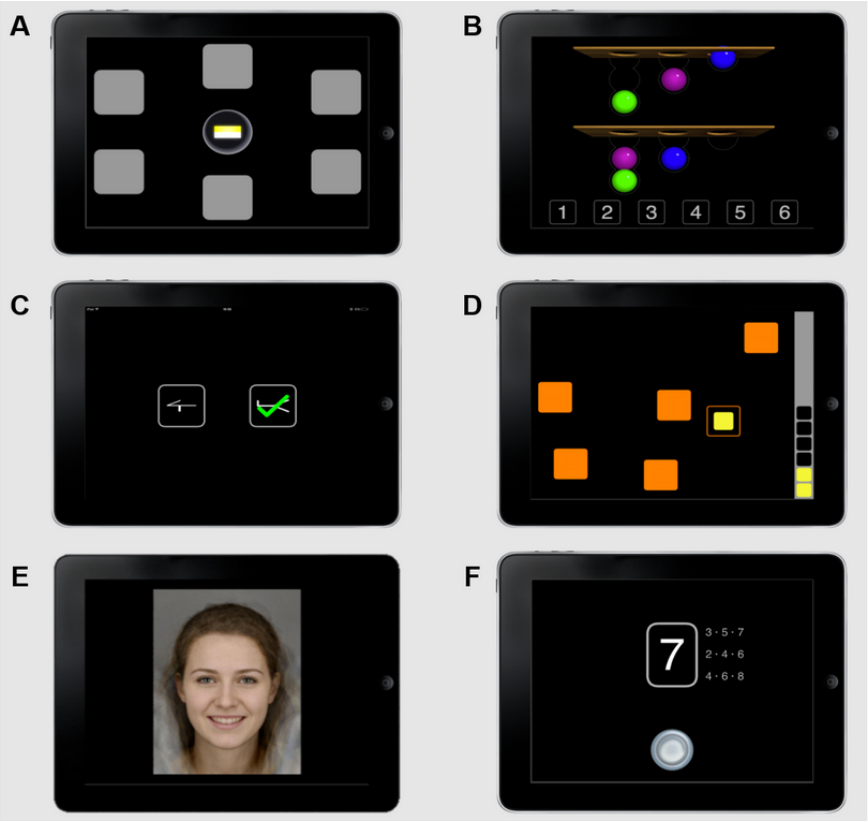
Web-based assessments were completed via the CANTAB Connect web-based testing feature [34]. This delivered assessments which, from the viewpoint of the participant, were identical to those administered in-person, with the exception that they were administered at home and on personal computing systems. Web-based testing was enabled only on desktop or laptop computers, and not on touch screen devices. Responses were logged using mouse or trackpad clicks. Identical to in-person assessments, test administration was automated, with on-screen text instructions and additional voiceover guidance for each task, training incorporated into tasks where required, and automatic transitions between tests. Web-based CANTAB tests are designed to be resistant to low bandwidth by preloading or caching of data, allowing tests to be run in offline mode in testing locations where internet connectivity is poor. The application code is designed for cross-browser support and uses ubiquitous HTML and JavaScript features to support commonly used platforms. Extensive automated and manual tests are carried out to test functionality across browsers and ensure that the tests operate correctly and record accurate data.

Distraction during web-based assessment was documented with inbuilt programming to log if tasks were completed in full-screen mode, or if the participant tabbed to another browser window during the task. Participants were also asked at the end of the testing session if they were distracted during testing, although the nature of the distraction was not queried. These different forms of distraction were logged, but not differentiated, in the study database during data collection.

Measures

A total of 7 CANTAB tests (Figure 1) were administered. Cognitive outcome measures include performance indices (eg, number of trials solved, number of errors, response sensitivity) and reaction time (response times). For both in-person and web-based assessments, tests were administered in the following order:

Figure 1. Screenshots of Cambridge Neuropsychological Test Automated Battery tests administered: (A) Paired Associate Learning, (B) One Touch Stockings of Cambridge, (C) Pattern Recognition Memory, (D) Spatial Working Memory, (E) Emotion Recognition Task, and (F) Rapid Visual Information Processing.



1. Paired associate learning (PAL) [22] is an 8-min test of visual episodic memory. The screen displays a number of boxes and shows the interior of each box in randomized order to briefly reveal patterns in some boxes. Patterns are then displayed in the middle of the device screen one at a time, and the participant must identify the box in which each pattern was originally located. If an error is made, boxes are opened in sequence again to remind participants of the pattern locations. The test begins with a practice trial, which includes 6 boxes in which there are 2 patterns. Once the practice trial is successfully completed, the test begins. The task increased in difficulty after each successfully completed stage, with trials including 2, 4, and 6 different patterns in 6 boxes, and finally 8 different patterns in 8 boxes. The task discontinues when a participant fails to locate all patterns after 4 attempts on the same trial. Key outcome measures included PAL Total Errors Adjusted, the total number of errors adjusted for the stages not completed due to early discontinuation, and PAL First Attempt Memory Score, the number of times a participant chooses the correct box on their first attempt across each stage.
2. One touch stockings of Cambridge (OTS) [35] is a 10-min test of executive function, assessing spatial planning and working memory, and based on the Tower of London test. The screen shows 2 displays, each containing 3 colored balls that look like stacks held in stockings or socks suspended from a beam. The target configuration is shown at the top of the screen and the starting arrangement below. The subject must determine the number of moves required to match the starting configuration to the target. One move consists of taking 1 ball from its current location and placing it in a stocking that has free space. Only the top ball in any stocking may be moved (the balls below are inaccessible until any balls above have been moved), and a ball placed in a stocking drops to the lowest free space available. Participants must solve each problem without moving the balls, by indicating the number of moves required by selecting a numbered box at the bottom of the screen. The task begins with 3 training trials. The first two show how the balls would be moved before participants select their

response, and the third only shows the solution when the participants' response is incorrect. Once training is completed, the task then progresses with increasing difficulty. Key outcomes included problems solved on first choice and median latency to correct response.

3. Pattern recognition memory immediate (PRM-I) [36] is a 3-min test of immediate visual pattern recognition. A series of 18 simple but abstract stimulus patterns are shown in the center of the screen for 3000 ms each. The screen then displays pairs of patterns, one novel pattern and one that was shown previously. The participants have to select patterns that they recognize from the presentation phase. Participants receive performance feedback in the form of a tick or cross after every response. Key outcome variables include the percentage of correct responses and median latency of correct responses.
4. Spatial working memory (SWM) [35] is a 4-min test of retention and manipulation of visuospatial information. Participants click on colored boxes presented on the screen to inspect their contents and reveal a token hidden below. They then move these tokens to a collection area on the right-hand side of the screen. The key task instruction is that tokens will not be located in the same box twice during each trial. Outcome measures include SWM Between Errors: the number of times the participant incorrectly revisits a box, calculated across all assessed 4, 6, and 8 token trials; and SWM Strategy: the number of unique boxes from which a participant starts a new search in the 6 and 8 box trials. More efficient searches are carried out by searching boxes in a fixed order [37]. The task discontinues after 20 failed inspections during 4-token trials, 30 failed inspections for 6-token trials, and 40 failed inspections for 8-token trials.
5. The emotion recognition task (ERT) [38] is a 7-min test measuring participants' ability to identify 6 basic facial emotion expressions along a continuum of expression magnitude. Participants fixate on a white "+" cross in the center of the screen for 1500 to 2500 ms, after which a face stimulus is displayed for 200 ms followed by a stimulus mask image for 250 ms. Participants then choose the most appropriate emotion from a list of 6 options (sadness, happiness, fear, anger, disgust, or surprise). Outcome measures included the total number of hits and median latency to correct responses.
6. Pattern recognition memory delayed (PRM-D) is a 2-min test of delayed visual pattern recognition. Patterns displayed for PRM-I are revisited and recognition is probed in the same manner as described in (3) after delay. In this study, the delay between PRM-I and PRM-D was approximately 12 min. Key outcome variables include the percentage of correct responses and median latency of correct responses.
7. Rapid visual information processing [39] (RVP) is a test of sustained attention lasting 7 min. Digits from 2 to 9 are presented successively at the rate of 100 digits per minute and in a pseudorandom order. Participants are asked to respond to target sequences of digits (eg, 3-5-7, 2-4-6, 4-6-8) as quickly as possible by clicking or pressing a button at the center of the device screen. The level of difficulty varies with either 1- or 3-target sequences that the participant must watch for at the same time. Outcome measures included a signal detection measure of response sensitivity to the target, regardless of response tendency (RVP A': expected range is 0-1) and the median response latency.

CANTAB test structures are identical for each administration, across both in-person and web-based assessments. However, for most CANTAB tests (OTS, PAL, RVP, PRM, and ERT), stimuli are allocated at random from a broader stimulus pool during each assessment, making it unlikely that participants complete the same problems more than once. For the SWM test, token locations are not fixed but instead programmed to respond to participants' performance and selection strategy, reducing the risk of participants being able to learn the location of tokens from one assessment to the next. These adaptive features aim to reduce practice effects on repeat testing and also mean that there are no set variants of the tests that can be compared in a group-wise fashion.

Statistical Analysis

Frequentist analyses including mixed models, regressions, correlational analysis, and ICCs were completed in SAS version 9.4. Statistical significance thresholds were set at $P \leq .05$ (two tailed). The Bayesian statistical analysis was carried out using JASP [40].

Outliers were identified using the methods recommended by Aguinis et al [41], first through visual plotting and then confirmed numerically, using a cutoff of 2.24 SD units above or below the mean. One data point was excluded from each of the following assessments: RVPA, RVP Median Latency to Correct Response, PRM Percentage Correct Immediate, and PRM Median Latency Immediate and Delayed (ranging 4.5-6.9 SD units from mean, all acquired during the web-based assessment).

To allow the comparison with test-retest reliabilities commonly reported in the literature [3,5,18,42], bivariate coefficients were computed to measure the strength of the linear association of outcome measures across test settings. Spearman rank correlations are reported because of the nonnormal distribution of data. To control for variation in the duration between assessments, partial correlations were completed, which examined correlations of test results between settings after covarying for the duration between tests.

However, although the correlational analysis reflects the degree to which paired observations follow a straight line, they do not inform regarding the slope of the line or whether the sets of observations capture the same metric or range of scores [43]. ICCs were selected as the primary reliability measure, because ICCs assume that the variables investigated share both their metric and variance and incorporate both random and systematic errors when calculating consistency between assessments [44,45]. ICCs therefore account for both consistency in performance (the degree of correlation) between test settings as well as capturing any systematic changes in the mean (the degree of agreement) [46]. Following guidance by Koo and Li [46] and justifications outlined in detail in Hansen et al [5], ICC was calculated based on a single-rating, absolute agreement,

two-way random effects model (ICC 2,1 [47]). ICC coefficients were computed using the %INTRACC macro for SAS [48]. In line with previous studies and interpretative recommendations for ICC, we used ρ≥.60 to indicate good reliability [18,27].

Mixed effects models simultaneously investigated differences between the test settings (in-person vs web-based) and time (first vs second assessment). Mixed effects models can evaluate multiple factors that affect the structure of the data and allow longitudinal effects (practice and learning effects) to be straightforwardly incorporated into the statistical model [49]. Outcome measures were entered individually into each model as dependent variables, and 2 mixed effects models were analyzed for each outcome measure. The first model examined only the fixed effects of test setting and time of assessment, with participants entered into the model as a random effect. A second model was used to examine the presence of covariates that may affect test performance across settings, and included additional fixed effects of age, an age-by-setting interaction, and distraction during web-based testing (dummy coded as 1=distracted, 0=not distracted). This second model tested whether age affected performance and interacted with assessment setting to affect test results, and whether distraction during web-based assessment contributed to differences in test results.

The normality of the distribution of residuals was examined, and where required data were transformed before data analysis. Transformations included log transformations for PAL Total Errors Adjusted, SWM Between Errors, OTS Problems Solved on First Choice, and OTS Median Latency to Correct response and square root transformation for PAL First Attempt Memory Score. For most variables, transformations were successful and a linear mixed model was carried out (SAS command PROC MIXED). For PRM-I and PRM-D percentage correct, transformations were not successful. These data were reverse transformed (calculated as the percentage correct subtracted from 100) and were analyzed with mixed models with gamma error distributions and log links (SAS command PROC GLIMMIX).

Evidence in favor of the null hypothesis was examined using a Bayesian approach [50]. The advantage of using the Bayes factor over classical significance testing is that it provides a comparison of how likely the null hypothesis is compared with the alternative hypothesis [31]. Bayesian paired samples *t* tests were conducted, and Bayes factor test statistics were extracted, alongside effect sizes (δ) and their 95% credible intervals, contrasting the likelihood of data fitting under the null hypothesis (H_0: no difference between test settings) with the alternate hypothesis (H_1: that there is a difference between test settings). A default Chauchy prior width of r=0.707 was selected, and a Bayes factor robustness check was completed to examine if the qualitative conclusions changed with reasonable variations to the prior width. Bayes factors (BF_{10}) were interpreted using a classification scheme adopted from Wagenmakers et al [51]: with Bayes factors below 1 seen as evidence for the null hypothesis (0.33-1: anecdotal evidence; 0.1-0.33: moderate evidence; <0.1 strong evidence for H_0), and Bayes factors above 1 seen as evidence for H_1.

Agreement between test settings was examined with Bland-Altman plots [52]. These plot the difference between assessments (eg, *A*−*B*) versus the average across paired measures (*A*+*B*/2), along with 95% limits of agreement [53]. The plots serve as a visual check that the magnitude of the differences is comparable throughout the range of measurement. Distributions of difference scores were assessed using Kruskal-Wallis tests, and where these were nonnormally distributed, raw data were log transformed before plotting and analysis. Other transformations were not considered, as these are not advised for this method of analysis [52,54]. Agreement is considered adequate when 95% of data points lie within limits of agreement [52]. Proportional bias was examined by regressing difference scores against mean scores to identify the tendency for the difference to increase or decrease with higher score magnitudes [55].

Results

Test Completion

Full test data were obtained from all participants with the exception of 2 individuals for whom the SWM test terminated early due a large number of errors made during web-based assessment. During in-person assessments, support from the examiner was required on 4 occasions (3 times for volume adjustment during PAL testing and once for additional instruction on the PRM immediate recognition task). Distraction, either through self-report or due to participants tabbing away from the assessment window during web-based assessments, was noted for 16 participants for PAL, ERT, OTS, and PRM-I tests and for 17 participants during SWM, RVP, and PRM-D tests.

Reliability

Bivariate correlation coefficients and ICCs are shown in Table 1. Spearman correlation coefficients across testing settings ranged from 0.39 to 0.73 (P<.01). ICCs ranged from 0.23 to 0.67 (P≤.05). A total of 5 tests had ICC coefficients meeting the cutoff at ≥0.60, with PAL Total Errors Adjusted just meeting requirements (exact ICC coefficient=0.595, rounded up), and above threshold coefficients for RVP A', SWM Between Errors, PRM-I Median Latency, and ERT Median Correct Reaction Time. Partial correlations of test results across settings after controlling for the duration between tests produced very similar results. These are shown in Multimedia Appendix 1.

Equivalence

Descriptive statistics and results from the mixed model assessing fixed effects of test setting and time are presented alongside the Bayesian analysis results in Table 2. Mixed models revealed no significant differences between in-person and web-based assessments for performance indices (P=.10 to .54). However, 3 of the 5 reaction time measures showed differences across test settings (response latencies for PRM-I, PRM-D, and ERT tasks), with web-based assessments yielding slower median response times (P<.001 to .03). Practice effects were seen for RVP and SWM performance indices, showing improvement on second administration (P<.01). Response latencies were faster on the second administration for OTS responses (P=.001).

Additional fixed effects of age, an age-by-setting interaction effect, and distraction were incorporated into mixed models. Age effects on test performance, showing a decline in test performance with increasing age, were found for all outcome measures with the exception of RVP A', the percentage of correct responses on PRM-I and PRM-D, and OTS Problems Solved on First Choice. No significant age-by-setting interactions were observed, indicating that test performance did not differ between in-person and web-based testing as a function of age, although there was a trend for slower reaction times on web-based testing for older participants on the PRM-I task (PRM-I Median Latency: $F_{1,45}$=4.01, P=.051; for all other tests F statistic range 0.02-2.49; P=.12 to .90). Effects of distraction were nonsignificant for most tests, but reached or neared significance thresholds for certain reaction time measures (ERT Median Correct Reaction Time: $F_{1,47}$=6.03, P=.02; RVP Median Reaction Time: $F_{1,46}$=3.78, P=.06).

Bayesian analyses supported the null hypothesis (H_0: no difference between test settings) over the alternate hypothesis: BF_{10}=0.161-0.54) for all performance indices. Applying the classification scheme adopted from Wagenmakers et al [51], support for the null hypothesis was anecdotal for 3 variables (PAL First Attempt Memory Score, SWM Strategy, and ERT Total Hits), and moderate for 6 other performance indices. No change in the qualitative conclusions was seen with reasonable variations in the prior width. The effect sizes were small (0.15-0.27).

The alternate hypothesis, reflecting a difference between test settings, was supported for 3 out of the 5 reaction time measures (response latencies on PRM-I, PRM-D, and ERT tasks), with support being between anecdotal for the PRM measures (BF_{10}=1.60-2.15) and very strong for ERT (BF_{10}=512557.32). Effect sizes were in the low-to-large range (0.04-1.69). Moderate support for the null hypothesis was seen for the RVP and OTS reaction time measures.

Table 1. Reliability analysis for outcome measures of Spearman correlation coefficients and intraclass correlations between test results obtained in-person and in web-based assessments.

Outcome variable	Spearman correlation		Intraclass correlation	
	Correlation coefficient	*P* value	Correlation coefficient	*P* value
PAL[a] total errors adjusted	0.54	<.001	0.60	<.001
PAL first attempt memory score	0.45	.001	0.51	<.001
OTS[b] problems solved on first choice	0.39	.005	0.40	.002
OTS median latency to correct	0.55	<.001	0.45	<.001
PRM-I[c] percentage of correct trials	0.40	.004	0.34	.008
PRM-I median latency	0.61	<.001	0.65	<.001
SWM[d] between errors	0.61	<.001	0.62	<.001
SWM strategy	0.50	<.001	0.49	<.001
ERT[e] total hits	0.54	<.001	0.57	<.001
ERT median correct reaction time	0.73	<.001	0.61	<.001
PRM-D[f] percentage of correct trials	0.49	<.001	0.49	<.001
PRM-D median latency	0.57	<.001	0.56	<.001
RVP[g] A'	0.71	<.001	0.67	<.001
RVP median latency	0.41	.003	0.23	.048

[a]PAL: paired associate learning.

[b]OTS: one touch stockings of Cambridge.

[c]PRM-I: pattern recognition memory immediate.

[d]SWM: spatial working memory.

[e]ERT: emotion recognition task.

[f]PRM-D: pattern recognition memory delayed.

[g]RVP: rapid visual information processing.

Table 2. Descriptive data for outcome variables and statistical results for equivalence analyses. Time at assessment (first vs second assessment) and test setting (in-person or web-based). Mixed effects model and Bayesian t test statistics

Outcome variable	Descriptive statistics				Mixed model test statistics				Bayesian paired *t* test statistics	
	Time of assessment, mean (SD)		Test setting, mean (SD)		First vs second assessment		In person vs web based		In person vs web based	
	First assessment	Second assessment	In-person	Web-based	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value	Bayes factor H_1	Effect size δ (95% Credible Intervals)
PAL[a] total errors adjusted	12.06 (13.76)	11.00 (13.21)	12.43 (14.53)	10.63 (12.36)	0.12 (1,49)	.73	0.99 (1,49)	.33	0.259	−0.19 (−0.58 to 0.18)
PAL first attempt memory score	14.49 (4.31)	14.57 (3.83)	14.14 (4.28)	14.92 (3.81)	0.25 (1,49)	.62	1.87 (1,49)	.18	0.383	0.25 (−0.13 to 0.63)
OTS[b] problems solved on first choice	11.73 (1.81)	11.90 (1.95)	11.69 (1.96)	11.94 (1.79)	0.06 (1,49)	.81	0.70 (1,49)	.41	0.221	0.15 (−0.21 to 0.53)
OTS median latency to correct (ms)	13933.22 (8130.39)	11525.04 (6651.39)	12718.31 (8124.91)	12764.24 (6878.99)	11.50 (1,49)	.001	2.18 (1,49)	.15	0.153	0.01 (−0.36 to 0.39)
PRM-I[c] percentage correct	92.48 (13.67)	92.17 (12.76)	92.98 (11.83)	91.77 (14.48)	0.09 (1,37)	.77	0.91 (1,37)	.35	0.290	0.18 (−0.56 to 0.18)
PRM-I median latency (ms)	1533.67 (367.19)	1587.89 (448.00)	1506.80 (376.83)	1615.29 (434.55)	0.25 (1,48)	.62	4.36 (1,48)	.04	1.60	0.41 (0.04 to 0.80)
SWM[d] between errors	7.80 (8.08)	4.92 (6.40)	6.96 (7.27)	5.82 (7.59)	7.59 (1,47)	.008	1.15 (1,47)	.29	0.229	0.16 (−0.55 to 0.20)
SWM strategy	7.04 (2.55)	5.67 (2.68)	6.71 (2.65)	6.02 (2.72)	12.50 (1,47)	<.001	0.71 (1,47)	.40	0.479	0.27 (−0.65 to 0.10)
ERT[e] total hits	30.67 (4.18)	30.53 (4.52)	31.06 (4.13)	30.14 (4.5)	0.06 (1,49)	.80	2.78 (1,49)	.10	0.54	0.29 (−0.67 to 0.06)
ERT median correct reaction time (ms)	1274.50 (414.45)	1370.55 (510.74)	1174.45 (419.26)	1470.60 (465.65)	0.04 (1,49)	.83	39.79 (1,49)	.001	512557.32	1.25 (0.82 to 1.69)
PRM-D[f] percentage correct	89.87 (12.73)	88.83 (15.49)	89.87 (14.17)	88.07 (14.14)	1.44 (1,31.2)	.24	0.63 (1,31.2)	.43	0.217	0.15 (−0.52 to 0.22)
PRM-D median latency (ms)	1731.65 (417.72)	1801.47 (463.71)	1698.15 (462.43)	1835.64 (409.44)	0.32 (1,48)	.58	4.90 (1,48)	.03	2.15	0.44 (0.05 to 0.83)
RVP[g] A’	0.92 (0.05)	0.95 (0.04)	0.94 (0.04)	0.94 (0.05)	29.29 (1,48)	<.001	0.38 (1,48)	.54	0.161	0.17 (−0.21 to 0.54)
RVP median latency (ms)	452.96 (84.52)	436.98 (71.60)	449.32 (72.05)	440.37 (84.66)	0.97 (1,48)	.32	0.31 (1,48)	.58	0.183	−0.11 (−0.47 to 0.27)

[a]PAL: paired associate learning.

[b]OTS: one touch stockings of Cambridge.

[c]PRM-I: pattern recognition memory immediate.

[d]SWM: spatial working memory.

[e]ERT: emotion recognition task.

[f]PRM-D: pattern recognition memory delayed.

[g]RVP: rapid visual information processing.

Agreement

Bland-Altman plots showed overall good agreement between test settings for performance indices (see Figure 2, for example, for PAL Total Errors Adjusted). Only 2 performance indices fell short of the requirement that 95% of the data points should lie within limits of agreement (PAL First Attempt Memory Score and SWM Strategy, with 94% and 92% of data points within limits of agreement, respectively). The PAL First Attempt Memory Score showed a proportional bias ($F_{1,50}$=7.43; P=.009; R^2=0.13), with lower mean scores being associated with greater difference between measurements (Figure 3). For all other performance measure plots, no bias was seen relating to the test setting, and difference magnitudes were comparable throughout the range of measurements. Performance data from PRM tasks and from SWM Between Errors could not be accurately visualized using Bland-Altman plots because of significant nonnormality of the difference scores that could not be corrected through logarithmic transformation.

Figure 2. Comparability of Paired Associate Learning Total Errors Adjusted across test settings. Density plot for (A) web-based assessment and (B) in-person assessment showing similar distributions; (C) scatterplot with reference line showing linear relationship between assessment settings (ρ=0.54); (D) Bland-Altman plot: mean difference (solid black line) is close to zero, showing no bias; dashed lines delimit limits of agreement. Comparable magnitudes of difference are seen throughout the range of measurements, and 96% of the data within limits of agreement.

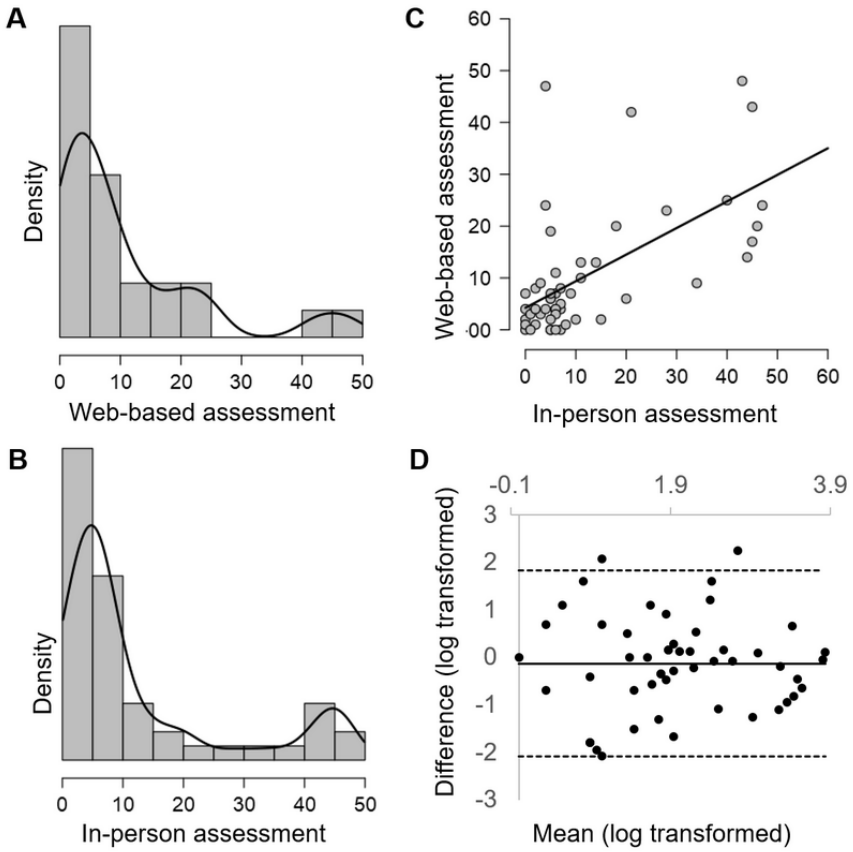
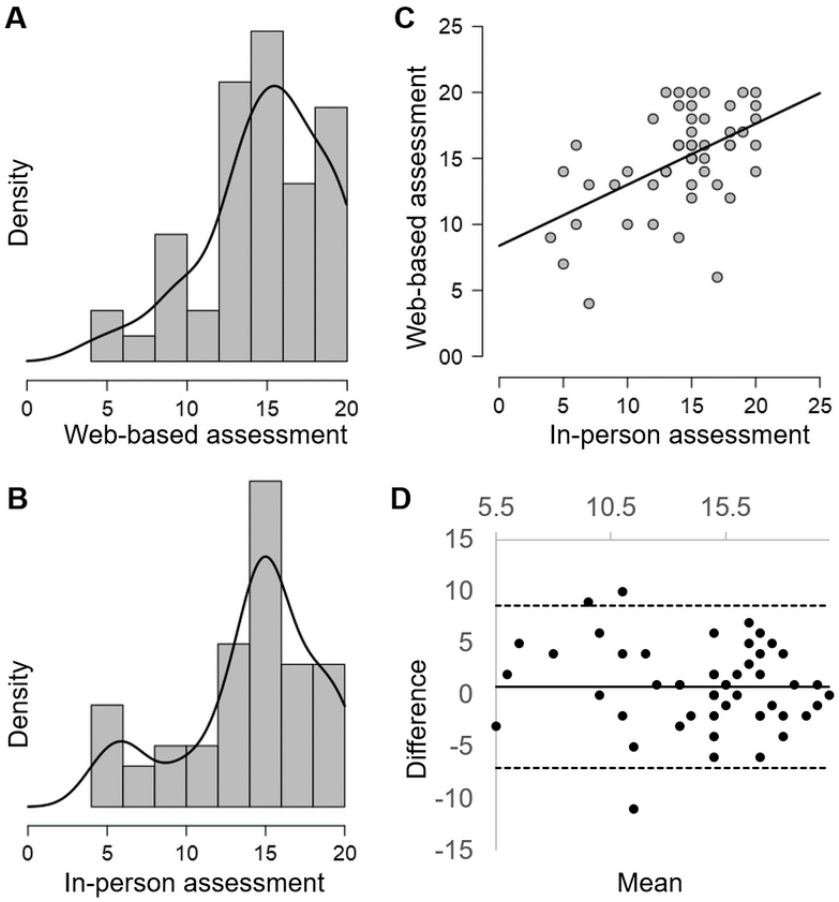
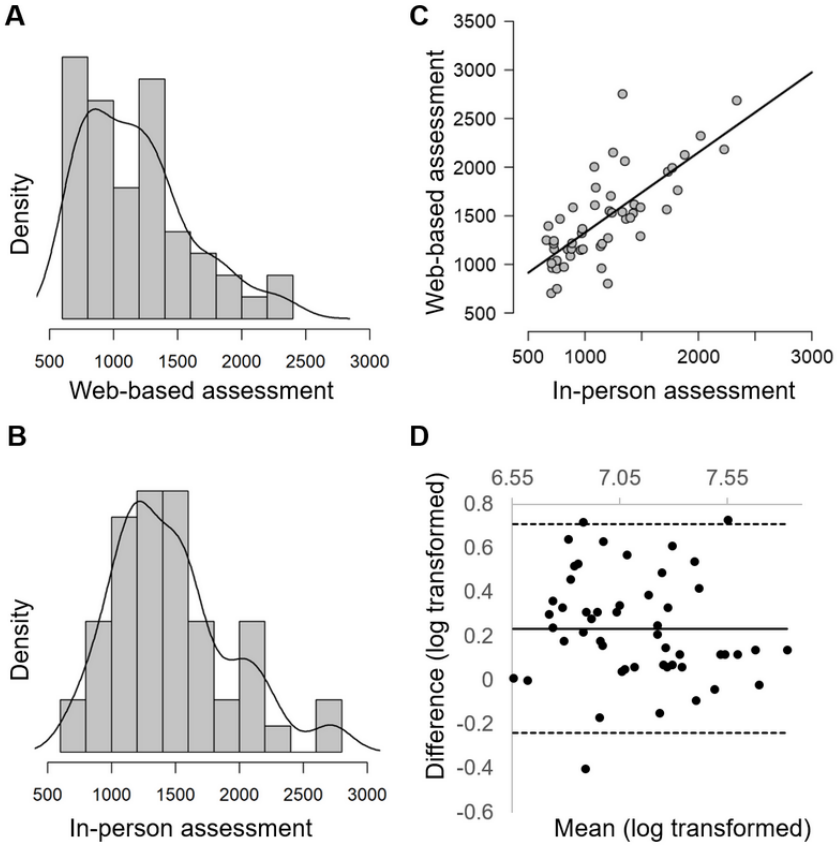


Figure 3. Comparison of Paired Associate Learning First Attempt Memory Score across test settings. Density plot for (A) web-based assessment and (B) in-person assessment showing similar distributions; (C) scatterplot with reference line showing linear relationship between assessment settings (ρ=0.45); (D) Bland-Altman plot: mean difference (solid black line) is close to zero, showing no bias; dashed lines delimit limits of agreement. Proportional bias is seen: greater differences at lower mean measurements and 94% of data within limits of agreement.



For reaction time measures, Bland-Altman plots reflected bias in test settings in PRM-I and PRM-D response latencies and ERT Median Correct Reaction Time (eg, Figure 4), confirming the findings from the mixed model and Bayesian analyses. Additionally, for all reaction times, 94% of the data points were within limits of agreement, falling short of the 95% cutoff. Visual inspection of the plots confirmed comparable magnitudes of difference throughout the range of measurements, and regression analyses revealed no proportional bias (R^2 range 0-0.05; P=.12 to .67).

Figure 4. Comparability of Emotion Recognition Task median correct reaction time (in ms) across test settings. Density plot for (A) web-based assessment and (B) in-person assessment, showing broader distribution of timings (range 500-3000 ms) and slower overall timings for web-based assessment compared to in-person assessment (range 500-2500 ms); (C) scatterplot with reference line showing strong linear relationship between assessment settings (ρ=0.73); (D) Bland-Altman plot: mean difference (solid black line) is shifted above zero, demonstrating bias; dashed lines show limits of agreement. Comparable magnitudes of difference are seen throughout the range of measurements, and 94% of the data within limits of agreement.



Discussion

This study examines the comparability of the widely used CANTAB administered unsupervised via the internet against a typical in-person lab-based assessment, using a counterbalanced within-subjects design. We imposed strict criteria for comparability, including satisfactory intersetting reliability, equivalence, and agreement across test settings. Overall, our results support the comparability of performance indices (errors, trials completed, and response sensitivity) acquired during web-based assessments. Reaction time measures show poorer comparability, with results revealing significant differences and poor agreement between test settings.

Bivariate correlation coefficients between the 2 modes of test administration ranged between 0.39 and 0.73, broadly in keeping with previous research comparing in-person and web-based assessment of other cognitive tasks [16,18,19]. The correlations reported here are similar to previously reported test-retest correlations in the CANTAB tests. An overview of test-retest correlations for CANTAB performance indices from previously published papers (and in different test populations) can be seen in Multimedia Appendix 1.

ICCs were higher for some tests than for others, with fair reliabilities (ICC ρ=0.40-0.49) seen for planning and executive function tasks (SWM Strategy and OTS performance measures). Previous research has shown that cognitive measures are subject to significant intraindividual variation [56]. A meta-analysis showed that test-retest reliabilities can differ depending on the tests completed and the cognitive functions that they tap into, with lower reliability typically seen for tests assessing executive functions and memory [57]. Poor reliability was seen for PRM-I percentage of correct trials in this study, which could be attributable to the low variance and high ceiling-level performance on this task in this healthy volunteer sample.

ICCs and Spearman correlations generally provided similar results, but showed greater discrepancy for reaction times, where there was a difference in the range and average between assessment settings. In these cases, ICCs typically presented a tempered correlation coefficient in comparison to Spearman correlations, reflecting that this statistic takes into account systematic error between assessments.

Learning effects are likely to have had an impact on concordance between test settings [16]. Practice effects with improvement on the second test administration were seen for 4 outcome measures (RVP A', SWM Strategy, SWM Between Errors, and OTS Median Latency to Correct response). Previous work has shown increased susceptibility to specific tests, in particular those assessing visual memory, to practice effects [58]. The novelty of a test, particularly in the executive function domain, is also thought to influence susceptibility to practice effects [59]. Owing to these effects, it is recommended that a

familiarization session, to reduce the immediate effect of novelty of tests and testing procedures, is used before baselining cognitive performance in clinical trials and other within-subject designs. Practice effects were not seen for the remaining outcome measures, which may be due to the use of alternate test stimuli [57]. In most CANTAB tests, stimuli are allocated at random from a broader stimulus pool during each assessment, reducing the likelihood that participants completed the same problems more than once.

Two out of 9 performance indices met all predefined criteria for comparability between measures. PAL Total Errors Adjusted and RVP A' test scores did not differ between test settings, showed good intersetting reliability, and showed acceptable agreement on Bland-Altman plots. Additionally, for SWM Between Errors, Bland-Altman analyses were not completed, but the intersetting reliability was good, and there was no evidence of performance differences between settings. These measures are therefore determined to have good overall comparability vis-à-vis typical in-person assessment (overview shown in Table 3).

Table 3. The overall assessment of web-based outcome measures with regard to 3 criteria.

Outcome variable	Reliability[a]	Equivalence[b]	Agreement[c]
Performance indices			
PAL[d] total errors adjusted	✓[e]	✓	✓
PAL first attempt memory score	x[f]	✓	x
OTS[g] problems solved on first choice	x	✓	✓
PRM-I[h] percentage of correct trials	x	✓	—[i]
SWM[j] between errors	✓	✓	—
SWM strategy	x	✓	x
ERT[k] total hits	x	✓	✓
PRM-D[l] percentage of correct trials	x	✓	—
RVP[m] A'	✓	✓	✓
Reaction time measures			
OTS median latency to correct	x	✓	x
PRM-I median latency	✓	x	x
ERT median correct reaction time	✓	x	x
PRM-D median latency	x	x	x
RVP median latency	x	✓	x

[a]: reliability criterion met where intraclass correlation coefficients ≥0.60.

[b]: equivalence criteria met where there is no significant difference between performance levels across test settings in mixed effects models, and data supporting the null hypothesis for Bayesian paired *t* tests).

[c]: agreement criteria met where ≥95% of data points lie within the limits of agreement on Bland-Altman plots, and there is no evidence of bias or proportional bias.

[d]PAL: paired associate learning.

[e]✓: criteria met.

[f]x: criteria not met.

[g]OTS: one touch stockings of Cambridge.

[h]PRM-I: pattern recognition memory immediate.

[i]—: analyses not completed.

[j]SWM: spatial working memory.

[k]ERT: emotion recognition task.

[l]PRM-D: pattern recognition memory delayed.

[m]RVP: rapid visual information processing.

Two additional performance indices were determined to have moderate comparability with respect to in-person assessment. The ERT Total Hits and OTS Problem Solved on First Choice outcome measures showed good equivalence and agreement, but below the threshold reliability indices. For the ERT Total Hits, the ICC fell just short of the imposed threshold (ICC coefficient=0.57).

Overall, none of the 5 web-based reaction time measures met more than one of the predefined comparability criteria,

indicating that response latency measures are less easily translated from the lab to the home. Acceptable correlations between in-person and web-based assessments were undermined by a lack of equivalence and agreement between the measures. Correlation coefficients examine the linear relationship and relative consistency between 2 variables (the consistency of the position or rank of individuals in one assessment relative to the other [45]) rather than the absolute agreement between measurements within individuals [52,55], and are therefore insensitive to differences in metrics or variance (Figure 4).

Differences between settings could be due to a variety of factors. First, web-based assessments were completed on laptop and desktop computers that participants had readily available to them at home or elsewhere. Differences in computing equipment across settings are likely to have had an impact on response times [12]. Second, additional variance may have been introduced by distractions in the home environment, in comparison with the formal lab-based testing environment. We attempted to monitor and control for distraction and found that distraction more strongly affected reaction time measures during web-based testing. At the same time, all 5 outliers excluded during the current analyses were obtained during web-based assessments. Missing data from 2 participants was due to additional errors during web-based testing on the SWM task, which precluded the accurate calculation of test performance scores. Susceptibility to distraction and resultant increases in variance of test outcome measures are important to bear in mind when considering web-based testing as a substitute for, or in addition to, in-person testing.

Limitations

The use of a healthy, relatively young, and highly educated sample may limit the generalization of findings to lesser-educated, clinical, or old-age samples. This research suggests that for the examined CANTAB performance indices, web-based assessments are likely to be a suitable alternative for similar samples. Further examination of the comparability of web-based assessment is now required in populations of clinical interest. In the longer-term, participants and patient groups with access restrictions may be the ones who benefit most from remote testing.

The study examined only the reliability of tests across settings and different devices, since all in-person tests were completed on touch screen iPads, and all web-based assessments on personal computers or laptops. Further research is required to examine whether reaction time data may be collected more consistently, where similar or the same devices are used across settings. Since the completion of this study, variance in workstation information is now routinely collected for CANTAB web-based tests, which allows for better determination of the effects of different workstations on test performance.

It is not clear how computer/device experience may have interacted with our results because we did not collect this information. However, our participants were recruited via Facebook, screened for inclusion online, and tested at home using their personal computing system, so it is likely that they had at least modest computer experience. Discrepancies between lab-based and web-based remote testing may be amplified for individuals with less computer experience, who may need to rely on the support of study staff to a greater extent.

The study was powered to detect moderate differences between test settings and was not adequately powered to identify subtle differences. Bayesian statistics were able to qualify the level of support for the null or alternate hypothesis, but much larger samples would be required to determine stronger evidence for the null hypothesis. Replication in a larger sample is now required to examine for the presence of any subtle differences between test settings.

Further work is now required to examine test-retest reliability for web-based assessments to identify whether test reliabilities are similar to those obtained during repeated in-person assessments. Our data show intersetting reliabilities, which are similar to previously reported test-retest reliabilities obtained during in-person assessments. Automated test scoring of performance indices, standardized across test administration and testing platforms, circumvent problems with rater-based variances in reliability. However, differences in computer hardware and software can impact reaction time data, and this must be borne in mind during web-based neuropsychological assessments.

Overview and Implications

This study compared web-based CANTAB tests with gold-standard in-person administered lab-based assessments. Performance indices obtained in person showed broad equivalence, good agreement, and significant linear relationships with those obtained during web-based assessments. Overall, this study provides evidence for the comparability of a range of performance outcome indices examined using web-based testing in a healthy adult sample. Certain performance indices showed better comparability than others and should therefore be preferable for use where comparability with typical in-person assessment is needed. Reaction time indices were not found to be comparable, and greater care is required in the interpretation of web-based latency results in relation to typical in-person assessments.

Acknowledgments

This study was financially supported by Cambridge Cognition, a digital health company specializing in computerized cognitive assessment, including CANTAB.

Conflicts of Interest

All authors are employed by Cambridge Cognition and have no other conflicts of interest to declare.

Multimedia Appendix 1
Comparison of bivariate (Spearman) correlation of test performance between settings, with partial correlations which covary for elapsed time (in days) between assessments and test retest reliabilities of relevant CANTAB performance indices from previously published research.
[DOC File , 58 KB - jmir_v22i8e16792_app1.doc]

References

1. Morrison GE, Simone CM, Ng NF, Hardy JL. Reliability and validity of the neurocognitive performance test, a web-based neuropsychological assessment. Front Psychol 2015;6:1652 [FREE Full text] [doi: 10.3389/fpsyg.2015.01652] [Medline: 26579035]
2. Haworth CM, Harlaar N, Kovas Y, Davis OS, Oliver BR, Hayiou-Thomas ME, et al. Internet cognitive testing of large samples needed in genetic research. Twin Res Hum Genet 2007 Aug;10(4):554-563. [doi: 10.1375/twin.10.4.554] [Medline: 17708696]
3. Feenstra HE, Vermeulen IE, Murre JM, Schagen SB. Online cognition: factors facilitating reliable online neuropsychological test results. Clin Neuropsychol 2017 Jan;31(1):59-84. [doi: 10.1080/13854046.2016.1190405] [Medline: 27266677]
4. Kraut R, Olson J, Banaji M, Bruckman A, Cohen J, Couper M. Psychological research online: report of board of scientific affairs' advisory group on the conduct of research on the internet. Am Psychol 2004;59(2):105-117. [doi: 10.1037/0003-066X.59.2.105] [Medline: 14992637]
5. Hansen T, Lehn H, Evensmoen H, Håberg A. Initial assessment of reliability of a self-administered web-based neuropsychological test battery. Comput Hum Behav 2016 Oct;63:91-97. [doi: 10.1016/j.chb.2016.05.025]
6. Barenboym DA, Wurm LH, Cano A. A comparison of stimulus ratings made online and in person: gender and method effects. Behav Res Methods 2010 Mar;42(1):273-285. [doi: 10.3758/BRM.42.1.273] [Medline: 20160306]
7. Gosling SD, Vazire S, Srivastava S, John OP. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. Am Psychol 2004;59(2):93-104. [doi: 10.1037/0003-066X.59.2.93] [Medline: 14992636]
8. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: joint position paper of the American academy of clinical neuropsychology and the national academy of neuropsychology. Clin Neuropsychol 2012;26(2):177-196 [FREE Full text] [doi: 10.1080/13854046.2012.663001] [Medline: 22394228]
9. Hoskins LL, Binder LM, Chaytor NS, Williamson DJ, Drane DL. Comparison of oral and computerized versions of the word memory test. Arch Clin Neuropsychol 2010 Nov;25(7):591-600. [doi: 10.1093/arclin/acq060] [Medline: 20719808]
10. Skitka LJ, Sargis EG. The internet as psychological laboratory. Annu Rev Psychol 2006;57:529-555. [doi: 10.1146/annurev.psych.57.102904.190048] [Medline: 16318606]
11. Schmand B. Why are neuropsychologists so reluctant to embrace modern assessment techniques? Clin Neuropsychol 2019 Feb;33(2):209-219. [doi: 10.1080/13854046.2018.1523468] [Medline: 30882285]
12. Parsons TD, McMahan T, Kane R. Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. Clin Neuropsychol 2018 Jan;32(1):16-41. [doi: 10.1080/13854046.2017.1337932] [Medline: 28590154]
13. Germine L, Nakayama K, Duchaine BC, Chabris CF, Chatterjee G, Wilmer JB. Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. Psychon Bull Rev 2012 Oct;19(5):847-857. [doi: 10.3758/s13423-012-0296-9] [Medline: 22829343]
14. Wilmer JB, Germine L, Chabris CF, Chatterjee G, Williams M, Loken E, et al. Human face recognition ability is specific and highly heritable. Proc Natl Acad Sci U S A 2010 Mar 16;107(11):5238-5241 [FREE Full text] [doi: 10.1073/pnas.0913053107] [Medline: 20176944]
15. Hansen TI, Haferstrom EC, Brunner JF, Lehn H, Håberg AK. Initial validation of a web-based self-administered neuropsychological test battery for older adults and seniors. J Clin Exp Neuropsychol 2015;37(6):581-594 [FREE Full text] [doi: 10.1080/13803395.2015.1038220] [Medline: 26009791]
16. Assmann KE, Bailet M, Lecoffre AC, Galan P, Hercberg S, Amieva H, et al. Comparison between a self-administered and supervised version of a web-based cognitive test battery: results from the nutrinet-santé cohort study. J Med Internet Res 2016 Apr 5;18(4):e68 [FREE Full text] [doi: 10.2196/jmir.4862] [Medline: 27049114]
17. Cromer JA, Harel BT, Yu K, Valadka JS, Brunwin JW, Crawford CD, et al. Comparison of cognitive performance on the cogstate brief battery when taken in-clinic, in-group, and unsupervised. Clin Neuropsychol 2015;29(4):542-558. [doi: 10.1080/13854046.2015.1054437] [Medline: 26165425]
18. Feenstra HE, Murre JM, Vermeulen IE, Kieffer JM, Schagen SB. Reliability and validity of a self-administered tool for online neuropsychological testing: the Amsterdam cognition scan. J Clin Exp Neuropsychol 2018 Apr;40(3):253-273. [doi: 10.1080/13803395.2017.1339017] [Medline: 28671504]
19. Silverstein SM, Berten S, Olson P, Paul R, Willams LM, Cooper N, et al. Development and validation of a world-wide-web-based neurocognitive assessment battery: WebNeuro. Behav Res Methods 2007 Nov;39(4):940-949. [doi: 10.3758/bf03192989] [Medline: 18183911]

20. Wild K, Howieson D, Webbe F, Seelye A, Kaye J. Status of computerized cognitive testing in aging: a systematic review. Alzheimers Dement 2008 Nov;4(6):428-437 [FREE Full text] [doi: 10.1016/j.jalz.2008.07.003] [Medline: 19012868]
21. CANTAB The most sensitive and validated cognitive research software available. Cambridge Cognition. URL: https://www.cambridgecognition.com/cantab [accessed 2020-07-13]
22. Barnett J, Blackwell A, Sahakian B, Robbins T. The paired associates learning (PAL) test: 30 years of CANTAB translational neuroscience from laboratory to bedside in dementia research. Curr Top Behav Neurosci 2016;28:449-474. [doi: 10.1007/7854_2015_5001] [Medline: 27646012]
23. Wykes T, Sturt E. The measurement of social behaviour in psychiatric patients: an assessment of the reliability and validity of the SBS schedule. Br J Psychiatry 1986 Jan;148:1-11. [doi: 10.1192/bjp.148.1.1] [Medline: 3082403]
24. Rathbone A, Shaw S, Kumbhare D. Package 'ICC.Sample.Size'. CRAN-R Project. 2015. URL: https://cran.r-project.org/web/packages/ICC.Sample.Size/ICC.Sample.Size.pdf [accessed 2020-07-09]
25. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.r-project.org/ [accessed 2020-07-13]
26. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. Stat Med 2012 Dec 20;31(29):3972-3981. [doi: 10.1002/sim.5466] [Medline: 22764084]
27. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994 Dec;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
28. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
29. Kühberger A, Fritz A, Scherndl T. Publication bias in psychology: a diagnosis based on the correlation between effect size and sample size. PLoS One 2014;9(9):e105825 [FREE Full text] [doi: 10.1371/journal.pone.0105825] [Medline: 25192357]
30. Richard FD, Bond CF, Stokes-Zoota JJ. One hundred years of social psychology quantitatively described. Rev Gen Psychol 2003 Dec;7(4):331-363. [doi: 10.1037/1089-2680.7.4.331]
31. Jarosz AF, Wiley J. What are the odds? A practical guide to computing and reporting Bayes factors. J Probl Solving 2014 Nov 7;7(1):9. [doi: 10.7771/1932-6246.1167]
32. SurveyMonkey. URL: https://www.surveymonkey.co.uk/ [accessed 2020-07-13]
33. Cognitive research: CANTAB Connect research. Cambridge Cognition. URL: https://www.cambridgecognition.com/products/cognitive-research [accessed 2020-07-09]
34. Web-based testing: conduct virtual research and test participants online. Cambridge Cognition. URL: https://www.cambridgecognition.com/products/cognitive-research/web-based-testing [accessed 2020-07-09]
35. Owen AM, Downes JJ, Sahakian BJ, Polkey CE, Robbins TW. Planning and spatial working memory following frontal lobe lesions in man. Neuropsychologia 1990;28(10):1021-1034. [doi: 10.1016/0028-3932(90)90137-d] [Medline: 2267054]
36. Owen AM, Sahakian BJ, Semple J, Polkey CE, Robbins TW. Visuo-spatial short-term recognition memory and learning after temporal lobe excisions, frontal lobe excisions or amygdalo-hippocampectomy in man. Neuropsychologia 1995 Jan;33(1):1-24. [doi: 10.1016/0028-3932(94)00098-a] [Medline: 7731533]
37. Rabbitt P, Lowe C. Patterns of cognitive ageing. Psychol Res 2000;63(3-4):308-316. [doi: 10.1007/s004269900009] [Medline: 11004884]
38. Ospina L, Shanahan M, Perez-Rodriguez M, Chan C, Clari R, Burdick K. Alexithymia predicts poorer social and everyday functioning in schizophrenia and bipolar disorder. Psychiatry Res 2019 Mar;273:218-226 [FREE Full text] [doi: 10.1016/j.psychres.2019.01.033] [Medline: 30658205]
39. Sahakian B, Jones G, Levy R, Gray J, Warburton D. The effects of nicotine on attention, information processing, and short-term memory in patients with dementia of the Alzheimer type. Br J Psychiatry 1989 Jun;154:797-800. [doi: 10.1192/bjp.154.6.797] [Medline: 2597885]
40. Jasp Team. URL: https://jasp-stats.org/ [accessed 2020-07-13]
41. Aguinis H, Gottfredson RK, Joo H. Best-practice recommendations for defining, identifying, and handling outliers. Organ Res Methods 2013 Jan 14;16(2):270-301. [doi: 10.1177/1094428112470848]
42. Ruano L, Sousa A, Severo M, Alves I, Colunas M, Barreto R, et al. Development of a self-administered web-based test for longitudinal cognitive assessment. Sci Rep 2016 Jan 8;6:19114 [FREE Full text] [doi: 10.1038/srep19114] [Medline: 26743329]
43. Vaz S, Falkmer T, Passmore AE, Parsons R, Andreou P. The case for using the repeatability coefficient when calculating test-retest reliability. PLoS One 2013;8(9):e73990 [FREE Full text] [doi: 10.1371/journal.pone.0073990] [Medline: 24040139]
44. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996 Jan;1(1):30-46 [FREE Full text] [doi: 10.1037/1082-989X.1.1.30]
45. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005 Mar;19(1):231-240. [doi: 10.1519/15184.1] [Medline: 15705040]
46. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]

47. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420] [Medline: 18839484]
48. Compute six intraclass correlation measures. SAS Institute Inc. URL: https://support.sas.com/kb/25/031.html#ref [accessed 2020-07-09]
49. Baayen R, Davidson D, Bates D. Mixed-effects modeling with crossed random effects for subjects and items. J Mem Lang 2008 Nov;59(4):390-412. [doi: 10.1016/j.jml.2007.12.005]
50. Mulder J, Wagenmakers E. Editors' introduction to the special issue 'Bayes factors for testing hypotheses in psychological research: practical relevance and new developments'. J Math Psychol 2016 Jun;72:1-5. [doi: 10.1016/j.jmp.2016.01.002]
51. Wagenmakers E, Love J, Marsman M, Jamil T, Ly A, Verhagen J, et al. Bayesian inference for psychology. Part II: example applications with JASP. Psychon Bull Rev 2018 Feb;25(1):58-76 [FREE Full text] [doi: 10.3758/s13423-017-1323-7] [Medline: 28685272]
52. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Mar 8;1(8476):307-310. [Medline: 2868172]
53. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999 Jun;8(2):135-160. [doi: 10.1177/096228029900800204] [Medline: 10501650]
54. Bland JM, Altman DG. The use of transformation when comparing two means. Br Med J 1996 May 4;312(7039):1153 [FREE Full text] [doi: 10.1136/bmj.312.7039.1153] [Medline: 8620137]
55. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015;25(2):141-151 [FREE Full text] [doi: 10.11613/BM.2015.015] [Medline: 26110027]
56. Salthouse TA, Nesselroade JR, Berish DE. Short-term variability in cognitive performance and the calibration of longitudinal change. J Gerontol B Psychol Sci Soc Sci 2006 May;61(3):P144-P151 [FREE Full text] [doi: 10.1093/geronb/61.3.p144] [Medline: 16670183]
57. Calamia M, Markon K, Tranel D. The robust reliability of neuropsychological measures: meta-analyses of test-retest correlations. Clin Neuropsychol 2013;27(7):1077-1105. [doi: 10.1080/13854046.2013.809795] [Medline: 24016131]
58. Calamia M, Markon K, Tranel D. Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. Clin Neuropsychol 2012;26(4):543-570. [doi: 10.1080/13854046.2012.680913] [Medline: 22540222]
59. Lowe C, Rabbitt P. Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: theoretical and practical issues. Cambridge neuropsychological test automated battery. International study of post-operative cognitive dysfunction. Neuropsychologia 1998 Sep;36(9):915-923. [doi: 10.1016/s0028-3932(98)00036-0] [Medline: 9740364]

Abbreviations

CANTAB: Cambridge Neuropsychological Test Automated Battery
ERT: emotion recognition task
ICC: intraclass correlation
OTS: one touch stockings of Cambridge
PAL: paired associate learning
PRM: pattern recognition memory
RVP: rapid visual information processing
SWM: spatial working memory

Edited by G Eysenbach; submitted 29.10.19; peer-reviewed by H Feenstra, F Lanfranchi; comments to author 20.02.20; revised version received 15.05.20; accepted 11.06.20; published 04.08.20.

Please cite as:
Backx R, Skirrow C, Dente P, Barnett JH, Cormack FK
Comparing Web-Based and Lab-Based Cognitive Assessment Using the Cambridge Neuropsychological Test Automated Battery: A Within-Subjects Counterbalanced Study
J Med Internet Res 2020;22(8):e16792
URL: https://www.jmir.org/2020/8/e16792
doi:10.2196/16792
PMID:32749999

Original Paper

Augmented Reality System for Digital Rectal Examination Training and Assessment: System Validation

Theerapat Muangpoon[1], MSc; Reza Haghighi Osgouei[1], PhD; David Escobar-Castillejos[1], PhD; Christos Kontovounisios[1], MD, PhD, FACS, FRCSE; Fernando Bello[1], PhD

Faculty of Medicine, Department of Surgery and Cancer, Imperial College London, London, United Kingdom

Corresponding Author:
Reza Haghighi Osgouei, PhD
Faculty of Medicine, Department of Surgery and Cancer
Imperial College London
Rm H38, ICCESS, 3rd Floor, Chelsea and Westminster Hospital
369 Fulham Road
London, SW10 9NH
United Kingdom
Phone: 44 7940419331
Email: r.haghighi-osgouei@imperial.ac.uk

Abstract

Background: Digital rectal examination is a difficult examination to learn and teach because of limited opportunities for practice; however, the main challenge is that students and tutors cannot see the finger when it is palpating the anal canal and prostate gland inside the patients.

Objective: This paper presents an augmented reality system to be used with benchtop models commonly available in medical schools with the aim of addressing the problem of lack of visualization. The system enables visualization of the examining finger, as well as of the internal organs when performing digital rectal examinations. Magnetic tracking sensors are used to track the movement of the finger, and a pressure sensor is used to monitor the applied pressure. By overlaying a virtual finger on the real finger and a virtual model on the benchtop model, students can see through the examination and finger maneuvers.

Methods: The system was implemented in the Unity game engine (Unity Technologies) and uses a first-generation HoloLens (Microsoft Inc) as an augmented reality device. To evaluate the system, 19 participants (9 clinicians who routinely performed digital rectal examinations and 10 medical students) were asked to use the system and answer 12 questions regarding the usefulness of the system.

Results: The system showed the movement of an examining finger in real time with a frame rate of 60 fps on the HoloLens and accurately aligned the virtual and real models with a mean error of 3.9 mm. Users found the movement of the finger was realistic (mean 3.9, SD 1.2); moreover, they found the visualization of the finger and internal organs were useful for teaching, learning, and assessment of digital rectal examinations (finger: mean 4.1, SD 1.1; organs: mean 4.6, SD 0.8), mainly targeting a novice group.

Conclusions: The proposed augmented reality system was designed to improve teaching and learning of digital rectal examination skills by providing visualization of the finger and internal organs. The initial user study proved its applicability and usefulness.

(J Med Internet Res 2020;22(8):e18637) doi:10.2196/18637

KEYWORDS

Augmented Reality; Digital Rectal Examination (DRE); Magnetic Tracker; Pressure Sensor; Medical Education; Usability

Introduction

Digital rectal examination (DRE) is a physical examination for detecting rectal and prostate abnormalities. Focusing on prostate examination, DRE requires that an index finger to palpate the prostate gland to detect abnormalities in gland size, tenderness, and surface texture. Even though it is a recommendation of the American Cancer Society to perform DRE to screen and detect prostate cancer in patients with colorectal symptoms, multiple studies [1,2] have found that, during their final year, medical students are not confident in their abilities to perform the examination. The lack of confidence in performing DRE among

medical students has mainly been attributed to not having adequate practice in medical schools [1].

A standard method to train DRE skills is to practice on a benchtop model. This model is a plastic human mannequin with skin-like rubber to represent the rectal canal with several plastic replaceable prostate models (Figure 1). Even though students can touch and feel the prostate gland through the rubber rectum, no visualization of finger movement or internal organs can be obtained because it lacks transparency. Similarly, this model does not provide enough information to examiners to assess the techniques used by students to perform DRE.

There have been several attempts to improve the visualization of DRE on a benchtop model. Early attempts include a training system using virtual reality technology, together with a Phantom haptic device [3] which displayed a simplified 3D model of kidneys, rectum, bladder, and prostate along with the virtual representation of the examining finger on the 2D monitor. This system was evaluated against the rubber model, and it was concluded that it could be a new way to train DRE if the realism of the haptic system was improved [3]. In another study [4], a similar approach used a haptic interface for palpating the prostate gland. In follow-up work and to improve the design of haptic-based learning tools, Granados et al [5] conducted a study to better understand palpation techniques of experts while conducting DRE on a real subject. Dobson et al [6] proposed a system using virtual reality technology to visualize the anatomy in the pelvic area [6]. For their system, the user had to wear special glasses to view the model in 3D which was displayed on the 67-inch×50-inch screen (VR ImmersaDesk). It was shown that the system helped medical students in gaining more understanding of the anatomy and results in better exam scores [6]. Rissanen et al [7] introduced Annotated Simulation Records for DRE, which focused on using virtual reality technology to reveal useful data from the sensor during DRE practice. In this system, the urologist selected the most useful parameters to be annotated to help in teaching DRE. This system was evaluated by medical students, and it was found that the numerical annotations helped them learn faster than verbal feedback. Balkissoon et al [8] introduced a DRE training system that consisted of a physical benchtop model and a 2D screen for visualizing the DRE in which multiple sensors were attached to the prostate gland in the benchtop model to measure user applied pressure on various locations. Visual information, including applied force, palpated area, and palpation at each location, was displayed during the examination. The study [8] showed that the sensors could help the instructor to observe and assess the performance of medical students performing DREs. In another model [9], a similar concept, embedding pressure sensors in the model, was followed.

Displaying information, such as finger position and pressure on a 2D screen, was demonstrated to be beneficial in understanding and performing DRE; however, the user experience was not ideal due to the lack of colocation between the benchtop model and the display.

In this paper, we present an augmented reality system for DRE visualization. It uses sensors attached to the examining finger to track its maneuvers and to monitor applied pressure. It also displays the DRE and essential information overlaid on the real benchtop model using an augmented reality device. The main goal was to improve the user experience of a widely available benchtop model. The paper first describes the visualization of the examining finger and internal organs, the step-by-step guidance for DRE, and the performance recording feature. Results, including performance measures and feedback from clinicians and medical students, are then reported, followed by discussion and conclusions.

Figure 1. A standard benchtop model used in medical schools for teaching and practicing digital rectal examination.

Methods

Hardware and Software

The proposed augmented reality system is used as an extension to a standard Rectal Examination Trainer (Mk2, Limbs & Things Inc) benchtop model (Figure 1, [10]). The model is a semirealistic representation of the buttocks, anus, and rectum, allowing for the practice of diagnostic skills associated with rectal examination. It includes additional rectal examination perineum, which contains two rectal pathologies (polyp and carcinoma). For prostate examination, five interchangeable prostates are provided: normal, bilateral benign, unilateral benign, bilateral carcinoma, and unilateral carcinoma. The model can be used for both digital examination of prostate and rectum, as well as for the insertion and use of anoscope and proctoscope.

We used the HoloLens (version 1, Microsoft Inc) as an augmented reality head-mounted display (Figure 2). The

HoloLens is immersive and see-through to help the user perceive the environment as realistic. It also enables interactions with holographic content. The HoloLens features an inertial measurement unit (accelerometer, gyroscope, and magnetometer), 4 environment understanding sensors (2 on each side), an energy-efficient depth camera with a 120°×120° angle of view, a 2.4-megapixel photographic video camera, a 4-microphone array, and an ambient light sensor [11].

HoloLens has been used in various medical simulations; VimedixAR (CAE Inc) was the first ultrasound simulator to integrate a HoloLens [12]; with this system, health care professionals manipulate representations of realistic anatomical parts and view the ultrasound beam in real time as it passes through human anatomy. With CAE LucinaAR, clinical learners can view 3D holograms of a fetus, as it descends the birth canal, and learn to manage complex deliveries [13]. HoloPatient (Microsoft Inc) is a mixed-reality learning tool for nursing, allied health, and medical schools that delivers simulated patient experiences [14]. Learning Heart (Spheregen) is a HoloLens application that assists students in understanding the physiology of the heart [15].

The 3D software in our system was developed using the Unity games engine (version 2018.3.7f1; Unity Technologies) [16] cross-platform authoring tool for creating 3D content.

Figure 2. Microsoft HoloLens, a head-mounted augmented reality device to display 3D virtual objects.

Real-Time Performance Visualization

To track and show the movement of the examining finger inside the benchtop model during the examination, we used a Trakstar magnetic tracking system (Northern Digital Inc) to obtain the position and orientation (pose) of the examining finger in real time, due to its ability to operate without line-of-sight [17]. It consists of a midrange magnetic field transmitter and a 6 degrees-of-freedom receiver (model 180), and it was connected to an electronics unit for amplification and digitization. The 6 degrees-of-freedom sensor has a position accuracy of 1.40 mm RMS and orientation accuracy of 0.50° RMS. The combination of the transmitter and the receiver allows tracking within a 30×40×30 cm^3 zone, which is large enough for tracking the examining finger inside the benchtop model. The sensor was attached to the finger using thin tape (Figure 3, [17]). Sensor data were read at 40 Hz by the computer via an API using a previously developed C++ plug-in [18]. Once read, sensor data was transferred to the HoloLens via Wi-Fi. Position and orientation were transformed into the real-world coordinate system with the help of Vuforia Engine (version 8.1.7) [19]. The transformation was achieved by using the HoloLens to track the pose of the image target in world coordinates (Figure 4). This transformation enabled synchronization and overlay of the virtual finger on the tracked real finger, and for the result to be seen through the HoloLens (Figure 5).

Figure 3. A Trakstar magnetic positioning sensor attached to the examining finger.

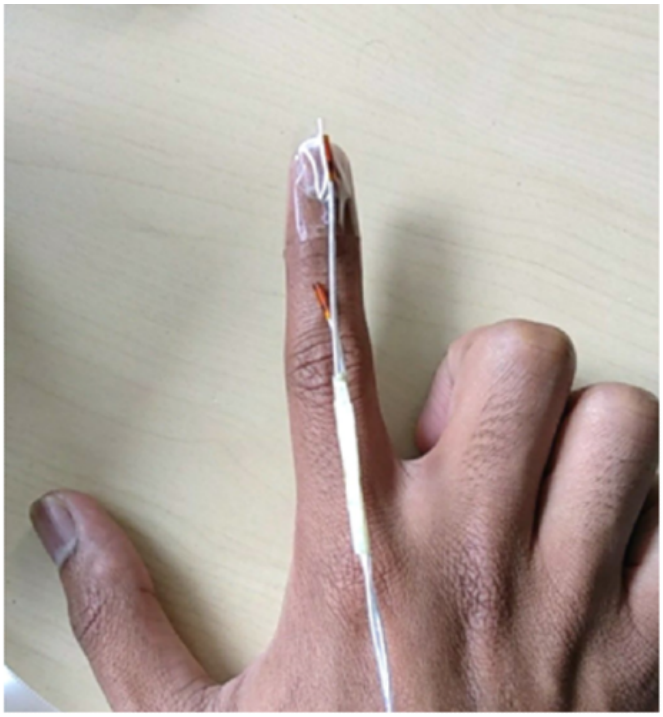


Figure 4. Conversion process of the position tracking coordinates to real-world (HoloLens) coordinates using Vuforia image tracking.

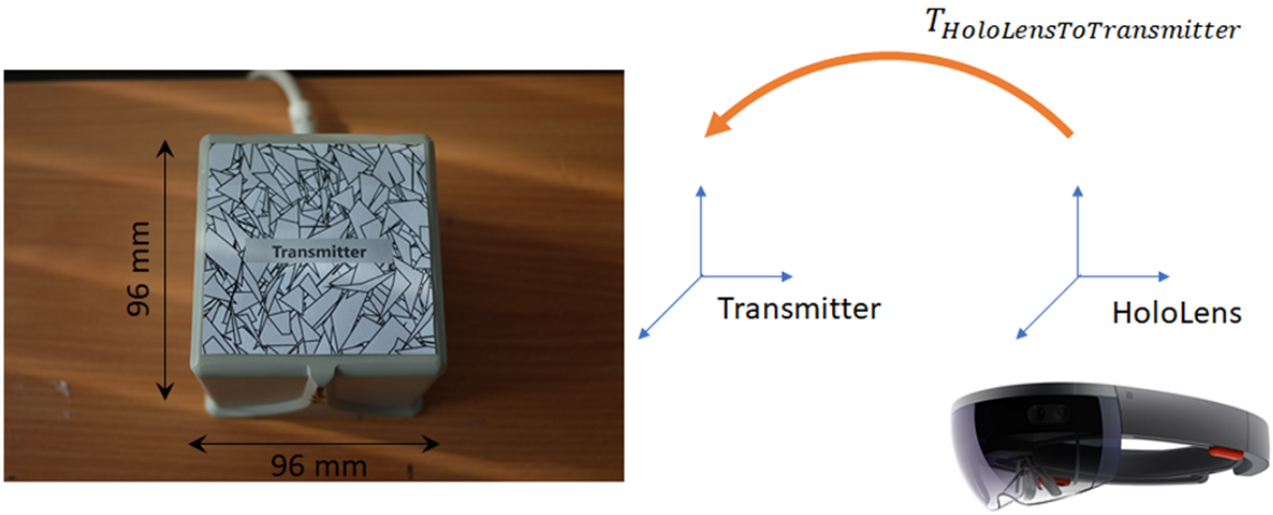


Figure 5. The virtual finger in blue overlaid onto the real examining finger with a blue medical glove.

To visualize the internal components of the benchtop model and to overlay the relevant virtual anatomy, it was necessary to first align the virtual model (Figure 6) with the physical benchtop model. For this purpose, the iterative closest point algorithm was used [20]; it takes the position of 7 anatomical landmarks as an input (Figure 7) and yields a transformation matrix. This matrix is used to rotate and translate the virtual benchtop model and virtual internal organs to align with the physical benchtop model (Figure 8). Once aligned, the user can visualize the movement of the examining finger and the model by directly looking at the model with the HoloLens (Figure 9).

Figure 6. 3D virtual benchtop model.

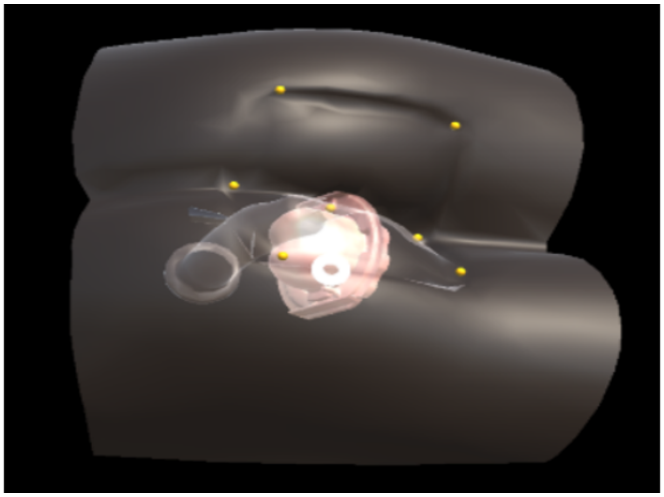


Figure 7. Red landmarks on a benchtop model used as inputs to the iterative closest point algorithm.

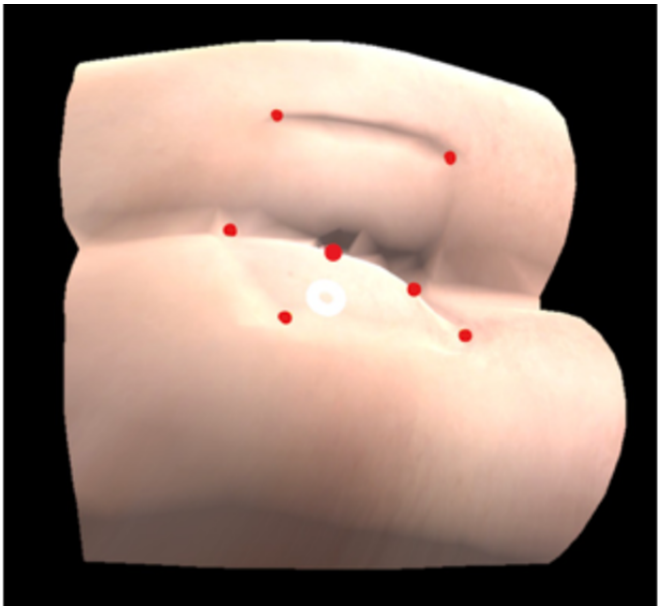


Figure 8. A virtual transparent benchtop model with a 3D virtual prostate inside is aligned with the physical benchtop model.

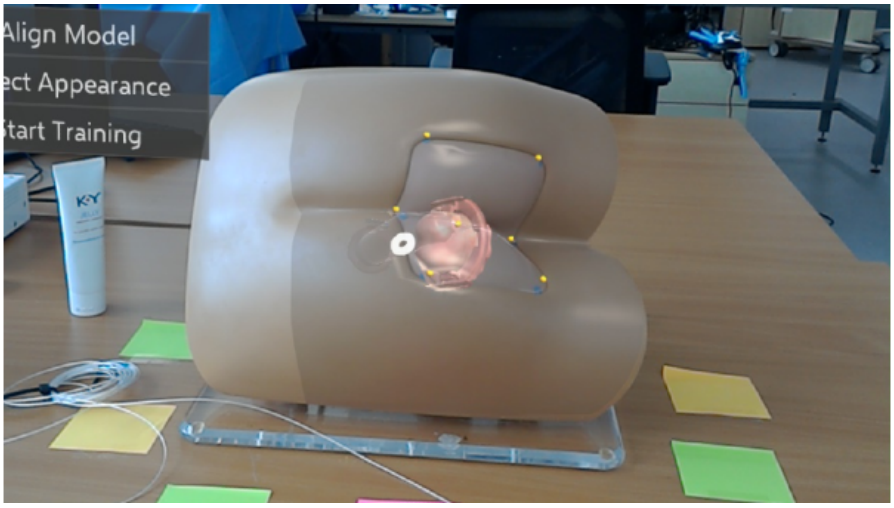
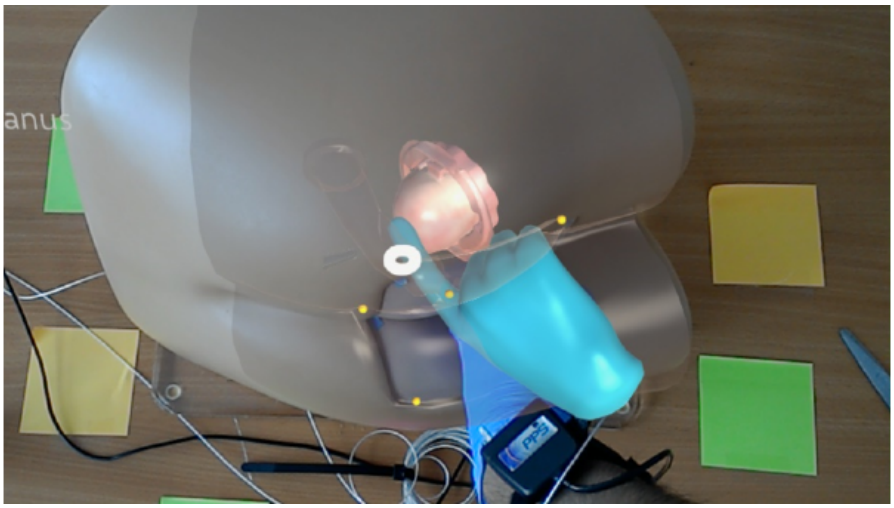


Figure 9. Real-time visualization of digital rectal examination using the proposed augmented reality system.



A FingerTPS force sensor [21] was also attached to the examining finger to estimate the pressure applied to the prostate during palpation. The force sensor is flat and thin and can be worn under a surgical glove (Figure 10, [21]). The force sensor data were transferred to the HoloLens using the same process as for the pose information from the tracking sensor. A force

visualizer, represented as a color bar with 3 regions showing different levels of pressure applied to the prostate gland, was displayed to provide real-time feedback (Figure 11).

Figure 10. A FingerTPS pressure sensor. It measures pressure applied during the prostate palpation.

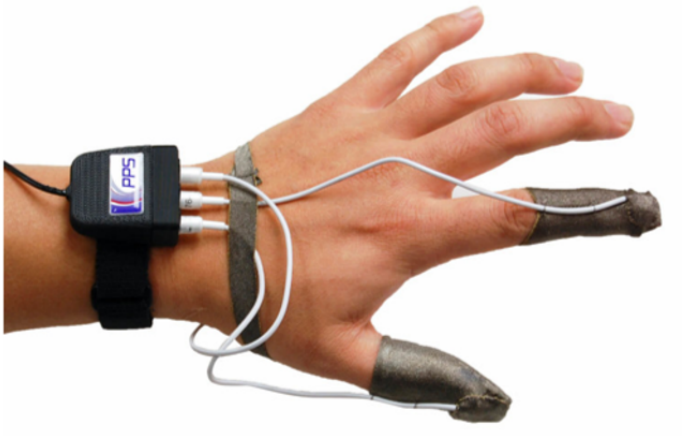
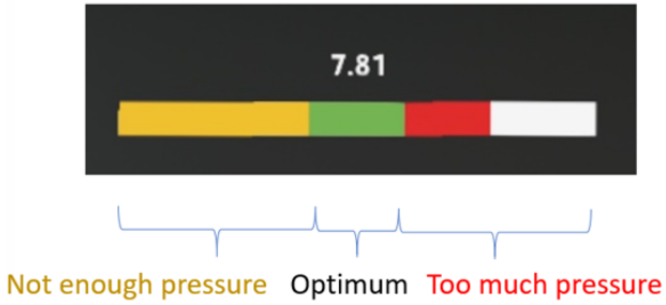


Figure 11. Real-time visualization of force applied by the examining finger.

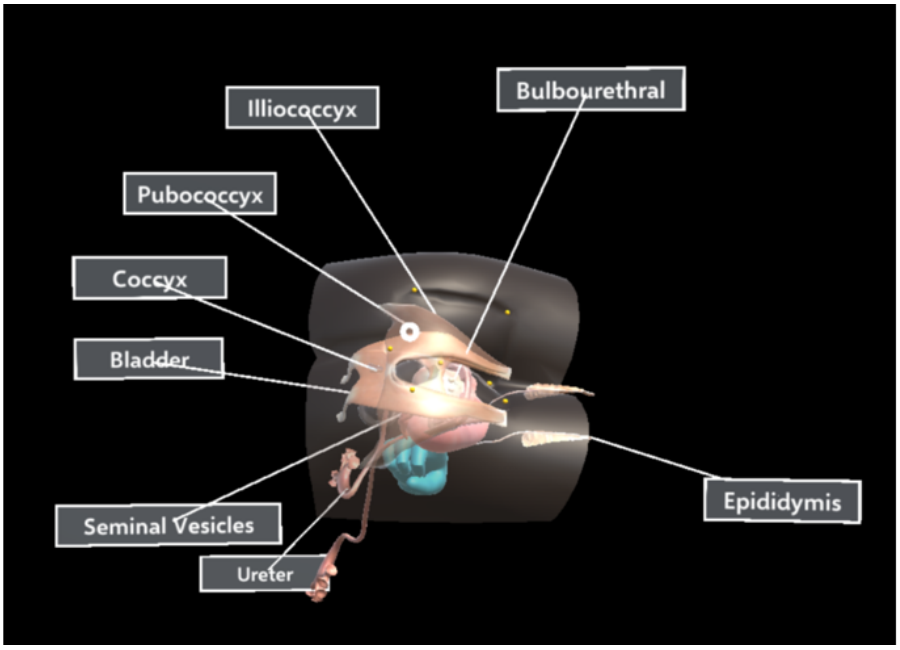


Internal Organ Labeling

Labels were used for displaying relevant information regarding internal anatomy inside the benchtop model; however, since there were several anatomical structures inside a small area, this made it difficult to label the anatomy in an effective way (in a way that avoids overlap with other labels or with the anatomy). Also, since our system allowed users to look at the anatomy from different perspectives, labels needed to be dynamically positioned. To address these issues, we implemented a view management system, capable of resolving occlusion among labels. Our labeling system used the labeling technique suggested by Tatzgern [22] (overcoming occlusion by limiting update rates, ie, not continuously moving the labels to separate them, but only when they occlude each other) combined with the theory by Hartzmann [23] to ensure that the labels were always near the object and that the lines from labels to objects did not cross each other. To achieve these criteria, the following steps are iteratively applied until all occlusions are resolved: (1) Place the label near the referenced object on the surface of the sphere, centered at the center of mass of the benchtop model. (2) Iterate through all labels to find all those that may be occluded by another label, establishing the side of the occlusion. (3) Move the occluded label to the opposite side to resolve the occlusion.

The labels were created using a ToolTip component of Unity, provided by the Microsoft Mixed Reality Toolkit (version 2.0.0 RC2) [24]. This component can rotate the label to always face the user, making it more readable (Figure 12).

Figure 12. Occlusion between labels is resolved by pushing them apart along the circumference of the sphere, whose center is at the center of mass of the benchtop model.



Step-By-Step Guidance

Our augmented reality system incorporated step-by-step guidance to help trainees follow the correct steps and trajectory during the examination. Steps were extracted from the cognitive task analysis study performed by Low-Beer et al [25]. The system automatically tracked the position of the finger inside the rectal canal and checked whether the trainee had correctly followed the step. The user interface was designed to be readable and to require the least possible head movement to see all the contents (Figure 13), despite the narrow field of view of the HoloLens display.

Figure 13. Step-by-step guidance next to the virtual benchtop model. DRE: digital rectal examination.

Performance Recording

In addition to step-by-step guidance, our augmented reality system also allowed recording and playback of the examination. The pose of the examining finger and the benchtop model were recorded so that, when played back, the virtual examining finger and the virtual benchtop model could be accurately displayed on the HoloLens. This feature is useful trainees as they can observe experts repeatedly, or they can analyze their own performance. It can also be used by an examiner to assess student performance from different angles.

Model Alignment Evaluation

The accuracy of the model alignment system was evaluated by performing an alignment task five times by an experienced user. After each alignment task was performed, the positions of each landmark on both the physical and virtual benchtop models were then measured using the magnetic tracking sensor. The error of the alignment system was then calculated as a distance

between a point on the virtual model and a corresponding point on the physical model.

Pilot User Study

An initial validation study was conducted. Clinicians (n=9; 6 men, 3 women; mean 37.8, SD 5.4 years of age) who routinely perform DRE, and medical students (n=10; 7 men, 3 women; mean 21.8, SD 2.4 years of age) were recruited. The study was approved by the National Health Service Patient Safety Agency Research Ethics Committee (09/H0701/68). Before the study, a consent form was signed by each participant. During the study, participants were asked to perform DREs on the benchtop model with the augmented reality system, wearing both the position tracking sensor and the pressure sensor under a standard surgical glove. They were also asked to pay attention to the information displayed on the HoloLens, such as the force bar and the guidance panel. Once finished, they were asked to watch the recorded performance. Afterward, participants were asked to complete an online questionnaire using a 5-point Likert scale from 1 (definitely disagree) to 5 (definitely agree) regarding the usefulness of the system (Table 1).

Table 1. Questions assessing the usefulness of the system.

Number	Question
Q1	The record of the expert's performance would be useful for DRE[a] teaching or learning
Q2	The movement of the examining finger would be useful for DRE training
Q3	Being able to visualize the internal organs in the benchtop model could help a trainee better understand DRE
Q4	The real-time visualization of force applied to the model would be useful for DRE training
Q5	The step-by-step guidance would be useful for DRE training
Q6	The movement of the examining finger inside the model is realistic and accurate
Q7	The virtual representation of benchtop model can be aligned accurately on the physical model
Q8	The AR[b] system is easy to use and understand
Q9	The AR system requires minimum movement to operate
Q10	During the performance, you feel tired or fatigued
Q11	The AR system can enhance current teaching and learning of DRE
Q12	I would recommend the AR system to be integrated into the medical curriculum

[a]DRE: digital rectal examination.

[b]AR: augmented reality.

Results

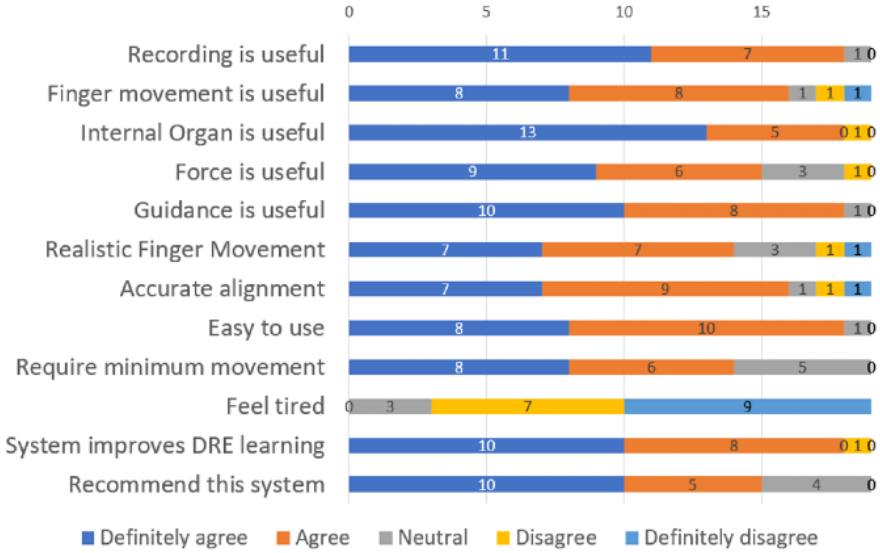
The augmented reality prototype system ran stably and achieved a frame rate of 60 fps, which is the highest possible frame rate of the HoloLens and Unity. The alignment between the virtual and real fingers and between the virtual and benchtop models was acceptable when visually inspected. The average alignment error of each landmark was an overall average of 1.73, 2.91, and 1.91 mm in the x, y, and z directions, respectively, or a root mean square of 3.9 mm (Table 2).

The usefulness of the system was assessed from the answers given to the questionnaire (Figure 14). Most participants would recommend this system to be integrated into medical school curriculum (mean 4.3, SD 0.8).

Table 2. Model alignment error.

Landmark	Alignment error (mm)		
	x, mean (SD)	y, mean (SD)	z, mean (SD)
1	0.46 (0.72)	1.96 (2.49)	0.48 (0.34)
2	2.38 (0.29)	2.24 (0.39)	2.02 (0.25)
3	2.04 (0.25)	0.78 (0.29)	4.38 (0.86)
4	1.02 (0.45)	5.68 (1.85)	3.06 (0.62)
5	1.18 (0.62)	6.40 (0.32)	0.96 (0.35)
6	1.90 (0.61)	1.20 (0.40)	1.92 (0.50)
7	2.98 (0.67)	2.12 (0.27)	0.52 (0.66)
Mean	1.73 (0.95)	2.91 (2.37)	1.91 (1.72)

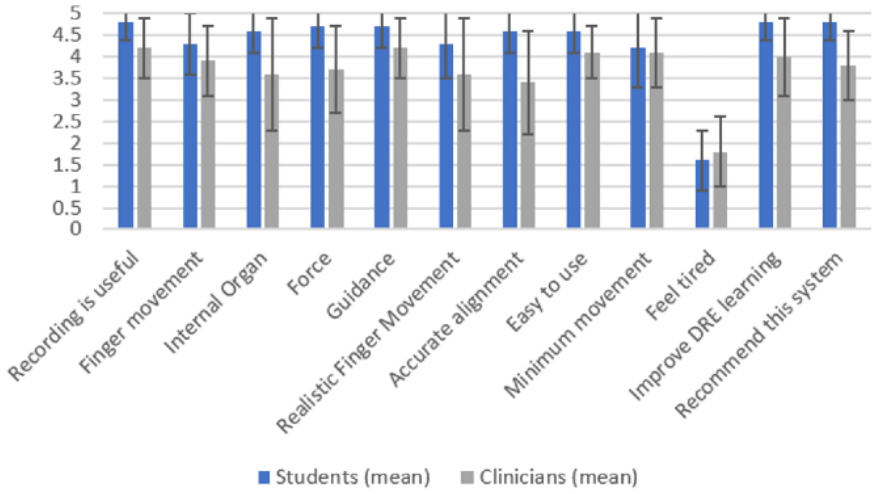
Figure 14. Results from the Likert-scale questionnaire. The numbers and hence the length of each bar indicate the number of participants choosing that rate. DRE: digital rectal examination.



According to the results, some features were more useful than others. For example, most participants agreed that performance recording was useful (mean 4.5, SD 0.6). Regarding the real-time feedback feature, most participants agreed that the visualization of an examining finger was useful (mean 4.1, SD 1.1). The highest score was achieved for the usefulness of the visualization of internal organs (mean 4.6, SD 0.8). Step-by-step guidance was also one of the most highly rated features (mean 4.5, SD 0.6). Regarding the usability of the system, participants responded that it was easy to use (mean 4.4, SD 0.6), and they did not feel fatigued after using the system (mean 1.7, SD 0.7). The alignment of the virtual benchtop model with the real benchtop model was also rated as very accurate (mean 4.1, SD 1.1).

The results from both groups, medical students and clinicians, are given in Figure 15. The scores obtained from the students were higher than those from the clinicians for all features (mean 0.63, SD 0.41).

Figure 15. Comparison between questionnaire results from medical students and clinicians (1 = definitely disagree, 5 = definitely agree) regarding the usefulness of the augmented reality system. DRE: digital rectal examination.



Discussion

Principal Findings

The results showed that medical students and clinicians were interested in the system and recognized the value of visualization of DRE, performance recording, and step-by-step guidance in improving the learning and teaching of DRE skills. Medical students could visualize the movement of the finger and the pressure applied to the prostate. This visualization and sensor data, combined with the step-by-step guidance, allowed them to receive feedback in real time while performing DRE. Using the record-and-playback feature, students could not only rewatch examinations performed by an expert, but examiners could also review student performance from multiple viewing angles.

Most users reported that the HoloLens and the tracking sensors were comfortable to wear. Even with a narrow field of view, the user interface did not appear cluttered, facilitating navigation through the menu and visualization of the performance, while at the same time facilitating step-by-step guidance on the left-hand side of the display. The model alignment was also perceived as accurate which demonstrated the use of the iterative

closest point algorithm, and the readings from the electromagnetic positioning sensor were valid.

User experience was improved by displaying all information on a head-mounted augmented reality display that also allowed the overlaying of the virtual finger on the real finger. This colocation allowed trainees to avoid having to change visual focus from the benchtop model to a separate display.

When comparing the feedback from medical students to the feedback from clinicians, it was observed that the mean ratings from medical students were higher than those from clinicians for all questions, with an average difference of 20%. A likely explanation is that experienced clinicians had higher expectations, and they were comparing the quality of the simulation with that of real cases because of their experience performing DREs. Having less experience, students would be more likely to require visualization and guidance, while clinicians would not because tactile feedback was adequate, given their level of experience. In addition, it may also reflect the increased acceptability and interest in this type of technology by younger generations.

While the proposed system was demonstrated to be beneficial, it has some limitations. Wearing and adjusting the HoloLens properly can take some time and require assistance. In addition, operating and interacting with the HoloLens by using gestures such as pinching, also requires practice. Regarding the sensors, properly wearing the tracking and pressure sensors is crucial, and at the moment, requires the presence of an assistant to be properly placed on the examining finger. In terms of implementation, the current alignment between the virtual and the physical benchtop model was done manually through 7 landmarks; however, automatic alignment would be faster, more accurate, and more convenient. With respect to features, a scoring system with real-time feedback would be valuable for teaching and assessing DRE skills. Apart from these, the benchtop model itself was reported to be much stiffer than real patients and generally unrealistic with limited anatomical landmarks.

Conclusions

This paper presents an augmented reality system for teaching and learning DRE that can be used with widely available benchtop models. It was designed to assist both trainees and teachers in learning, teaching, and practicing DRE, as well as to allow examiners to assess student technique. With colocation of the virtual and real models, students only need to focus on the benchtop model to visualize all important information. Even though the results from the user study are positive, further research is needed to evaluate the system. This would include more robust quantitative analysis with a larger number of participants and varying levels of experience. The augmented reality system could be improved by using the second version of the HoloLens which offers a larger field of view and resolution, and it could also be improved by using a haptic-based instead of standard benchtop model so that a wider variety of prostate glands and abnormalities may be generated and used during the examination. Such a model would also be able to directly track finger movement and estimate pressure applied without the need for external sensors. Finally, formative and summative assessment of DRE performance will be an important component of the next version of our system.

Acknowledgments

This work was partly funded by a Cancer Research UK Early Detection Project—Improving early detection of colorectal cancer through an innovative highly realistic simulation environment (award reference: C63674/A26161). The authors also would like to acknowledge and express their gratitude to Mr Agni Lahiri and Dr Neerav Joshi for their assistance during the pilot user study.

Conflicts of Interest

None declared.

References

1. Popadiuk C, Pottle M, Curran V. Teaching digital rectal examinations to medical students. Academic Medicine 2002;77(11):1140-1146. [doi: 10.1097/00001888-200211000-00017]
2. Wolf AMD, Wender R, Etzioni R, Thompson I, D'Amico A, Volk R, American Cancer Society Prostate Cancer Advisory Committee. American Cancer Society guideline for the early detection of prostate cancer: update 2010. CA Cancer J Clin 2010;60(2):70-98 [FREE Full text] [doi: 10.3322/caac.20066] [Medline: 20200110]
3. Burdea G, Patounakis G, Popescu V, Weiss R. Virtual reality-based training for the diagnosis of prostate cancer. IEEE Trans Biomed Eng 1999;46(10):1253-1260. [doi: 10.1109/10.790503]
4. Granados A, Mayer E, Norton C, Ellis D, Mobasheri M, Low-Beer N, et al. Haptics modelling for digital rectal examinations. : Springer, Cham; 2014 Presented at: International Symposium on Biomedical Simulation; 2014 Oct 16; Strasbourg, France p. 40-49. [doi: 10.1007/978-3-319-12057-7_5]
5. Granados A, Cox S, Low-Beer N, Higham J, Kneebone R, Bello F. Taking the pressure off the patient - understanding digital rectal examinations on a real subject. IEEE Trans Biomed Eng 2020:1-1. [doi: 10.1109/tbme.2020.2971826]
6. Dobson HD, Pearl RK, Orsay CP, Rasmussen M, Evenhouse R, Ai Z, et al. Virtual reality. Diseases of the Colon & Rectum 2003;46(3):349-352. [doi: 10.1007/s10350-004-6554-9]

7. Rissanen M, Kume N, Kuroda Y, Kuroda T, Yoshimura K, Yoshihara H. Asynchronous teaching of psychomotor skills through vr annotations: evaluation in digital rectal examination. Studies in Health Technology and Informatics 2008;132(16):411-416 [FREE Full text]
8. Balkissoon R, Blossfield K, Salud L, Ford D, Pugh C. Lost in translation: unfolding medical students' misconceptions of how to perform a clinical digital rectal examination. The American Journal of Surgery 2009 Apr;197(4):525-532. [doi: 10.1016/j.amjsurg.2008.11.025]
9. Wang W, Gerling GJ, Childress RM, Martin ML. Quantifying palpation techniques in relation to performance in a clinical prostate exam. IEEE Trans Inform Technol Biomed 2010 Jul;14(4):1088-1097. [doi: 10.1109/titb.2010.2041064]
10. Male rectal examination trainer - standard. Limbs & Things. URL: https://limbsandthings.com/uk/products/60170/60170-male-rectal-examination-trainer-standard [accessed 2020-07-17]
11. Microsoft Hololens. Hololens Gen 1. URL: https://www.wevolver.com/wevolver.staff/hololens.gen.1 [accessed 2020-07-17]
12. VimedixAR ultrasound simulator. URL: https://caehealthcare.com/hololens/ [accessed 2020-07-17]
13. CAE Lucina - Validated High-Fidelity Maternal/Fetal Training. URL: https://caehealthcare.com/patient-simulation/lucina/ [accessed 2020-07-17]
14. Introducing HoloPatient. URL: https://www.pearson.com/uk/web/pearsontq/news-and-blogs/2017/04/introducing-holopatient.html [accessed 2020-07-17]
15. SphereGen's 'Learning Heart' - Ted Dinsmore AR in ACTION. URL: https://www.youtube.com/watch?v=fSz2Ce1YcQ0 [accessed 2020-07-17]
16. Unreal Real-Time Development Platform. URL: https://unity.com/ [accessed 2020-07-17]
17. 3D Guidance driveBAY and trakSTAR. NDI Measurement Sciences. URL: http://www.ndigital.com/msci/products/drivebay-trakstar [accessed 2020-07-17]
18. Grandos A, Perhác J, Rosby L, Lee Y, Tan G, Tan T, et al. See-through visualisation for training and assessing unsighted physical examinations. 2017 Presented at: 13th Workshop on Virtual Reality Interactions and Physical Simulations; 2017 Apr 23; Lyon, France p. 85-92. [doi: 10.2312/vriphys.20171087]
19. Vuforia Engine. URL: https://developer.vuforia.com/ [accessed 2020-07-17]
20. Besl P, McKay N. A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell 1992 Feb;14(2):239-256. [doi: 10.1109/34.121791]
21. Finger TPS - Finger And Hand Pressure Measurement. URL: https://pressureprofile.com/body-pressure-mapping/finger-tps [accessed 2020-07-17]
22. Tatzgern M, Kalkofen D, Grasset R, Schmalstieg D. Hedgehog labeling: view management techniques for external labels in 3D space. 2014 Presented at: IEEE Virtual Reality (VR); 2014 Mar 29; Minneapolis, USA p. 27-32. [doi: 10.1109/VR.2014.6802046]
23. Hartmann K, Götzelmann T, Ali K, Strothotte T. Metrics for functional and aesthetic label layouts. 2005 Presented at: International Symposium on Smart Graphics; 2005 Aug 22; Frauenwörth, Germany p. 115-126. [doi: 10.1007/11536482_10]
24. Mixed Reality Toolkit (MRTK). URL: https://microsoft.github.io/MixedRealityToolkit-Unity/README.html [accessed 2020-07-17]
25. Low-Beer N, Kinnison T, Baillie S, Bello F, Kneebone R, Higham J. Hidden practice revealed: using task analysis and novel simulator design to evaluate the teaching of digital rectal examination. The American Journal of Surgery 2011 Jan;201(1):46-53. [doi: 10.1016/j.amjsurg.2010.09.004]

Abbreviations

DRE: digital rectal examination
RMS: root mean square

Edited by G Eysenbach; submitted 09.03.20; peer-reviewed by C Friedrich, M Aksoy; comments to author 18.04.20; revised version received 09.05.20; accepted 13.05.20; published 13.08.20.

Please cite as:
Muangpoon T, Haghighi Osgouei R, Escobar-Castillejos D, Kontovounisios C, Bello F
Augmented Reality System for Digital Rectal Examination Training and Assessment: System Validation
J Med Internet Res 2020;22(8):e18637
URL: https://www.jmir.org/2020/8/e18637
doi:10.2196/18637
PMID:32788146

Original Paper

Perceptions of Digital Health Education Among European Medical Students: Mixed Methods Survey

Felix Machleid[1,2*], BA, MD; Robert Kaczmarczyk[1,2], MD; Doreen Johann[3], BA; Justinas Balčiūnas[1,4]; Beatriz Atienza-Carbonell[1,5], MD; Finn von Maltzahn[1,6]; Lina Mosch[1,7*]

[1]Association Européenne des Étudiants en Médecine (EMSA), Brussels, Belgium

[2]School of Medicine, Technical University of Munich, Munich, Germany

[3]Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany

[4]Lithuanian University of Health Sciences, Kaunas, Lithuania

[5]Medical Faculty, University of Valencia, Valencia, Spain

[6]Medical Faculty, Rheinisch-Westfälische Technische Hochschule Aachen University, Aachen, Germany

[7]Charité - Universitätsmedizin Berlin (corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health), Berlin, Germany

[*]these authors contributed equally

Corresponding Author:
Lina Mosch
Charité - Universitätsmedizin Berlin (corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health)
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 15757206906
Email: lina.mosch@charite.de

Abstract

Background: Digital health technologies hold promise to enhance patient-related outcomes, to support health care staff by reducing their workload, and to improve the coordination of care. As key users of digital health technologies, health care workers are crucial to enable a meaningful digital transformation of health care. Digital health literacy and digital skills should become prerequisite competencies for health professionals to facilitate the implementation and leverage the potential of digital technologies to improve health.

Objective: We aimed to assess European medical students' perceived knowledge and opinions toward digital health, the status of digital health implementation in medical education, and the students' most pressing needs.

Methods: The explanatory design of our mixed methods study was based on an online, anonymous, self-administered survey targeted toward European medical students. A linear regression analysis was used to identify the influence of the year of medical studies on the responses. Additional analysis was performed by grouping the responses by the self-evaluated frequency of eHealth technology use. Written responses to four qualitative questions in the survey were analyzed using an inductive approach.

Results: The survey received a total of 451 responses from 39 European countries, and there were respondents for every year of medical studies. The majority of respondents saw advantages in the use of digital health. While 40.6% (183/451) felt prepared to work in a digitized health care system, more than half (240/451, 53.2%) evaluated their eHealth skills as poor or very poor. Medical students considered lack of education to be the reason for this, with 84.9% (383/451) agreeing or strongly agreeing that more digital health education should be implemented in the medical curriculum. Students demanded introductory and specific eHealth courses covering data management, ethical aspects, legal frameworks, research and entrepreneurial opportunities, role in public health and health systems, communication skills, and practical training. The emphasis lay on tailoring learning to future job requirements and interprofessional education.

Conclusions: This study shows a lack of digital health-related formats in medical education and a perceived lack of digital health literacy among European medical students. Our findings indicate a gap between the willingness of medical students to take an active role by becoming key players in the digital transformation of health care and the education that they receive through their faculties.

(J Med Internet Res 2020;22(8):e19827) doi:10.2196/19827

KEYWORDS

medical students; medical education; eHealth; mixed method; health workforce; digital literacy; curriculum

Introduction

Health care systems around the world are facing challenges connected to an aging population, multimorbidity, an increase of preventable noncommunicable diseases, and health workforce shortages [1-3]. Digital health technologies are seen as a key solution to address these challenges, reinforced by the public health emergency of coronavirus disease 2019 (COVID-19), by having the potential to change the way health services are delivered and promoting the health and well-being of millions of citizens [1,4-6]. For instance, open source technologies have enabled low-cost dissemination and access to data and health information, telehealth technologies have offered communication channels for citizens and health care workers besides physical consultations, and nanotech products have been developed to improve diagnosis and treatment of COVID-19 [7]. However, extensive and sustainable implementation of digital technologies, both into specific clinical settings [8,9] and into national health systems [4,10,11], has been advancing slowly.

Health care professionals play a crucial role in assisting their patients in using digital health technologies appropriately [12,13]. Thus, the need to improve the digital competencies of health workers and citizens to take advantage of digital technologies and facilitate implementation has been emphasized frequently on the international policy level [1,12,14-16].

Major barriers to the successful implementation of digital health technologies are (1) the lack of coordinated, formal education and (2) health care professionals' skepticism and unwillingness toward implementing digital technologies [17-19]. Engaging with these challenges, medical education, and an effective culture of learning could drive the meaningful digitalization of health care [14]. However, to effectively introduce respective topics in medical education, the needs of key stakeholders and the status of the medical curriculum should be considered.

In this paper, we present and discuss the results of a European-wide study assessing the expectations and needs of medical students regarding digital health competencies and the implementation of digital health in the medical curriculum.

Methods

Setting

The European Medical Students' Association (*Association Européenne des Étudiants en Médecine*; EMSA) is a nonprofit, nongovernmental organization representing the voice of medical students from over 110 faculties in 30 countries across Europe [20]. Data were collected according to the European Union General Data Protection Regulation 2016/679 [21].

Study Design

We conducted a cross-sectional mixed methods online survey, following an explanatory approach [22-24]. The survey questions were developed during four online discussions after conducting literature research and collecting feedback from external collaborators. We developed 48 questions, of which 29 were quantitative (including 6 questions on demographic data), and 17 were qualitative (see Multimedia Appendix 1).

Data Collection and Analysis

Data were collected via an informed consent–based online survey in English from June 2018 until August 2018. The target group consisted of medical students in Europe from the first to the seventh year of studies. The survey was distributed via medical faculty mailing lists mapped by EMSA, social media channels, and personal connections.

For the statistical analysis, R (version 3.5.0) and R Studio (version 1.1a) were used [25]. A linear regression analysis was used to identify the influence of the year of medical studies on the responses. Additional analysis was performed by grouping the students depending on the answers to the question "How often are you using eHealth technologies (for example health apps) in your daily life?" P values<.05 were deemed statistically significant. For questions where the possible answers ranged from 0 ("strongly disagree") to 6 ("strongly agree"), the answer options "undecided" and "I do not feel informed enough" were treated adjacently in the linear regression model, as they were situated exactly in the middle of the extremes 0 and 6.

For the linear regression model, the answers "undecided" and "I feel not informed enough" were arbitrarily located adjacent to one another and between the first two and subsequent two options in questions where both these options were included. The dependent variables represented the answers to each question, the independent variable, if not stated otherwise, was the year of medical studies, ranging from first to seventh. Most quantitative questions included an ordinal response format (eg, a Likert-type scale), and in one case, a categorical response format was used. Likert-type scale results were rounded to the next significant digit.

Written responses to four qualitative questions in the survey were analyzed using an inductive approach [26]. The coding and categorization of responses was performed using MaxQDA (version 2020; VERBI GmbH) qualitative data analysis software (Multimedia Appendix 2, Multimedia Appendix 3, Multimedia Appendix 4, and Multimedia Appendix 5) [27]. The results were summarized in paragraphs of continuous text, according to the established code system.

To maintain reflexivity, the research team discussed and established codes and coding in three face-to-face discussions and documented the process of analysis in research diaries throughout the study.

Results

Summary

Our study sample consisted of students (n=451) from 39 European countries in all years of medical studies, most aged 18-24 years (344/451, 76.3%). Respondents indicated a need for more eHealth implementation into medical curriculum (agree or strongly agree: 383/451, 84.9%) and a subjective lack of digital skills among the surveyed medical students (evaluated eHealth skills as poor or very poor: 240/451, 53.2%). One quarter (110/451, 24.4%) stated that their faculty provides no eHealth-related courses at all. Surveyed students reported courses on ethical discussions (247/451, 54.8%), research opportunities (211/451, 46.8%), computer science (138/451, 30.6%), usage of eHealth technologies (83/451, 18.4%), start-up possibilities (69/451, 15.3%), and other (36/451, 8.0%).

Qualitatively, we found variability in the responses to the question “Please define eHealth in your own words,” both in the level of detail, scope, and specification. According to the respondents, implementing eHealth in their curriculum would prepare students for their future working environment, keep education up-to-date, and reduce their doubts about eHealth. The already high density of learning content in the medical curriculum was seen as a counterpoint to this. The feeling of being ready to work in a digitized health care system was based on the students’ own technical interests and motivations rather than on adequate training in their faculties. The respondents emphasized being willing to learn about eHealth. They indicated a need for an introduction to eHealth and for specific courses (data management, practical training with eHealth technologies, courses on informatics, ethical aspects, legal frameworks, research and entrepreneurial opportunities, role in public health and health systems, communication skills). Content-wise, the interest lay in learning about recent developments and technologies, health information systems, and artificial intelligence in health. Also, suggestions for cross-disciplinary courses, for teaching eHealth as a separate discipline were made, and that learning should be tailored to future job requirements and interprofessional education.

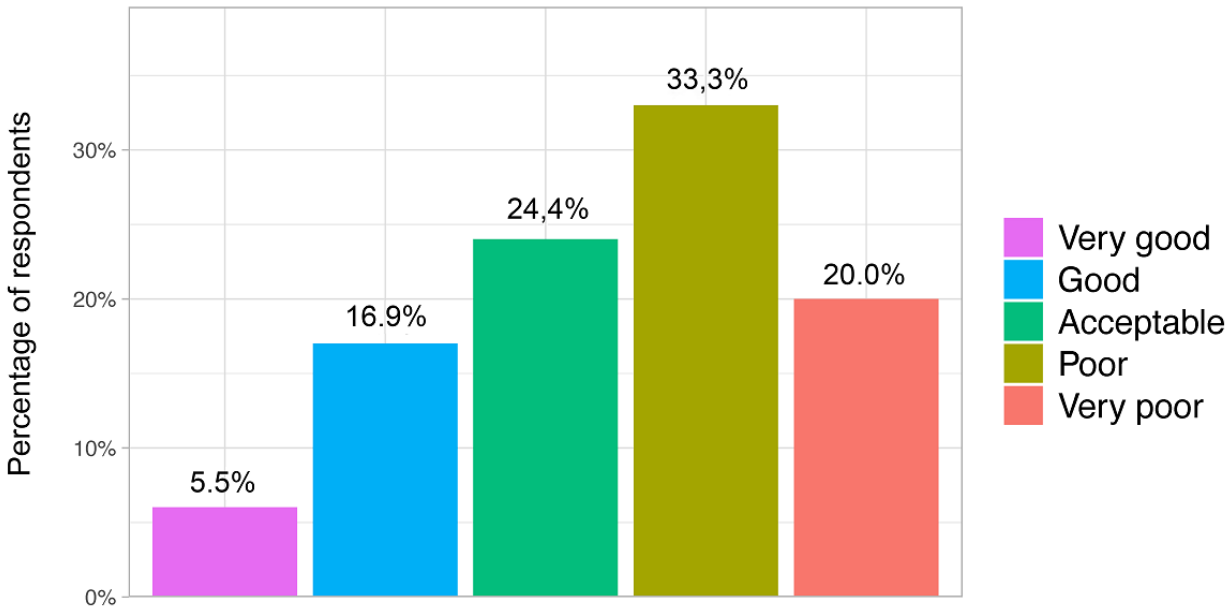
Demographic Data

In total, 459 replies were received, of which 2 (0.4%) were left empty and 6 (1.3%) were not willing to participate after reading the initial survey description, resulting in 451 (98.3%) respondents. Our sample was evenly distributed between the first and the sixth year of medical studies. There were fewer seventh-year medical students (22/451, 4.9%) since medical programs with a duration of seven years exist in only a few European countries. The majority of respondents were between 18 and 24 years old (344/451, 76.3%), followed by 25-34 years (98/451, 21.7%). In total, we received responses from 39 countries in the European region with most responses coming from Germany (134/451, 29.0%), Portugal (49/451, 10.9%), and Turkey (39/451, 8.6%).

Quantitative Results

In general, more than half of the respondents (239/451, 53.0%) strongly agreed or agreed on being familiar with the term *eHealth*. Together, almost two-thirds (274/451, 60.8%) of the respondents claimed to never use eHealth technologies or only every other week. Overall, they had positive expectations toward eHealth: they saw mainly or more advantages in mHealth (362/451, 80.3%), telehealth (314/451, 69.6%), and big data (302/451, 67.0%). The respondents strongly agreed or agreed (272/451, 60.3%) that health care professionals should be responsible for eHealth knowledge and skills of their patients. More than half of the respondents (240/451, 53.2%) evaluated their eHealth skills regarding working with eHealth technologies as poor or very poor (Figure 1).

Figure 1. Self-evaluation in response to the statement: "I evaluate my eHealth skills (eg, working with clinical decision support systems, remote patient monitoring systems, artificial intelligence, applications in radiology) as...".



Respondents who used eHealth technologies more frequently evaluated their eHealth skills on average as better (P<.001) (Figure 2). A majority of respondents (183/451, 40.6%) stated they felt prepared for working in a digitized health care system. Regarding the implementation of eHealth in medical education, we found that 81.8% (369/451) of the students received between 0 and 5 hours of eHealth training during their medical studies. The majority of students (383/451, 84.9%) agreed or strongly agreed that eHealth should be increasingly included in the medical curriculum; 8.9% (40/451) were undecided (Figure 3).

Regarding eHealth-related topics provided by the faculty, students stated they received courses on ethical discussions (247/451, 54.8%), followed by research opportunities (211/451, 46.8%), computer science (138/451, 30.6%), usage of eHealth technologies (83/451, 18.4%), start-up possibilities (69/451, 15.3%), and other (36/451, 8.0%). One-quarter (110/451, 24.4%) of the respondents stated that their faculty provided no eHealth-related courses at all. One-third (149/451, 33.0%) said that they were not informed enough to answer this question (Figure 4).

Figure 2. Self-evaluation of the respondents' eHealth skills in relation to their time spent using eHealth technologies in response to the statement: "I evaluate my eHealth skills (eg, working with clinical decision support systems, remote patient monitoring systems, artificial intelligence, applications in radiology) as...".

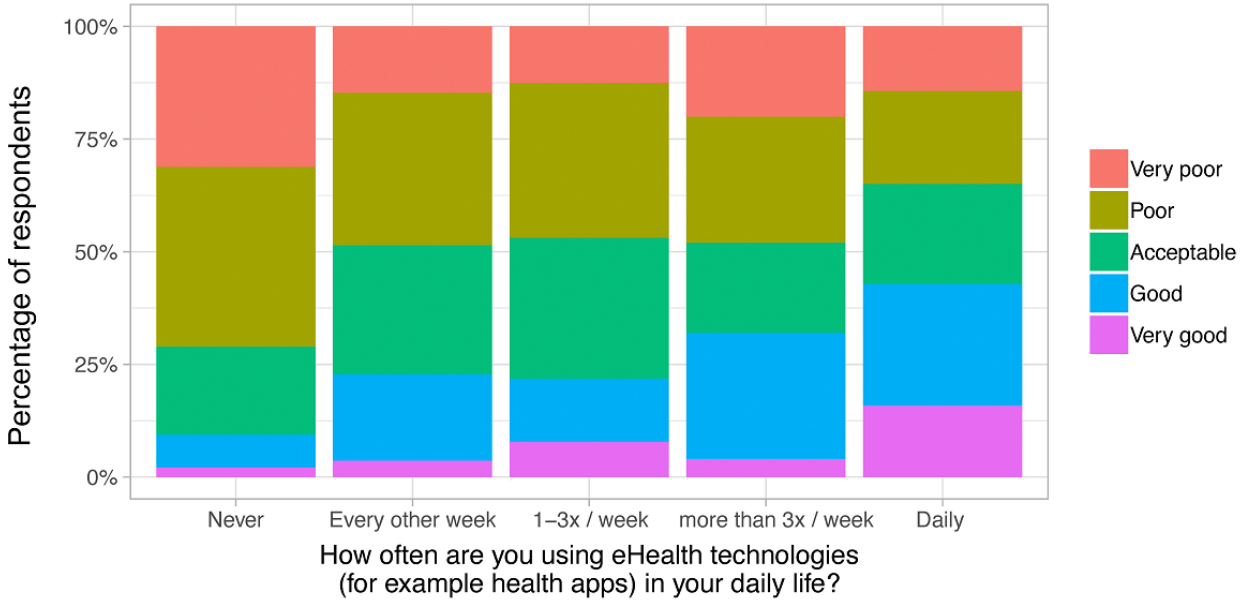


Figure 3. Self-evaluation in response to "Regarding the statement 'I would like eHealth to be more implemented in the medical curriculum,' do you...?".

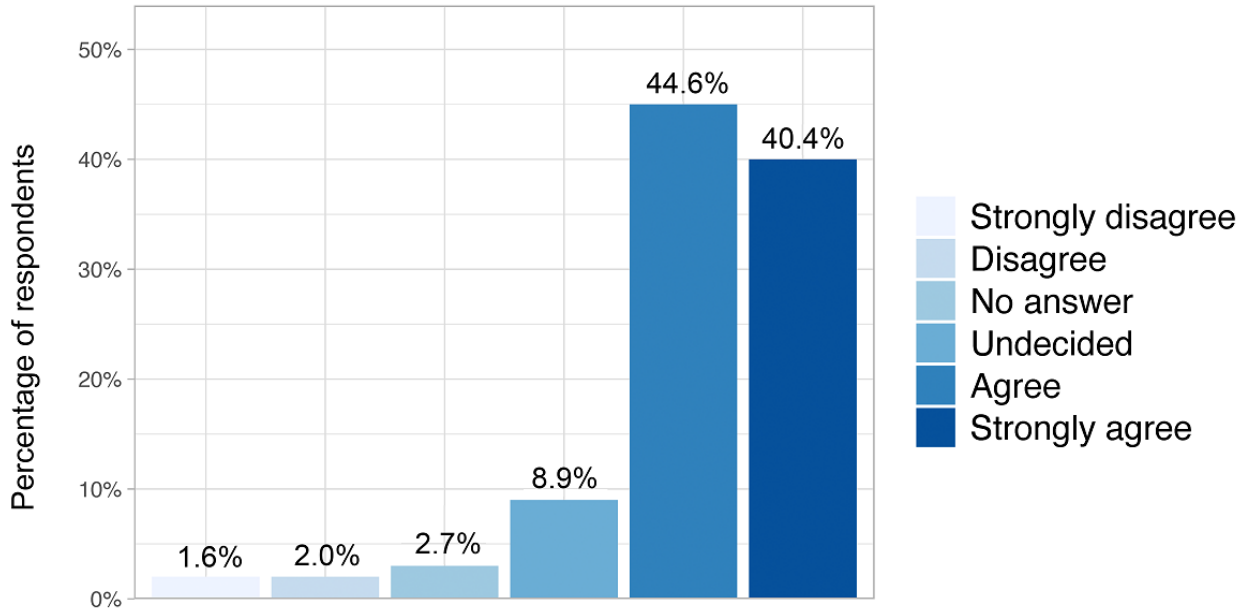
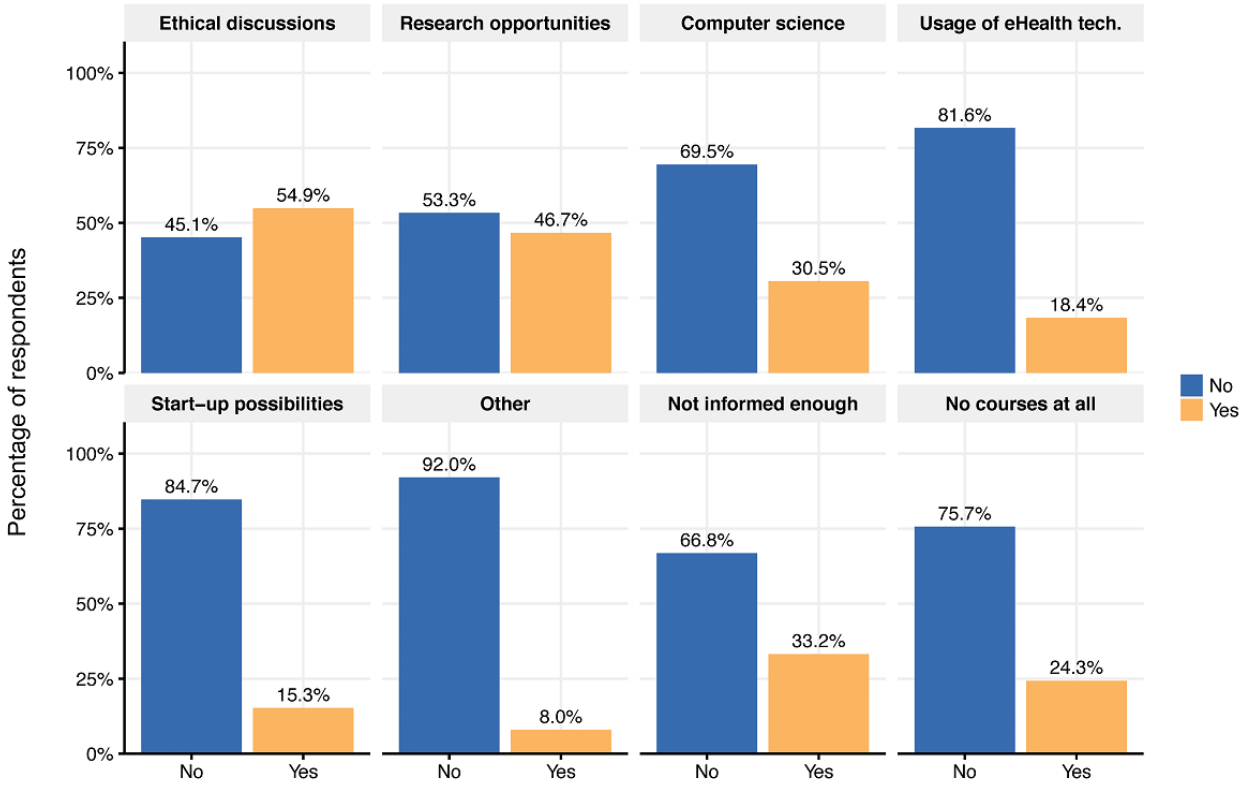


Figure 4. Overview of eHealth-related topics provided by the respondents' faculty (ie, response to the question: "What eHealth-related topics does your faculty provide courses on?).



Qualitative Results

Defining eHealth

Regarding the task "Please define eHealth in your own words," 30.3% (137/451) did not give a definition and 1.5% (7/451) admitted not knowing how to define the term *eHealth*. The resulting 68.1% (307/451) provided definitions that were coded and categorized using a grounded theory approach. Of these, 41.7% (128/307) provided a definition coded as *Usage of technologies in health* (Figure 5, Multimedia Appendix 2). The specifications were related to the purpose of technology application in health: (1) delivery of health, (2) monitoring and documentation of health, (3) communication (between health care professionals as well as between doctors and patients), and (4) public health (eg, health promotion, patient-centered health care); 18.9% (58/307) of the definitions fell into the category *Health care services* for patients and health care professionals where eHealth was seen either as a complement to doctors or as a replacement for them (eg, image recognition, clinical decision support systems); 15.6% (48/307) provided definitions of eHealth assigned the code *Technologies affecting the health care sector*, naming examples of software, hardware, internet tools, and health apps. The last category, eHealth as a *Field of medicine* putting digital technologies into practice and research accounted for 7.1% (22/307). The remaining 16.6% (51/307) could not be allocated to any of the preceding categories and fell under the category Other.

Figure 5. Responses to the statement: "Please define eHealth in your own words." The percentages are in relation to the number of definitions given.

Reasoning For and Against eHealth in Medical Education

The task "Regarding the statement 'I would like eHealth to be more implemented in the medical curriculum,' do you strongly agree, agree, undecided, disagree, strongly disagree? Why?" resulted in a total amount of 151 qualitative answers (151/451, 33.5%). The majority of all respondents (383/451, 84.9%) wished for a stronger implementation of eHealth into the medical curriculum. The reasons provided were that it would prepare them for their future work environment and that learning about eHealth in medical school would be part of keeping the curriculum up-to-date with the latest developments in medicine.

> *Medicine is a science that should always follow and walk hand by hand with the progress in other resources. Medical school should adapt those changes and provide the best knowledge to the upcoming physicians.* [survey response]

Furthermore, medical students justified their wish for personal benefits such as digital health literacy as a job qualification and to maintain physicians' responsibilities and power within the hospital environment and society.

> *...doctors have to be in charge not IT companies, doctors have to fight for their interest; this is only possible if we have the knowledge and skills...* [survey response]

Respondents stated that the implementation of eHealth into medical education could decrease the students' doubts about eHealth technologies. Education on eHealth would drive its effective implementation into health care systems and innovation processes in health. Respondents who were undecided said this was because of a lack of capacities due to an already high workload in medical studies. Some felt that both the quality of education and the resources at their university were not sufficient to sustain the implementation of digital health into the curriculum. Among the reasons for respondents disagreeing were arguments that eHealth topics were already implemented in their curriculum or that digital health literacy should have already been taught in earlier education.

> *No, this starts much earlier. Dealing with modern technology needs to be taught properly in schools. At university level, students should have basic abilities to deal with ehealth on their own!* [survey response]

Reasons for Feeling or Not Feeling Prepared to Work With eHealth Technologies

The results to the question of whether the respondents felt prepared to work in a digitized health care system (and why or why not) showed mixed opinions. The majority of all respondents (183/451, 40.6%) stated they felt prepared; however, this was not based on adequate training in their faculty but rather on their own technical skills and interests. The respondents agreed that eHealth should be addressed in more medical curricula, both technically-theoretically and practically.

> *The generation now studying medicine was raised during digitalization so we are very skilled in using its products.* [survey response]
>
> *My ability to use electronic devices and software in general is good, still it would be nice to be prepared for specific software and situations. In university digitization is not considered at all, neither in teaching nor in patient care.* [survey response]

Respondents not feeling prepared for working in a digitized health care system (130/451, 28.8%) justified this with no or only little eHealth education in their curriculum. They emphasized their willingness to learn more about eHealth.

> *I haven't received enough eHealth knowledge or practical skills in order to be able to work appropriately in a digitized health care system. In my opinion, I do not feel completely prepared for working in a digitalized health care system as there is a lack of training in our medical curricula and poor sources in our health care system.* [survey response]

Respondents who were undecided (129/451, 28.6%) stated that they did not experience any or enough training to evaluate their expertise. They, however, were willing to gain new or extend existing skills:

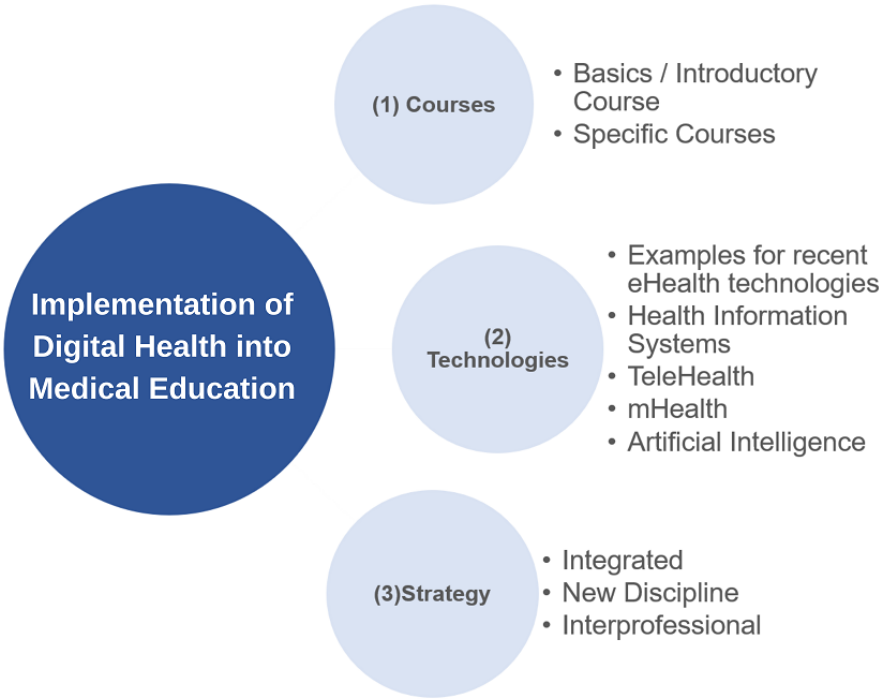
> *I'm not sure what a digitized health care system is and I don't remember if I ever had to interact with it. However, I think it's something I might be able to handle with my present knowledge.* [survey response]

Opinions on How to Implement eHealth in Medical Education

The question "What eHealth-related courses would you like to have in your university's curriculum? (free text)" received 451 answers. Of those, 14.2% (64/451) stated they did not know which courses should be implemented, 9.1% (41/451) gave no answer, and 4.0% (18/451) did not want to have any courses on eHealth. The remaining 72.7% (328/451) were coded and divided into the following categories: Courses, Technologies, and Strategy (Figure 6); hereby, one response could be affiliated with several categories.

Regarding *Courses*, respondents pointed out their need for an introduction to eHealth. For specific courses, education on data management including big data analysis, data sharing, and data security was mentioned most often. Furthermore, practical training with eHealth technologies and courses in computer science (eg, programming languages, app development), the operating principles of eHealth technologies ethical aspects, legal frameworks, research and entrepreneurial opportunities related to eHealth, and the role of eHealth in public health and in health systems were requested. Several respondents wished for courses focusing on communication skills, in particular on how to advise and guide patients using digital health tools. Regarding *Technologies*, respondents wanted to learn about recent examples of eHealth technologies (eg, telehealth and mHealth applications, virtual reality simulations, and robotic applications), health information systems (eg, electronic patient records), and artificial intelligence in health care (eg, clinical decision support systems). Regarding *Strategy* (of implementing eHealth into the medical curriculum), respondents suggested the introduction of cross-disciplinary courses (eg, eHealth in radiology, cardiology), or on the other hand, that eHealth be made a new individual discipline. In both cases, it was suggested that training should be tailored to future job requirements and taught interprofessionally, by involving information technology specialists.

Figure 6. Free text categories derived from responses to the question: "What specific eHealth courses would you like to be implemented in your university's curriculum?".



Discussion

Principal Findings

The results of this survey illustrated a gap between digital health literacy among medical students and the lack of training and education, despite the willingness of medical students to become key players in the digital transformation of health care. Qualitative analysis of eHealth definitions showed that the student understanding of the subject of eHealth varied considerably. Still, we found all domains of eHealth as described previously by Shaw et al [28] to be covered by the responses. The study showed that student needs in terms of digital health education and training are global and specific; students in the study wished for the implementation of courses ranging from programming languages, legal aspects, and guiding patients using digital health technologies. Students also demanded training with already established health information technology such as health information systems. We found mixed opinions on whether eHealth education should be implemented as a separate discipline or in an integrated form. The respondents highlighted the importance of interprofessional education.

Digital Health Literacy and Skills for the Future Health Workforce

Among the major factors enabling the successful implementation of eHealth are digital health literacy, digital skills of health care professionals, along with health care professionals' trust in the potential of novel digital health solutions [9,29]. Furthermore, education and training of health care professionals have been identified as key facilitators of digital health implementation [4,8]. The promotion of these factors requires a multifaceted and diverse approach in order to engage with different stakeholders [1,12,16].

Identifying health care professionals and health care students as central connectors within a challenging digitalization of the health care sector is paramount to its efficacy. In accordance with previous research on the digital skills of health care professionals [12,16,30], our survey showed a gap between the overall willingness of students to become key players in a meaningful digitalization of health care and the competencies and skills they have acquired through their learning.

Future-Proof Health Care Curriculum

In 2019, the Topol Review [14] identified the top digital technologies affecting 80% of the health workforce in 2040: digital medicine (eg, telemedicine and mHealth), artificial intelligence, robotics, and genomics. Future health care professionals have to become aware of the ethical and patient safety considerations posed by the digital transformation of health care [31]. Additionally, the shift away from technical tasks toward a more patient-centered medicine accompanied by the change in the traditional doctor-patient relationship requires different communication skills [13]. Continuous education and training of health care professionals regarding digital health literacy and skills help to ensure the most effective application of digital health to clinical workflows [32,33]. However, medical students do not necessarily need to become experts in programming and data science as the medical profession will probably continue to be a social one. In line with these considerations and recent publications [14,34,35], the findings of this study indicate that medical education should incorporate basic courses on digital medicine, artificial intelligence in health care, genomics, and data science.

Including the various dimensions of digital health in health curricula is a challenging, yet urgently needed task. Education providers can and should build upon preliminary work [36-41]. Priorities and action plans for the improvement of information technology skills in the EU health care workforce have recently been set out [37].

The recognition of digital skills as key competencies by national training commissions and examination authorities is an essential factor for driving the implementation of this content in health care education [42]. To date, accreditation of digital health literacy and skills in national medical education frameworks is lacking [36,42]. The renowned CanMEDs framework, outlining core competencies for future physicians and serving as orientation for medical educators globally, mentions the term *digital* only in the context of communication and documentation of health data [43]. Our findings support suggestions to revise the CanMEDs and other national frameworks according to the requirements that arise in relation to digital health technologies [13,35,42,44].

Practicing physicians are educators whose proficiency in digital health has a direct impact on the learning outcomes of undergraduates [37]. Therefore, teaching digital health literacy and skills should follow a holistic approach and be integrated into both undergraduate and continuing medical education [45-48].

Interprofessional Collaboration and International Best Practice Exchange

The creation of future-proof health curricula involves the integration of interprofessional collaboration and supranational coordination and monitoring. Evolving through the disruptive changes brought about by digitalization, health care engages many different professions [44,49]. Topics related to engineering, computer science, and entrepreneurship have become increasingly relevant for medical professionals [49,50]. Our study indicates that medical students are eager to learn interprofessionally and wish to include input from other fields to their education.

In Europe, several networks have been established to support the digitalization of health care and its implementation in health care education [51-54]. Approaches coordinated by the European Commission or EU member states take into account the heterogeneity of European health care systems and the different pace of digitalization in EU member states [16,51]. Additionally, the European exchange of best practices on an institutional level is focused on driving advances in digital health literacy for health care professionals and digital health implementation [14,55]. For instance, platforms such as the European Institute of Innovation and Technology Health, the European Deans' meeting Training Future-Proof Doctors for the Digital Society, and respective medical education conferences [31,53,56,57] bring together experts and stakeholders in the field. Such

platforms for interprofessional collaboration and best practice exchange as well as cross-disciplinary training are essential to ensure continuous improvements in a rapidly changing field.

Limitations

Our survey was answered by first- to seventh-year medical students throughout Europe. Some of the questions may not have been fully answered by first-year medical students, as some may not yet have had a complete overview of their medical curriculum. Moreover, it is important to mention that the students' perception of their curriculum might differ from the actual courses offered by their university. The lack of awareness among students suggests faculties need to promote already existing courses.

Our findings may even overestimate the digital health literacy of European medical students. This may be due to two aspects. First, the voluntary mode of participation may suggest that respondents were already aware of and interested in digital health. Thus, it is possible that they do not represent the average knowledge and skill level of medical students regarding digital health and are already more confident about their e-skills abilities than their colleagues would be. Second, people with weak skills tend to overestimate their skills and expected performance [58]. Further research on the eHealth literacy of medical students using validated assessment tools [59,60] would be necessary in order to obtain a clearer and more objective picture.

Due to the limited number of respondents, our results can be considered neither complete nor representative of the European region. Rather, the findings should be seen as a starting point for further research in this area. In a follow-up study, various results could be specifically queried again and linked to current developments in health care education. In addition, findings can be an impetus for policymakers and stakeholders in health care education to revise their approach to meet the needs of future health care professionals with regard to digital skills and eHealth literacy.

The study was conducted in English only, which may have resulted in misunderstandings, inaccuracies, and flawed answers due to a language barrier. Despite the definitions of technical terms given in the survey, their complexity and ambiguity may have caused difficulties in understanding, and subsequently, in answering survey questions. Furthermore, when responding to a Likert scale, it has been shown that respondents were more likely to pick answers in the middle of the scale when responding to an English-language questionnaire compared to a questionnaire in their native language. In turn, questionnaires in a respondents' native language provoked a higher level of extreme responses [61]. Next to this, the field of eHealth is multidimensional and often not directly labeled as such [28,62], which could have led to misunderstandings among respondents, and thus, biased the survey results.

Conclusion

This study was the first pan-European approach assessing the needs of medical students regarding digital health literacy and digital skills in medical education. We revealed that the majority of European medical students have a positive attitude toward the digitalization of health care and are willing to play an active role and take responsibility especially as mediators of digital health literacy to patients. However, we also found a lack of knowledge and skills regarding the adequate use and evaluation of digital health technologies, attributed to a lack of respective topics in the medical curriculum. We showed the students' demand for new, additional teaching concepts ranging from technical-theoretical issues (data management, computer science, legal, and ethical aspects) to practical training with specific technologies and patient communication.

The apparent gap between the overall willingness of medical students to become key players in the digitalization of health care and the education they receive poses a significant challenge to the successful implementation of digital technologies into health care settings. Education providers and policymakers should acknowledge the central role of future health care professionals in health innovation, develop interprofessional concepts ensuring continuous learning, and evaluate them in a continual exchange among themselves and with their students, adapting to the latest scientific and technological developments.

Further research on the specific needs of health care professionals is necessary as new challenges in the growing field of digital health continuously arise. The medical curriculum is essential to create preparatory experiences regarding digital health literacy and digital skills before students enter their professional life. Our findings support the call for faculties and medical education institutions to collaboratively establish targeted, customized, and efficient education and training on digital health.

Acknowledgments

We express our gratitude to the respondents of our survey.

We would like to thank the following students who have contributed to our work: Paulius Povilonis, Cristina-Nelida Bora, Maria Roque, Aykut Uçar, Dilara Tüzün, Friederike Münn, Sebastian Weyn Banningh, Philippa Lantwin, Leu Huang, Sophia Hodgkinson, Jolan van Maaren, Orsolya Süli, and Tuğçe Çetin. Furthermore, we would like to thank advisors Felix Balzer, Akira-Sebastian Poncette, Armin Ott, Bianca Jansky, Michael Rosenblatt, Carole Rouaud, Jehona Krasniqi for their support in the conception of the study and survey design.

Authors' Contributions

FM, JB, and LM conceived the study. The survey was developed by FM, JB, FVM, and LM. RK performed the statistical analysis. LM and DJ conducted the qualitative data analyses. FM, RK, and LM wrote the manuscript, supported by JB, BA, and FVM. LM and FM supervised all parts of the study. All authors critically reviewed and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Catalog of all survey questions.
[PDF File (Adobe PDF File), 237 KB - jmir_v22i8e19827_app1.pdf]

Multimedia Appendix 2
Codes and codings for "Please define eHealth in your own words.".
[PDF File (Adobe PDF File), 222 KB - jmir_v22i8e19827_app2.pdf]

Multimedia Appendix 3
Codes and codings for "Regarding the statement 'I would like eHealth to be more implemented in the medical curriculum,' do you... strongly agree, agree, undecided, disagree, strongly disagree, no answer - Why?".
[PDF File (Adobe PDF File), 160 KB - jmir_v22i8e19827_app3.pdf]

Multimedia Appendix 4
Codes and codings for "Regarding the statement 'I feel prepared for working in a digitized health care system,' do you... strongly agree, agree, undecided, disagree, strongly disagree, no answer - Please elaborate.".
[PDF File (Adobe PDF File), 123 KB - jmir_v22i8e19827_app4.pdf]

Multimedia Appendix 5
Codes and codings for "What specific eHealth-related courses would you like to have in your university’s curriculum?".
[PDF File (Adobe PDF File), 274 KB - jmir_v22i8e19827_app5.pdf]

Multimedia Appendix 6
Supplementary figures.
[PDF File (Adobe PDF File), 508 KB - jmir_v22i8e19827_app6.pdf]

References

1. European C. Communication on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society. https://ec.europa.eu.: European Commission; 2018 Apr 25. URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51628 [accessed 2020-04-03]
2. Richard S, John R B, Ties B, Somnath C. Health in an ageing world—what do we know? The Lancet. 2005 Nov. URL: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(14)61597-X/fulltext [accessed 2019-06-22]
3. Non CDI. Noncommunicable diseases. World Health Organization. 2018 Jun 01. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2019-06-22]
4. World Health Organization. WHO Guideline: recommendations on digital interventions for health system strengthening. World Health Organization. WHO guideline; 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1 [accessed 2020-04-02]
5. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. The Lancet Digital Health 2020 Apr;2(4):e166-e167. [doi: 10.1016/s2589-7500(20)30054-6]
6. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12:m998. [doi: 10.1136/bmj.m998]
7. Kritikos M. Ten technologies to fight coronavirus. EPRS | European Parliamentary Research Service.: European Parliament; 2020 Apr. URL: https://www.europarl.europa.eu/RegData/etudes/IDAN/2020/641543/EPRS_IDA(2020)641543_EN.pdf [accessed 2020-05-22]
8. Wachter R. Making IT Work: Harnessing the Power of Health Information Technology to Improve Care in England. Report of the National Advisory Group on Health Information Technology in England.: The National Advisory Group on Health Information Technology in England; 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf [accessed 2019-06-22]

9. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implementation Sci 2016 Oct 26;11(1):146. [doi: 10.1186/s13012-016-0510-7]
10. Hüsers J, Hübner U, Esdar M, Ammenwerth E, Hackl WO, Naumann L, et al. Innovative Power of Health Care Organisations Affects IT Adoption: A bi-National Health IT Benchmark Comparing Austria and Germany. J Med Syst 2017 Feb;41(2):33 [FREE Full text] [doi: 10.1007/s10916-016-0671-6] [Medline: 28054195]
11. Bertelsmann Stiftung. Summary #SmartHealthSystems - Focus Europe. https://www.bertelsmann-stiftung.de/en/publications.: Bertelsmann Stiftung; 2019. URL: https://www.bertelsmann-stiftung.de/fileadmin/files/Projekte/Der_digitale_Patient/VV_SHS_Europe_eng.pdf [accessed 2020-04-02]
12. European HP. Digital Skills for Health Profesionals. European Health Parliament. 2016. URL: https://www.healthparliament.eu/wp-content/uploads/2017/09/Digital-skills-for-health-professionals.pdf [accessed 2020-04-02]
13. Masters K. Preparing medical students for the e-patient. Medical Teacher 2017 May 22;39(7):681-685. [doi: 10.1080/0142159x.2017.1324142]
14. Topol E. Preparing the healthcare workforce to deliver the digital future Internet. NHS Health Education England. 2019 Feb. URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf [accessed 2019-06-22]
15. European Commission. Transatlantic eHealth/health IT Cooperation Roadmap. https://ec.europa.eu.: European Comission; 2013 Mar 15. URL: https://ec.europa.eu/digital-single-market/en/news/transatlantic-ehealthhealth-it-cooperation-roadmap [accessed 2019-06-22]
16. Keane L. D6.3 Report on eSkills for Professionals Information Note; WP 6 Enhancing Continuity of Care. 15th eHN meeting, June 2019.: eHAction - Joint Action supporting the eHealth Network; 2019 Jun. URL: https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20190611_co322_en.pdf [accessed 2019-09-13]
17. The Evidence Centre for Skills for Health. How do new technologies impact on workforce organisation? https://www.skillsforhealth.org.uk/. 2011. URL: https://www.skillsforhealth.org.uk/index.php?option=com_mtree&task=att_download&link_id=101&cf_id=24 [accessed 2019-06-22]
18. GOVUK Department for Digital, Culture, Media & Sport. Digital skills and inclusion - giving everyone access to the digital skills they need Internet. GOVUK Department for Digital, Culture, Media & Sport. 2017 Mar. URL: https://www.gov.uk/government/publications/uk-digital-strategy/2-digital-skills-and-inclusion-giving-everyone-access-to-the-digital-skills-they-need [accessed 2019-06-22]
19. Parry E, Battista V. The impact of emerging technologies on work: a review of the evidence and implications for the human resource function. Emerald Open Res 2019 Jan 28;1:5. [doi: 10.12688/emeraldopenres.12907.1]
20. European Medical Students' Association. European Medical Students' Association - Association Européenne des Étudiants en Médecine (EMSA). https://emsa-europe.eu/. URL: https://emsa-europe.eu/ [accessed 2019-09-13]
21. Regulation. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46. https://eur-lex.europa.eu/. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04 [accessed 2019-06-22]
22. Jansen H. The Logic of Qualitative Survey Research and its Position in the Field of Social Research Methods. Forum Qual Sozialforschung Forum Qual Soc Res 2010 Mar 21;Vol 11:No 2 (2010;Visualising Migration and Social Division:2-18 [FREE Full text] [doi: 10.17169/fqs-11.2.1450]
23. Guetterman TC, Babchuk WA, Howell Smith MC, Stevens J. Contemporary Approaches to Mixed Methods–Grounded Theory Research: A Field-Based Analysis. Journal of Mixed Methods Research 2017 Jun 02;13(2):179-195. [doi: 10.1177/1558689817710877]
24. Watling CJ, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. Med Teach 2012 Jan;34(10):850-861. [doi: 10.3109/0142159X.2012.704439] [Medline: 22913519]
25. The R Project for Statistical Computing. https://www.r-project.org/. URL: https://www.r-project.org/ [accessed 2019-06-22]
26. Boyatzis R. Transforming Qualitative Information: Thematic Analysis and Code Development. London and New Delhi: SAGE; 1998:-978.
27. MAXQDA 2020. MAXQDA 2020. Berlin, Germany: VERBI Software; 2019. URL: https://maxqda.com [accessed 2020-04-02]
28. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a Conceptual Model for eHealth: Qualitative Study with Key Informants. J Med Internet Res 2017 Oct 24;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]
29. Lennon MR, Bouamrane M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for Delivering Digital Health at Scale: Lessons From a Longitudinal Qualitative Evaluation of a National Digital Health Innovation Program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]
30. Stellefson M, Hanik B, Chaney B, Chaney D, Tennant B, Chavarria EA. eHealth literacy among college students: a systematic review with implications for eHealth education. J Med Internet Res 2011 Dec;13(4):e102 [FREE Full text] [doi: 10.2196/jmir.1703] [Medline: 22155629]
31. European Commission. Digital is transforming medical doctors? daily work: deans and student doctors agree on common principles to adapt medical doctors? training Internet. European Commission. 2019. URL: https://ec.europa.eu/

digital-single-market/en/news/digital-transforming-medical-doctors-daily-work-deans-and-student-doctors-agree-common [accessed 2019-06-22]
32. Torous J, Chan S, Luo J, Boland R, Hilty D. Clinical Informatics in Psychiatric Training: Preparing Today's Trainees for the Already Present Future. Acad Psychiatry 2018 Oct;42(5):694-697. [doi: 10.1007/s40596-017-0811-4] [Medline: 29047074]
33. Cresswell KM, Lee L, Mozaffar H, Williams R, Sheikh A. Sustained User Engagement in Health Information Technology: The Long Road from Implementation to System Optimization of Computerized Physician Order Entry and Clinical Decision Support Systems for Prescribing in Hospitals in England. Health Serv Res 2016 Oct 07;52(5):1928-1957. [doi: 10.1111/1475-6773.12581]
34. Fridsma DB. Health informatics: a required skill for 21st century clinicians. BMJ 2018 Jul 12:k3043. [doi: 10.1136/bmj.k3043]
35. Rampton V, Mittelman M, Goldhahn J. Implications of artificial intelligence for medical education. The Lancet Digital Health 2020 Mar;2(3):e111-e112. [doi: 10.1016/s2589-7500(20)30023-6]
36. Gray K, Dattakumar A, Maeder A, Butler-Henderson K, Chenery H. Advancing ehealth education for the clinical health professions: final report. Australia, Office for Learning and Teaching. 2014. URL: https://research.bond.edu.au/files/26149651/Advancing_Ehealth_Education_for_the_Clinical_Health_Professions_Final_Report_2014.pdf [accessed 2019-09-12]
37. Li S, Bamidis PD, Konstantinidis ST, Traver V, Car J, Zary N. Setting priorities for EU healthcare workforce IT skills competence improvement. Health Informatics J 2017 Apr 25;25(1):174-185. [doi: 10.1177/1460458217704257]
38. Mesko B, Győrffy Z, Kollár J. Digital Literacy in the Medical Curriculum: A Course With Social Media Tools and Gamification. JMIR Medical Education 2015 Oct 01;1(2):e6. [doi: 10.2196/mededu.4411]
39. Milano CE, Hardman JA, Plesiu A, Rdesinski RE, Biagioli FE. Simulated electronic health record (Sim-EHR) curriculum: teaching EHR skills and use of the EHR for disease management and prevention. Acad Med 2014 Mar;89(3):399-403 [FREE Full text] [doi: 10.1097/ACM.0000000000000149] [Medline: 24448035]
40. Brunner M, McGregor D, Keep M, Janssen A, Spallek H, Quinn D, et al. An eHealth Capabilities Framework for Graduates and Health Professionals: Mixed-Methods Study. J Med Internet Res 2018 May 15;20(5):e10229 [FREE Full text] [doi: 10.2196/10229] [Medline: 29764794]
41. Varghese J, Röhrig R, Dugas M. Which competencies in medical informatics are required by physicians? An update of the catalog of learning objectives for medical students. GMS Med Inform Biom Epidemiol German Medical Science GMS Publishing House 2020;16(1):-3 [FREE Full text] [doi: 10.3205/mibe000205]
42. Edirippulige S, Brooks P, Carati C, Wade VA, Smith AC, Wickramasinghe S, et al. It's important, but not important enough: eHealth as a curriculum priority in medical education in Australia. J Telemed Telecare 2018 Oct 22;24(10):697-702. [doi: 10.1177/1357633x18793282]
43. Frank J, Snell L, Sherbino J. CanMEDS 2015: physician competency framework. In: Royal College of Physicians and Surgeons of Canada. Ottawa, Canada: Royal College of Physicians and Surgeons of Canada; 2015:1-978.
44. Ahmadvand A, Gatchel R, Brownstein J, Nissen L. The Biopsychosocial-Digital Approach to Health and Disease: Call for a Paradigm Expansion. J Med Internet Res 2018 May 18;20(5):e189. [doi: 10.2196/jmir.9732]
45. Hilty DM, Turvey C, Hwang T. Lifelong Learning for Clinical Practice: How to Leverage Technology for Telebehavioral Health Care and Digital Continuing Medical Education. Curr Psychiatry Rep 2018 Dec 12;20(3):15. [doi: 10.1007/s11920-018-0878-y] [Medline: 29527637]
46. Cullen MW, Geske JB, Anavekar NS, McAdams JA, Beliveau ME, Ommen SR, et al. Reinvigorating Continuing Medical Education: Meeting the Challenges of the Digital Age. Mayo Clinic Proceedings 2019 Dec;94(12):2501-2509. [doi: 10.1016/j.mayocp.2019.07.004]
47. Schmuter G. Envirotyping in Medical Schools: A Movement Toward Digital Health Literacy. American Journal of Ophthalmology 2019 Nov;207:427-428. [doi: 10.1016/j.ajo.2019.06.028]
48. Stathis K, Li S, Traver V, Zary N, Bamidis P. Actions to Empower Digital Competences in Healthcare Workforce: A Qualitative Approach. Stud Health Technol Inform Netherlands: IOS Press 2017:253-256. [doi: 10.3233/978-1-61499-781-8-253]
49. Nelson R, Staggers N. Health informatics: an interprofessional approach. In: Health informatics: an interprofessional approach. Amsterdam, Netherlands: Elsevier; 2018:a.
50. Salminen L, Lindberg E, Gustafsson M, Heinonen J, Leino-Kilpi H. Entrepreneurship Education in Health Care Education. Education Research International 2014;2014:1-8. [doi: 10.1155/2014/312810]
51. European Commission. eHealth Network. https://ec.europa.eu/. URL: https://ec.europa.eu/health/ehealth/policy/network_en [accessed 2019-07-22]
52. eHAction - Joint Action to Support the eHealth Network Internet. eHAction. URL: http://ehaction.eu/ [accessed 2019-09-12]
53. EIT Health - about Internet. EIT Health. URL: https://eithealth.eu/who-we-are/ [accessed 2019-09-12]
54. European C. 2020 Thematic Network Digital Skills for future-proof doctors – Digital Doc. https://ec.europa.eu/. URL: https://ec.europa.eu/health/sites/health/files/policies/docs/ev_20200618_ag_en.pdf [accessed 2020-06-18]
55. Abbott PA, Foster J, Marin HDF, Dykes PC. Complexity and the science of implementation in health IT--knowledge gaps and future visions. Int J Med Inform 2014 Jul;83(7):e12-e22. [doi: 10.1016/j.ijmedinf.2013.10.009] [Medline: 24444700]

56. AMEE - conferences. AMEE. URL: https://amee.org/conferences [accessed 2019-09-12]
57. Society for Simulation in Europe. SESAM. URL: https://www.sesam-web.org [accessed 2019-09-12]
58. Dunning D, Johnson K, Ehrlinger J, Kruger J. Why People Fail to Recognize Their Own Incompetence. Curr Dir Psychol Sci 2016 Jun 23;12(3):83-87. [doi: 10.1111/1467-8721.01235]
59. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing Competencies Needed to Engage With Digital Health Services: Development of the eHealth Literacy Assessment Toolkit. J Med Internet Res 2018 May 10;20(5):e178 [FREE Full text] [doi: 10.2196/jmir.8347] [Medline: 29748163]
60. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
61. Harzing A. Response Styles in Cross-national Survey Research. Int'l Jnl of Cross Cultural Management 2016 Jul 24;6(2):243-266. [doi: 10.1177/1470595806066332]
62. Chris S, Christian N. How Should We Define eHealth, and Does the Definition Matter? Studies in Health Technology and Informatics 2012;180:881-884 [FREE Full text] [doi: 10.3233/978-1-61499-101-4-881]

Abbreviations

EMSA: The European Medical Students' Association

Edited by G Eysenbach; submitted 04.05.20; peer-reviewed by W Ahmed, S Mehta Ambalal; comments to author 12.06.20; revised version received 24.06.20; accepted 25.06.20; published 14.08.20.

Please cite as:
Machleid F, Kaczmarczyk R, Johann D, Balčiūnas J, Atienza-Carbonell B, von Maltzahn F, Mosch L
Perceptions of Digital Health Education Among European Medical Students: Mixed Methods Survey
J Med Internet Res 2020;22(8):e19827
URL: http://www.jmir.org/2020/8/e19827/
doi:10.2196/19827
PMID:32667899

Original Paper

The Acceptability and Impact of the Xploro Digital Therapeutic Platform to Inform and Prepare Children for Planned Procedures in a Hospital: Before and After Evaluation Study

Lucy Bray[1], BA, MSc, PhD; Ashley Sharpe[1], BSc, MSc; Phillip Gichuru[1], BSc, MSc, PhD; Peter-Marc Fortune[2], BSc, MA, PhD; Lucy Blake[1], BSc, MPhil, PhD; Victoria Appleton[1], BSc, Mres, PhD

[1]Faculty of Health, Social Care and Medicine, Edge Hill University, Ormskirk, United Kingdom

[2]Royal Manchester Children's Hospital, Manchester, United Kingdom

Corresponding Author:
Lucy Bray, BA, MSc, PhD
Faculty of Health, Social Care and Medicine
Edge Hill University
St Helens Road
Ormskirk, L39 4QP
United Kingdom
Phone: 44 01695 657231
Email: brayl@edgehill.ac.uk

Abstract

Background: There is increasing interest in finding novel approaches to improve the preparation of children for hospital procedures such as surgery, x-rays, and blood tests. Well-prepared and informed children have better outcomes (less procedural anxiety and higher satisfaction). A digital therapeutic (DTx) platform (Xploro) was developed with children to provide health information through gamification, serious games, a chatbot, and an augmented reality avatar.

Objective: This before and after evaluation study aims to assess the acceptability of the Xploro DTx and examine its impact on children and their parent's procedural knowledge, procedural anxiety, and reported experiences when attending a hospital for a planned procedure.

Methods: We used a mixed methods design with quantitative measures and qualitative data collected sequentially from a group of children who received standard hospital information (before group) and a group of children who received the DTx intervention (after group). Participants were children aged between 8 and 14 years and their parents who attended a hospital for a planned clinical procedure at a children's hospital in North West England. Children and their parents completed self-report measures (perceived knowledge, procedural anxiety, procedural satisfaction, and procedural involvement) at baseline, preprocedure, and postprocedure.

Results: A total of 80 children (n=40 standard care group and n=40 intervention group) and their parents participated in the study; the children were aged between 8 and 14 years (average 10.4, SD 2.27 years) and were attending a hospital for a range of procedures. The children in the intervention group reported significantly lower levels of procedural anxiety before the procedure than those in the standard group (two-tailed $t_{63.64}$=2.740; P=.008). The children in the intervention group also felt more involved in their procedure than those in the standard group (t_{75}=−2.238; P=.03). The children in the intervention group also reported significantly higher levels of perceived procedural knowledge preprocedure ($t_{59.98}$=−4.892; P=.001) than those in the standard group. As for parents, those with access to the Xploro intervention reported significantly lower levels of procedural anxiety preprocedure than those who did not ($t_{68.51}$=1.985; P=.05). During the semistructured *write and tell* interviews, children stated that they enjoyed using the intervention, it was fun and easy to use, and they felt that it had positively influenced their experiences of coming to the hospital for a procedure.

Conclusions: This study has shown that the DTx platform, Xploro, has a positive impact on children attending a hospital for a procedure by reducing levels of procedural anxiety. The children and parents in the intervention group described Xploro as improving their experiences and being easy and fun to use.

(J Med Internet Res 2020;22(8):e17367) doi:10.2196/17367

KEYWORDS

health literacy; augmented reality; children; procedure; health; artificial intelligence

Introduction

Children can find visiting a hospital to be a stressful and disorientating experience. Research shows that children can experience high levels of anxiety when attending a hospital for procedures [1], normally due to fear of the unknown, medical examinations, pain, separation from parents, uncertainty, and loss of control [2,3]. These fears and worries can result from poor preparation and information [4,5], and can result in children becoming distressed and uncooperative during a clinical procedure [6]. Anxiety and distress linked to undergoing a procedure can be long-lasting and have implications on children's subsequent health care encounters and long-term health outcomes [7]. Children's distress and noncooperation during procedures can also result in longer appointments, delays to appointments, and referrals to psychological services, all of which have cost implications for health service providers. Children and their parents identify an unmet need for information about hospital procedures and interventions [8,9] and that such information would be valuable to help them know what to expect and how best to prepare themselves for a procedure [10].

Traditional forms of preparation for children include leaflets and books [11], and while these have been shown to have some benefit, computer- and app-based interventions have been highlighted as being best placed to deliver preparation information for children coming to the hospital for a planned procedure [12]. There is evidence that computer- and app-based interventions are helpful in educating and preparing children for health experiences, but to date these have focused predominantly on admissions for surgery [13-16] or one particular context (eg, radiology [17]). Although these interventions are valuable, children may encounter numerous clinical procedures, health care professionals, and environments while visiting the hospital. There is a need to develop and robustly evaluate interventions that address the more common interactions and procedures that children encounter within hospitals. Therefore, it was expected that an accessible and child-centered intervention to familiarize and inform children about a broad range of experiences, environments, and health professionals would increase value and improve children's procedural health literacy and experiences.

Health literacy refers to a person's ability to access and gain information, understand this information, and use it to communicate and be involved in making health choices and decisions [18,19]. In relation to the Xploro intervention, we proposed that a child's health literacy would be enhanced by accessing meaningful information (knowledge) through the digital therapeutic (DTx) platform, which they can understand and use to familiarize and prepare themselves for their hospital visits (reduced procedural anxiety, improved experiences, and increased involvement in their procedure). The Xploro intervention has been developed with children based on an *information-pull* design. This design acknowledges that children learn and gain knowledge optimally by actively accessing information and constructing their own understandings through engaging with multiple elements (gamification, augmented reality, serious games, and a chatbot) to influence multiple aspects of a procedure (anxiety, knowledge, involvement, and satisfaction). There is currently a lack of robust research evaluating the use and impact of platforms such as Xploro with children attending a hospital.

This before and after study aims to assess the acceptability of Xploro DTx and examine its impact on children and their parents' health literacy, perceptions of procedural knowledge, procedural anxiety, procedural involvement, procedural satisfaction, and reported experiences when attending a hospital for a planned procedure.

Methods

Study Design

The study was designed as a before and after evaluation study comprising 2 separate groups of children. Data were collected sequentially from a group of children who received standard hospital information (before group) and from a group of children who received the DTx intervention (after group). As the DTx platform can be considered a complex intervention (multiple interacting components with multiple aims), we conducted a before and after study design to enable us to ascertain whether the intervention was acceptable and accessible to children and their parents within a health care setting and evaluate whether the outcomes were favorable [20]. This study would help inform whether further research to assess the effectiveness of the intervention in a larger study could and should be done.

The study used a mixed methods design consisting of structured quantitative measures and qualitative interviews. Self-completion questionnaire booklets were completed separately by children aged 8 to 14 years and their parent or carer at 3 time points: baseline (3 to 5 days before attending a hospital for the procedure), before the procedure (on arrival at the hospital), and after their procedure (within 5-10 min after completing their procedure). The questionnaire booklets collected self-report data on procedural anxiety, perceptions of knowledge, involvement, and satisfaction. The design of the questionnaire booklet was informed by consultations with 10 children and 4 parents to ensure that the directions, language, and measures were easily understood. The children, during this consultation, highlighted that the use of multiple measures or lengthy questionnaires may add to a child's anxiety when attending a hospital for a procedure, so the study was designed to ensure that participant burden was kept to a minimum.

Short qualitative interviews, focused around a *write and tell* [21] activity sheet, were conducted with children and their parents after the procedure. The interviews explored what children and parents perceived went well during the procedure and what might have made it better. For the children in the intervention group, the interviews also sought their opinions and experiences of using the DTx platform (Xploro) and aimed

to gain insight into how children used the intervention within the context of a hospital visit and how it impacted their procedural experiences. The collection of qualitative data from all the children and their parents involved in the study was in recognition of the value of these data as part of an evaluation of a complex intervention [22].

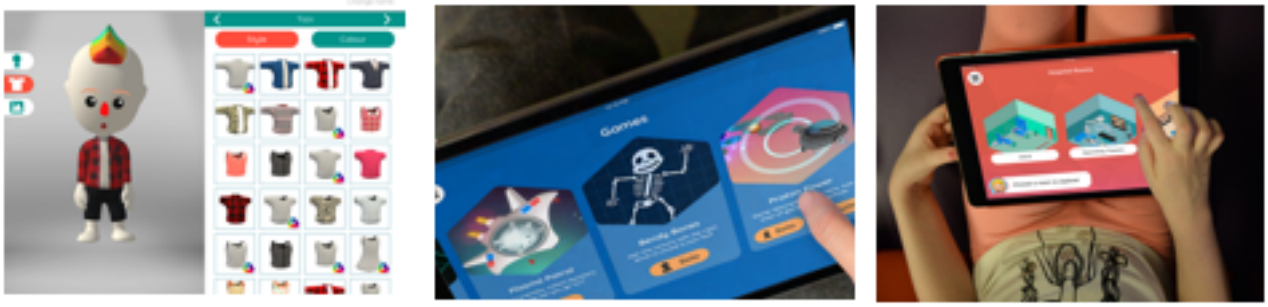
Participants and Recruitment

Participants were children undergoing a planned procedure in a children's hospital in North West England and their attending parents. Children were recruited using convenience sampling for the standard care group between September 2018 and January 2019 and the intervention group between January and June 2019. Eligible children were those aged between 8 and 14 years who were to attend a hospital for a planned clinical procedure without a moderate or severe cognitive impairment or a referral to psychological services for procedural anxiety. The age of the children recruited for the study was determined by the target population for the Xploro intervention. Researchers positioned themselves within outpatient and inpatient departments on different days of the week and worked with clinical teams to identify children who were due to return for a clinical procedure within the next few weeks. Eligibility was determined by the clinical team, who initially approached the family with information about the study.

Study Intervention

Xploro is a DTx platform that uses augmented reality, gameplay, and artificial intelligence to deliver health information to children (Figure 1). The intervention was developed in response to the personal experience of the founder to a lack of engaging information for children attending a hospital for procedures and treatments. The DTx platform provides information about health environments (wards and operating theaters), key health staff, and hospital equipment. The DTx platform includes an avatar that children can customize and acts as a guide and chatbot. The DTx platform also includes several serious games with health themes. The DTx platform includes information about the procedure, environment and staff, and information on sensory aspects of a procedure (what a child may feel or experience) as well as information to help a child build-coping strategies [9].

Figure 1. Different components of the Xploro Digital Therapeutic (DTx) platform.



Xploro development has been informed by a user-centered design or person-based approach [23]. Children (n=105), health professionals (n=19), and parents (n=27) were involved in a previous qualitative study to inform the content (chatbot questions and answers, games, and language used), navigation (working through different elements or easily locating the procedure of relevance to them), and access to the DTx app (parents as gatekeepers and importance of timing of procedural information) [9]. This previous work was conducted in acknowledgment of the importance of qualitative research in informing the development of DTx interventions to ensure that they are engaging, acceptable, and effective [24] and that an investment in developmental work is essential before the formal evaluation of complex interventions [20]. Xploro is designed to supplement, not replace, normal forms of information provision and communication between children and health professionals. For this study, Xploro was accessed by children in the intervention group on a preloaded iPad delivered to each family at least 3 days before the planned procedure.

Measures

Self-report questionnaires were completed by children and their parents at baseline (3 to 5 days before the planned procedure), before the procedure (within the hospital immediately before the procedure), and after the procedure (within the hospital and up to 10 min after the procedure was completed; Figure 2; Table 1).

Children and parents' self-reported ratings were collected for their procedural (state) anxiety (10-point visual analog scale [VAS]), trait anxiety (10-point VAS), procedural knowledge levels (10-point VAS), procedural satisfaction (10-point VAS), and procedural involvement (5-point Likert scale). These were single-item scales.

Figure 2. Flow diagram of the study processes.

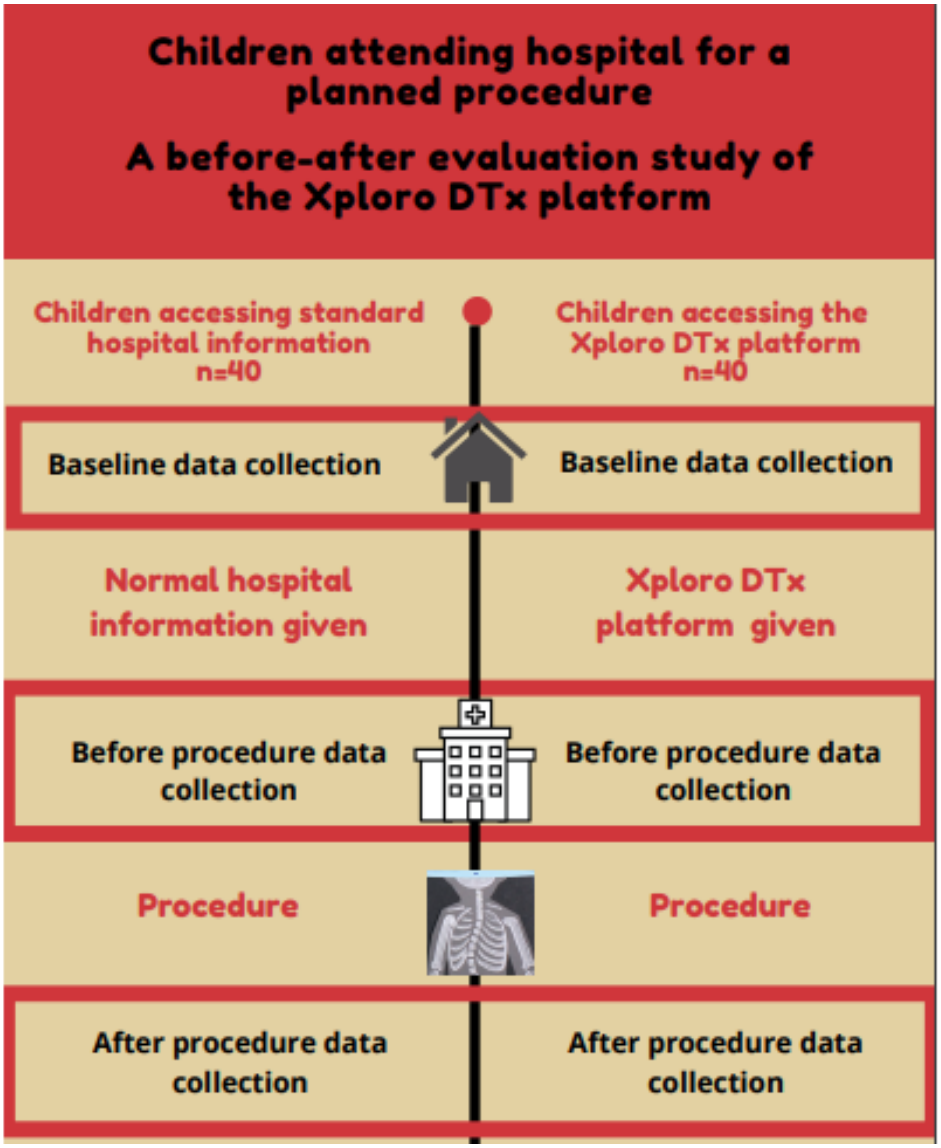


Table 1. Measures used in the study.

Child measures	Parent measures	Points of measurement
Trait anxiety measure		
• "How you would usually feel at home?" • 0 (completely relaxed and calm) to 10 (completely worried and anxious; adapted from Kleiber and McCarthy, 2006 [25])	• "How worried or anxious are you normally about things?" • 0 (not at all anxious or worried) to 10 (very anxious or worried) • Self-developed	• Baseline
Procedural anxiety measure		
• "How do you feel now?" • 0 (completely relaxed and calm) to 10 (completely worried and anxious; adapted from Kleiber and McCarthy, 2006 [25])	• "How anxious or worried are you about your child's procedure?" • 0 (not at all anxious or worried) to 10 (very anxious or worried) • Self-developed	• Baseline • Before the procedure
Perception of procedural knowledge measure		
• "How much you know or understand about what will happen to you today at the hospital?" • 0 (know nothing) to 10 (know everything) • Self-developed	• "How much do you know about your child's procedure?" • 0 (know nothing) to 10 (know everything) • Self-developed	• Baseline • Before the procedure
Procedural satisfaction measure		
• "How satisfied or happy you are with what happened today?" • 0 (not at all happy/satisfied) to 10 (completely happy/satisfied) • Self-developed	• "Overall how satisfied were you with your child's procedure today?" • 0 (not at all satisfied) to 10 (completely satisfied; adapted from Spencer and Franck, 2005 [26])	• After the procedure
Procedural involvement measure		
• "Were you involved, as much as you wanted to be, in decisions about your procedure?" • a. 0: No b. 1: Yes, sort of c. 2: Yes, definitely • Children could also respond: "I did not want or need to be involved and these responses were not included in the analyses"; adapted from Toomey et al, 2015 [27])	• N/A[a]	• After the procedure
Intervention engagement measure		
• "Which parts of the app did you look at /go through?" (Tick against a list of the different elements) • "Did you like this part?" (yes/no)	• N/A[a]	• After the procedure

[a]N/A: not applicable.

Ethics Approval

Ethics approval was received from the authors' institution (FOH 194) and the Health Research Authority (18/WA/0277).

Statistical Analysis

Independent *t* tests were conducted to compare parents and children's scores on the following measures between the 2 groups (those using the DTx platform intervention and those using standard procedural information when attending a hospital for a planned procedure): trait anxiety, procedural anxiety, perception of procedural knowledge and procedural satisfaction.

Where significant differences were found between the standard and intervention groups, two-tailed paired *t* tests were performed to determine whether there were any differences in parents and children's scores from baseline to before the procedure. To examine those who benefited the most from Xploro, the following exploratory group comparisons were conducted: (1) those who had invasive procedures versus those who had noninvasive procedures and (2) children who had 3 or more visits to the hospital and those who had fewer than 3 visits to the hospital.

Descriptive statistics have been presented regarding engagement with the intervention.

All analyses were conducted using SPSS version 25 (IBM Corp), and P=.04 was considered statistically significant.

Qualitative Analysis

The text responses of children and their parents were inductively analyzed by 2 researchers using content analysis processes [28], where the responses were coded and then organized into broad themes.

Results

Participant Characteristics

Of the 80 children and parents eligible to participate, all were successfully recruited, and all of them completed the baseline, preprocedural, and postprocedural data collection processes. Participants were aged between 8 and 14 years (average 10.5 years for the standard group and 12 years for the intervention group) and were attending a hospital for a range of procedures. The characteristics between the 2 groups were similar, for example, age, gender, previous hospital experience, and parental education (Table 2).

Table 2. Characteristics of the participants.

Characteristics	Usual hospital information (before) group, n (%)	Xploro digital therapeutic platform (intervention) group, n (%)
Child's age, (years)		
8-10	16 (40)	20 (50)
11-14	24 (60)	20 (50)
Gender		
Male	20 (50)	16 (40)
Female	20 (50)	24 (60)
Type of procedure		
Noninvasive (x-ray and ultrasound)	14 (35)	9 (23)
Invasive (surgery cannulation and blood tests)	26 (65)	31 (78)
Previous hospital procedure		
<3 previous hospital experiences	22 (55)	26 (65)
>3 previous hospital experiences	18 (45)	14 (35)
Parent educational level		
Primary	1 (3)	0 (0)
Secondary	22 (55)	27 (68)
Graduate	12 (30)	11 (28)
Postgraduate	4 (10)	2 (5)

Parents and Children's Levels of Trait Anxiety

At baseline, levels of general anxiety were similar for both parents and children in the 2 groups. Specifically, parents' levels of anxiety did not differ between the standard (mean 2.15, SD 0.770) and intervention groups (mean 2.10, SD 0.709; t_{78}=0.302; P=.76). Similarly, children's levels of trait anxiety did not differ between the standard (mean 2.35, SD 0.700) and intervention groups (mean 2.38, SD 0.49; $t_{69.85}$=−1.85; P=.85).

Feeling Less Worried: Reducing Procedural Anxiety

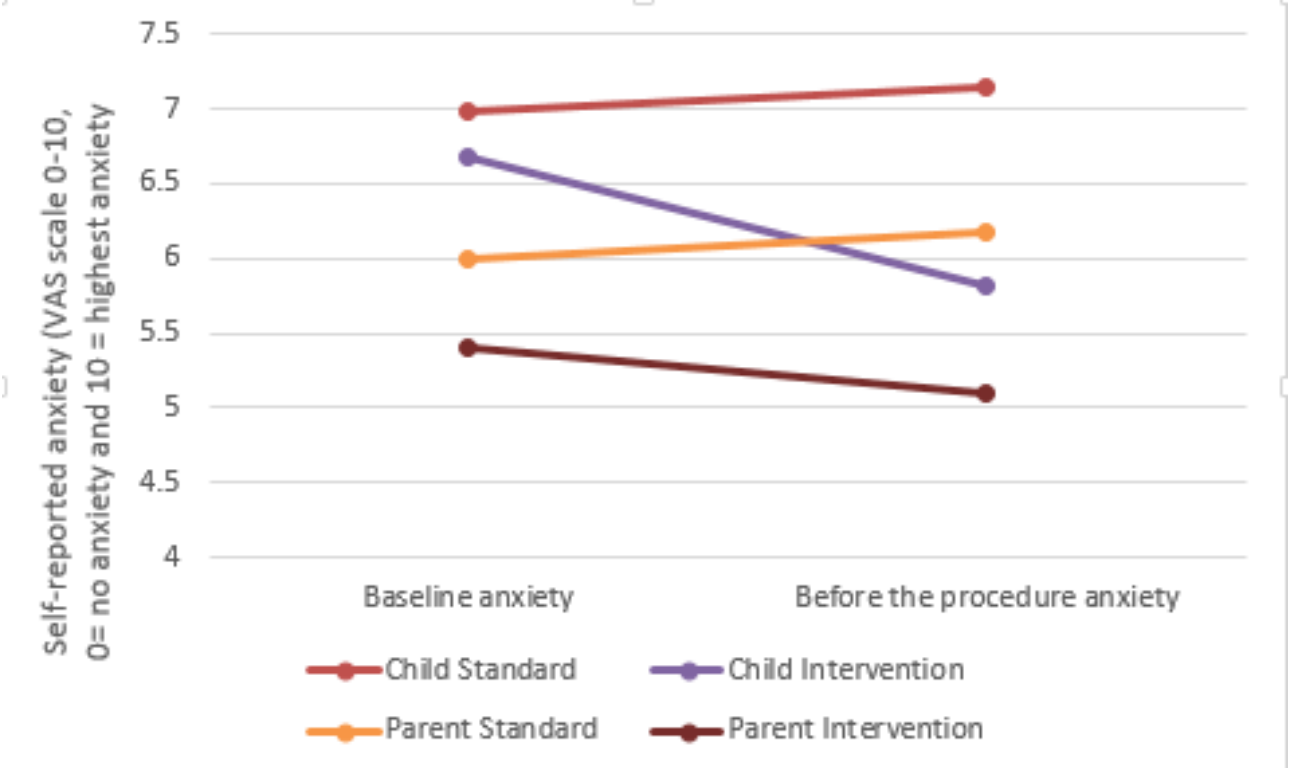
This study aimed to examine the impact of the intervention (DTx platform) compared with standard procedural information on children and their parents' self-reported procedural anxiety when attending a hospital for a planned procedure. Children and parents were asked to rate their procedural anxiety on a single 10-point VAS.

At baseline, there were no differences in procedural anxiety between the standard and intervention groups for both parents and children (Table 3). However, before the procedure, the mean procedural anxiety scores of both children and parents in the intervention group were significantly lower than those in the standard group (Figure 3).

Table 3. Procedural anxiety levels of children and their parents in the standard and intervention groups.

Timepoint	Children				Parents			
	Standard, mean (SD)	Intervention, mean (SD)	*t* value (*df*)	*P* value	Standard, mean (SD)	Intervention, mean (SD)	*t* value (*df*)	*P* value
Baseline	6.98 (2.70)	6.68 (1.51)	0.613 (78)	.54	6.00 (2.97)	5.40 (2.193)	1.056 (78)	.31
Before procedure	7.15 (2.63)	5.82 (1.57)	2.740 (63.64)	.008	6.18 (2.836)	5.10 (1.919)	1.985 (68.51)	.05

Figure 3. Self-reported procedural anxiety.



We then conducted two-tailed paired *t* tests to determine whether there were any differences in parents and children's levels of procedural anxiety from baseline to before the procedure, depending on the nature of the procedure (eg, invasive vs noninvasive procedure). The only statistically significant difference in scores occurred in the intervention group. For children undergoing invasive procedures, levels of anxiety decreased from baseline (mean 6.61, SD 1.50) to before the procedure (mean 5.84, SD 1.54; t_{30}=2.555; *P*=.02). Similarly, for children undergoing noninvasive procedures, anxiety decreased from baseline (mean 6.89, SD 1.62) to preprocedure (mean 5.78, SD 1.79; t_8=2.857; *P*=.02). For parents, anxiety decreased from baseline (mean 5.71, SD 2.27) to before the procedure (mean 5.32, SD 2.02) for those whose children were undergoing invasive procedures only (t_{30}=2.834; *P*=.008). The Xploro DTx platform therefore reduced children's levels of procedural anxiety regardless of the kind of procedure they were undergoing, whereas for parents, the intervention was most beneficial to those whose children were undergoing invasive procedures.

We then examined whether there were any differences in parents and children's levels of procedural anxiety from baseline to before the procedure, depending on the number of times children reported previous visits to the hospital (eg, less than 3 or more than 3). The only statistically significant difference was found in the intervention group. For those who had visited the hospital 3 times or less, children's levels of anxiety decreased from baseline (mean 6.77, SD 1.31) to preprocedure (mean 5.54, SD 1.30; t_{25}=7.273; *P*<.001). The parents of children who had visited the hospital 3 times or less also experienced a reduction in anxiety from baseline (mean 5.54, SD 1.79) to preprocedure (mean 5.15, SD 1.52; t_{25}=2.813; *P*=.009). Therefore, in terms of anxiety, Xploro benefitted children and parents who had less exposure to hospital environments compared with those who had more exposure.

Feeling I Know More: Increasing the Perception of Procedural Knowledge

This study aims to explore the impact of using a DTx platform compared with standard procedural information on children and their parents' self-reported perception of procedural knowledge**.** Self-reported procedural knowledge refers to how much knowledge the child or parent believed they had about the procedure, and this was assessed on a single VAS.

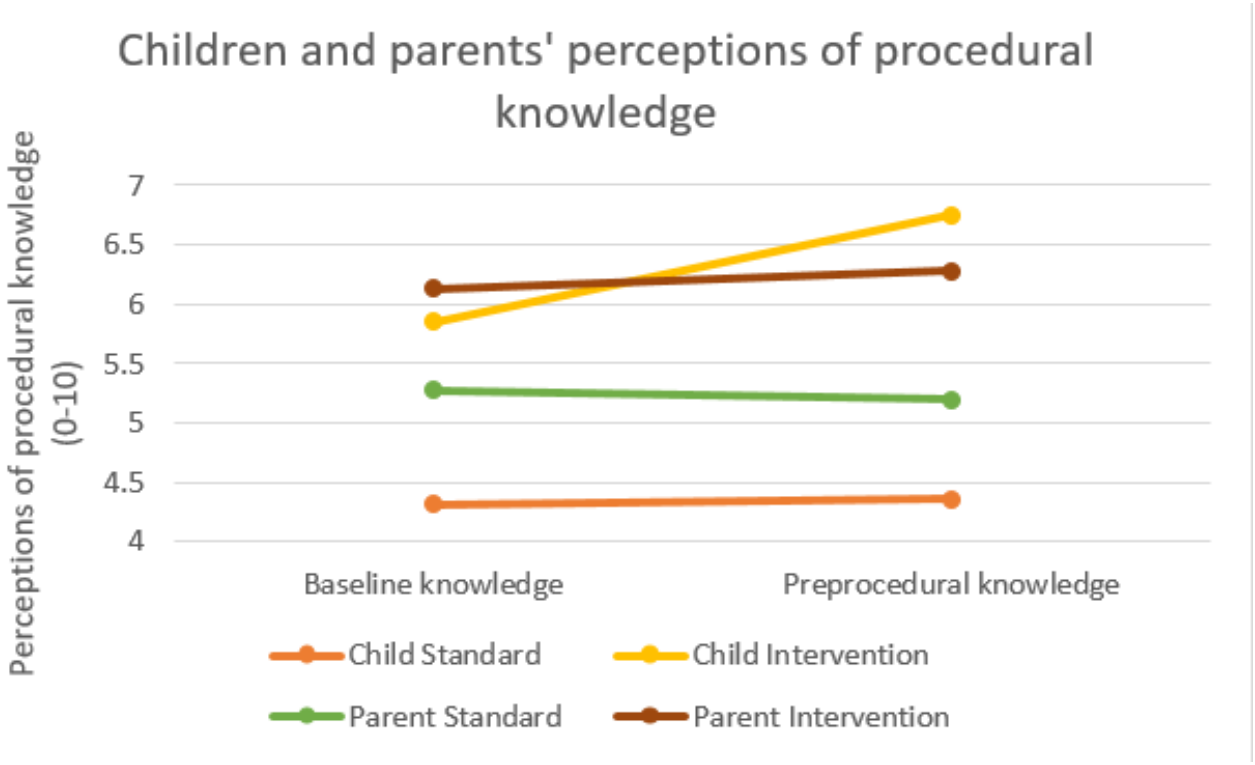
At baseline, children in the intervention group rated themselves as having significantly higher levels of perceived procedural knowledge compared with children in the control group (Table 4; Figure 4). Similarly, parents perceived procedural knowledge

at baseline was higher in the intervention group than in the standard group, although this difference was only marginally significant (Table 4). These findings were unexpected given that the children and parents in the intervention group did not have access to the DTx platform when the baseline measurements were obtained.

Table 4. Perceived procedural knowledge of children and their parents in the standard and intervention groups.

Timepoint	Children				Parents			
	Standard, mean (SD)	Intervention, mean (SD)	*t* value (*df*)	*P* value	Standard, mean (SD)	Intervention, mean (SD)	*t* value (*df*)	*P* value
Baseline	4.32 (2.66)	5.85 (1.83)	−2.866 (78)	.006	5.28 (2.29)	6.13 (1.70)	−1.888 (78)	.06
Before procedure	4.36 (2.66)	6.75 (1.51)	−4.892 (59.98)	<.001	5.200 (2.267)	6.28 (1.26)	−2.621 (78)	.01

Figure 4. Children and parents' reported perceptions of procedural knowledge.



We then conducted a two-tailed paired *t* test to determine whether there were any differences in parents and children's levels of perceived procedural knowledge from baseline to before the procedure, depending on the nature of the procedure (eg, invasive procedure vs noninvasive procedure). The only statistically significant difference in scores occurred in the intervention group. For children undergoing an invasive procedure, there was a statistically significant increase in perceptions of procedural knowledge from baseline (mean 6.10, SD 1.81) to before the procedure (mean 6.78, SD 1.38; t_{30}=−3.760; *P*=.001). Similarly, for children undergoing noninvasive procedures, there was a statistically significant increase in perceptions of procedural knowledge from baseline (mean 5.00, SD 1.73) to before the procedure (mean 6.33, SD 1.94; t_8=−2.412; *P*=.04). Xploro was therefore of the greatest benefit to children, regardless of the kind of procedure under.

We then examined whether there were any differences in parents and children's levels of perceptions of procedural knowledge from baseline to before the procedure, depending on the number of times the children had visited the hospital (eg, less than 3 or more than 3). The only statistically significant difference was found in the intervention group. For children who had fewer than 3 visits, there was a statistically significant increase in perceived procedural knowledge from baseline (mean 5.69, SD 1.81) to before the procedure (mean 6.54, SD 1.45; t_{25}=−3.528; *P*=.002). Similarly, for those who had 3 or more visits, there was a statistically significant increase in perceived procedural knowledge from baseline (mean 6.14, SD 1.92) to preprocedure (mean 7.14, SD 1.61; t_{13}=−2.646; *P*=.02). In terms of procedure knowledge, the Xploro DTx platform therefore benefitted children and parents perceived procedural knowledge regardless of their previous exposure to a hospital.

Feeling Happier About What Happened: Procedural Satisfaction

Children and their parents' self-reported satisfaction of undergoing the hospital procedure was rated after the procedure on a single 10-point VAS. Both parents and children's procedural satisfaction scores were higher in the intervention group compared with the standard group, although these differences were not statistically significant (Table 5).

Table 5. Procedural satisfaction of children and their parents in the standard and intervention groups.

Timepoint	Children				Parents			
	Standard, mean (SD)	Intervention, mean (SD)	*t* value (*df*)	*P* value	Standard, mean (SD)	Intervention, mean (SD)	*t* value (*df*)	*P* value
After the procedure	5.98 (2.787)	6.88 (2.002)	-1.659 (70.79)	.10	6.63 (2.457)	6.80 (1.800)	-0.363 (78)	.72

Feeling More Involved: Increasing Procedural Involvement

The study aimed to understand whether children and parents who received the intervention would report higher levels of procedural involvement. Children were asked after the procedure to rate their perceived levels of involvement on a 3-point Likert scale. After the procedure, the mean procedural involvement scores of children in the intervention group were significantly higher than those in the standard group (t_{75}=−2.238; *P*=.03).

Using It and Liking It: Children's Engagement With the Intervention

This study also aimed to understand the levels of engagement a child had with the intervention (DTx platform). Although software was not used to determine clear statistical levels of engagement, the questionnaire booklets asked the children to retrospectively record their levels of engagement with the intervention. This enabled us to gain some understanding of which parts of the DTx platform were used. All of the children who accessed and used the intervention valued the content and enjoyed the various components, particularly the customized avatar and chatbot. Table 6 identifies the specific components of the intervention that the children used.

The children were asked to rate how much they liked the different components of Xploro that they used. Only 20 out of the 40 children in the intervention group completed this section, but those who did reported that they liked all the different components they accessed.

Table 6. Children's engagement with the different components of the intervention (n=40).

Different components of the intervention	Children reporting using each component, n (%)
Avatar	40 (100)
Chatbot	40 (100)
Ward orientation	36 (90)
Equipments	
MRI[a]	22 (55)
Monitor	24 (60)
X-ray machine	24 (60)
LINAC[b]	24 (60)
CT[c] scan	24 (60)
Ultrasound	4 (10)
Who is who (staff)	28 (70)
Games	
Heart race	28 (70)
Germ buster	28 (70)
And relax	26 (65)
Operating room	30 (75)
Anesthesia room	30 (75)
Ward	39 (98)
Recovery room	25 (63)

[a]MRI: magnetic resonance imaging.

[b]LINAC: linear particle accelerator.

[c]CT: computed tomography.

Qualitative Experiences of Using the Intervention

It Helped Me to Understand What Would Happen: Experiences of Children Using the Intervention

During the semistructured *write and tell* interviews, the children stated that they enjoyed using the intervention and felt that it had positively influenced their experiences of coming to the hospital for a procedure. The interviews demonstrated that the intervention benefitted the children by providing them with information and knowledge about the procedure they were about to undergo while at home "it is so good, like being to see the hospital while you are still at home" (P24). One child stated that the DTx platform helped them "know things I didn't know before, I didn't know all the tests would be from one needle I thought it was loads" (P12), while another talked about how the intervention "is great, it taught me lots and I knew what all the numbers on the machine were and made me feel less worried" (P3). Even children who had previous experience with a procedure valued the information: "It [Xploro] is good, I have had an MRI before but I didn't know what it was doing while I was in it–but now I know" (P1).

The children also reported using the intervention to distract themselves and take their minds off their procedure. This was particularly linked to the games that were viewed positively by the children: "they were fun and helped to distract me" (P37); "I liked playing the games on the iPad" (P36); "it helped when I was having my stitches out as I could play on it and not think about it" (P21).

The interviews also helped to identify the important roles that parents play in facilitating or disabling children's access to information: "my mum only let me play on it last night" (P20); "if it was a game in real life I would ask my mum for it" (P14).

It Helped Us Talk About What Would Happen: Experiences of Parents of Their Child Using the Intervention

The parents discussed how their child had enjoyed using Xploro and how the intervention had helped provide children with information about the procedure: "it helped him learn lots about the scan" (P13) and "it [Xploro] is great, she now knows more than me" (P3); parents described how without the intervention their child may have been lacking in information "the nurse didn't explain so without it [Xploro] she wouldn't have had a clue" (P19). The increase in knowledge was reported by parents as reducing their child's anxiety "she feels more relaxed as she knows what things look like and knows about the anaesthetic machine" (P39).

Parents also reported that the intervention helped them talk about the planned procedure with their child, "playing with the app allowed us to talk about what would happen, otherwise I wouldn't know how to approach it" (P1). Despite the intervention being focused on providing information to children, some parents found it really useful to them, "It [Xploro] is amazing, I used it more than him" (P7), although some parents did not appreciate that the focus of the intervention was on information provision as well as being fun, "if I had known it was about learning and not just games I would have let her use it sooner" (P20). The engaging components were viewed positively by parents, "she liked the interactive bits, it looks good" (P37) and "he loved making the man" (P7).

Discussion

Principal Findings

We report the results of a before and after evaluation study to examine the acceptability of a DTx platform (Xploro) and the impact of this intervention on children undergoing a clinical procedure and their parents. To the best of our knowledge, this is the first study to evaluate the impact of a DTx platform using gamification, serious games, a chatbot, and an augmented reality avatar with children undergoing a wide range of hospital procedures. This study provides preliminary evidence that the intervention (Xploro) reduced procedural anxiety of children and their parents and improved children's perceptions of procedural knowledge and involvement. All the children who accessed and used the DTx platform described how they valued the content, enjoyed using the various components, and reported that it improved their procedural experience by helping them be more prepared, less anxious, and more distracted. These findings suggest that the Xploro intervention has the potential to improve children's procedural health literacy and address the need for meaningful and accessible procedural information. This is the first study to focus on the influence of a digital health intervention on children's procedural health literacy. To our knowledge, this is the first study to evaluate a DTx platform by measuring multiple concepts including children's procedural anxiety, perceptions of knowledge, procedural satisfaction, and perceived procedural involvement.

The DTx platform provides information about a broad range of hospital procedures, environments, and professionals; this is important as children can often encounter many different experiences when visiting the hospital, which may not be addressed by current digital health education and preparation interventions that focus on one specific procedure, for example, admission for surgery [13-16,29], or are designed around one specific hospital setting [30]. These interventions are of value but may not address the questions and concerns children have about procedures identified in previous research [9].

Children particularly valued that the Xploro intervention enabled them to customize and individualize their learning experiences; they could access different components of Xploro based on their individual preferences and information needs. Previous research indicates that children wish to receive information specific to them and are more likely to engage with and understand information delivered in a child-centered way [10]. It is increasingly recognized within the literature, particularly in relation to children, that health literacy is facilitated when an individual can tailor information to their needs and have the opportunity to process, question, and apply information to their individual circumstances [9] and actively construct knowledge and understanding [31]. This tailoring and individualization of information takes time, and interventions need to assist and support children in doing this. The Xploro intervention, underpinned by an *information-pull* approach, facilitated children to determine and address their own information needs

and actively engage with the multiple elements of the intervention to shape their own learning experience. Digital health interventions that children can use flexibly and actively to pace their own learning are potentially of most value, based on evidence showing how children require different procedural information and education depending on their cognitive development and a previous experience [32]. Studies exploring the use of virtual reality interventions have also highlighted the importance of active user engagement and how passive interaction with an intervention is linked to poorer impact and experience [33].

There is an increasing need for high-quality and evidence-based digital solutions to the National Health Service's challenges [34]. This is against the backdrop of the need for improved innovation and implementation [24] to meet these health care challenges. The team are committed to developing a digital health intervention for children, which has been rigorously developed and evaluated based on a child-centered approach. This study reports on the findings of a before and after study, which provides evidence that a larger effectiveness study would be of value.

Strengths and Limitations

One of the main strengths of the study were the methods used to ensure children's views and self-reported experiences were central to the investigation. Further strengths are the similarities in the demographic characteristics between the intervention and standard care groups and the demonstrated suitability of the recruitment and data collection processes.

A potential limitation is that the study participants were not randomly allocated to the different groups, and the study was conducted in one hospital setting. The analyses presented are exploratory in nature. Given the number of statistical tests that have been conducted and the relatively modest sample size, generalizations should be made with caution. The study only used single-item scales to measure the concepts of procedural anxiety, perception of knowledge, satisfaction, and procedural involvement; the use of multi-item scales may have provided more robust data. Future studies should validate these preliminary findings in a larger controlled study, which also examines the implementation of the intervention within clinical services.

Conclusions

Large numbers of children undergo clinical procedures every day within a hospital setting. Many of these children are currently not well-prepared or well-informed about what will happen, which can lead to high levels of procedural anxiety and distress. This study has shown that the Xploro DTx platform improved children's reported procedural involvement and perceived procedural knowledge. The intervention reduced both parents and children's levels of procedural anxiety before the procedure. The children and parents in the intervention group described Xploro as easy to use and fun, which helped them to know more before the planned procedure and how to cope during the procedure.

Acknowledgments

This work was supported by an Innovate UK grant awarded to the lead partner Corporation Pop, developers of the Xploro app.

Conflicts of Interest

None declared.

References

1. Kain ZN, Mayes LC, Caldwell-Andrews AA, Karas DE, McClain BC. Preoperative anxiety, postoperative pain, and behavioral recovery in young children undergoing surgery. Pediatrics 2006 Aug;118(2):651-658. [doi: 10.1542/peds.2005-2920] [Medline: 16882820]
2. Claar RL, Walker LS. Functional assessment of pediatric pain patients: psychometric properties of the functional disability inventory. Pain 2006 Mar;121(1-2):77-84 [FREE Full text] [doi: 10.1016/j.pain.2005.12.002] [Medline: 16480823]
3. Koukourikos K, Tzeha L, Pantelidou P, Tsaloglidou A. The importance of play during hospitalization of children. Mater Sociomed 2015 Dec;27(6):438-441 [FREE Full text] [doi: 10.5455/msm.2015.27.438-441] [Medline: 26889107]
4. Rodriguez EM, Dunn MJ, Zuckerman T, Vannatta K, Gerhardt CA, Compas BE. Cancer-related sources of stress for children with cancer and their parents. J Pediatr Psychol 2012 Mar;37(2):185-197 [FREE Full text] [doi: 10.1093/jpepsy/jsr054] [Medline: 21841187]
5. Duff AJ, Gaskell SL, Jacobs K, Houghton JM. Management of distressing procedures in children and young people: time to adhere to the guidelines. Arch Dis Child 2012 Jan;97(1):1-4. [doi: 10.1136/archdischild-2011-300762] [Medline: 22082500]
6. Bray L, Carter B, Snodin J. Holding children for clinical procedures: perseverance in spite of or persevering to be child-centered. Res Nurs Health 2016 Feb;39(1):30-41. [doi: 10.1002/nur.21700] [Medline: 26575179]
7. Kahana SY, Feeny NC, Youngstrom EA, Drotar D. Posttraumatic stress in youth experiencing illnesses and injuries: an exploratory meta-analysis. Traumatology 2006;12(2):148-161. [doi: 10.1177/1534765606294562]
8. Smith L, Callery P. Children's accounts of their preoperative information needs. J Clin Nurs 2005 Feb;14(2):230-238. [doi: 10.1111/j.1365-2702.2004.01029.x] [Medline: 15669932]

9. Bray L, Appleton V, Sharpe A. The information needs of children having clinical procedures in hospital: will it hurt? Will I feel scared? What can I do to stay calm? Child Care Health Dev 2019 Sep;45(5):737-743 [FREE Full text] [doi: 10.1111/cch.12692] [Medline: 31163093]
10. Bray L, Appleton V, Sharpe A. 'If I knew what was going to happen, it wouldn't worry me so much': children's, parents' and health professionals' perspectives on information for children undergoing a procedure. J Child Health Care 2019 Dec;23(4):626-638. [doi: 10.1177/1367493519870654] [Medline: 31431048]
11. Tsao Y, Kuo H, Lee H, Yiin S. Developing a medical picture book for reducing venipuncture distress in preschool-aged children. Int J Nurs Pract 2017 Oct;23(5):e12569. [doi: 10.1111/ijn.12569] [Medline: 28726263]
12. Fortier MA, Kain ZN. Treating perioperative anxiety and pain in children: a tailored and innovative approach. Paediatr Anaesth 2015 Jan;25(1):27-35 [FREE Full text] [doi: 10.1111/pan.12546] [Medline: 25266082]
13. Eijlers R, Dierckx B, Staals LM, Berghmans JM, van der Schroeff MP, Strabbing EM, et al. Virtual reality exposure before elective day care surgery to reduce anxiety and pain in children: a randomised controlled trial. Eur J Anaesthesiol 2019 Oct;36(10):728-737 [FREE Full text] [doi: 10.1097/EJA.0000000000001059] [Medline: 31356373]
14. Kim J, Chiesa N, Raazi M, Wright KD. A systematic review of technology-based preoperative preparation interventions for child and parent anxiety. Can J Anaesth 2019 Aug;66(8):966-986. [doi: 10.1007/s12630-019-01387-8] [Medline: 31098960]
15. Buffel C, van Aalst J, Bangels A, Toelen J, Allegaert K, Verschueren S, et al. A web-based serious game for health to reduce perioperative anxiety and pain in children (CliniPup): pilot randomized controlled trial. JMIR Serious Games 2019 Jun 1;7(2):e12431 [FREE Full text] [doi: 10.2196/12431] [Medline: 31199324]
16. Ryu J, Park S, Park J, Kim J, Yoo H, Kim T, et al. Randomized clinical trial of immersive virtual reality tour of the operating theatre in children before anaesthesia. Br J Surg 2017 Nov;104(12):1628-1633. [doi: 10.1002/bjs.10684] [Medline: 28975600]
17. Liszio S, Masuch M. Virtual Reality MRI: Playful Reduction of Children's Anxiety in MRI Exams. In: Proceedings of the 2017 Conference on Interaction Design and Children. 2017 Presented at: IDC'17; June 27-30, 2017; Stanford, CA, USA. [doi: 10.1145/3078072.3079713]
18. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
19. Peerson A, Saunders M. Health literacy revisited: what do we mean and why does it matter? Health Promot Int 2009 Sep;24(3):285-296. [doi: 10.1093/heapro/dap014] [Medline: 19372101]
20. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new medical research council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
21. Noonan RJ, Boddy LM, Fairclough SJ, Knowles ZR. Write, draw, show, and tell: a child-centred dual methodology to explore perceptions of out-of-school physical activity. BMC Public Health 2016 Apr 14;16:326 [FREE Full text] [doi: 10.1186/s12889-016-3005-1] [Medline: 27080384]
22. Mannell J, Davis K. Evaluating complex health interventions with randomized controlled trials: how do we improve the use of qualitative methods? Qual Health Res 2019 Apr;29(5):623-631. [doi: 10.1177/1049732319831032] [Medline: 30871431]
23. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
24. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
25. Kleiber C, McCarthy AM. Evaluating instruments for a study on children's responses to a painful procedure when parents are distraction coaches. J Pediatr Nurs 2006 Apr;21(2):99-107. [doi: 10.1016/j.pedn.2005.06.008] [Medline: 16545670]
26. Spencer C, Franck LS. Giving parents written information about children's anesthesia: are setting and timing important? Paediatr Anaesth 2005 Jul;15(7):547-553. [doi: 10.1111/j.1460-9592.2005.01500.x] [Medline: 15960637]
27. Toomey SL, Zaslavsky AM, Elliott MN, Gallagher PM, Fowler FJ, Klein DJ, et al. The development of a pediatric inpatient experience of care measure: child HCAHPS. Pediatrics 2015 Aug;136(2):360-369 [FREE Full text] [doi: 10.1542/peds.2015-0966] [Medline: 26195542]
28. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
29. Liguori S, Stacchini M, Ciofi D, Olivini N, Bisogni S, Festini F. Effectiveness of an app for reducing preoperative anxiety in children: a randomized clinical trial. JAMA Pediatr 2016 Aug 1;170(8):e160533. [doi: 10.1001/jamapediatrics.2016.0533] [Medline: 27294708]
30. Ask Oli & Alder Play App. United Kingdom National Health Service. URL: https://alderhey.nhs.uk/parents-and-patients/alder-play-app [accessed 2020-03-23]

31. Hirsh-Pasek K, Zosh JM, Golinkoff RM, Gray JH, Robb MB, Kaufman J. Putting education in 'educational' apps: lessons from the science of learning. Psychol Sci Public Interest 2015 May;16(1):3-34. [doi: 10.1177/1529100615569721] [Medline: 25985468]
32. Jaaniste T, Hayes B, von Baeyer CL. Providing children with information about forthcoming medical procedures: a review and synthesis. Clin Psychol Sci Pract 2007 Jun;14(2):124-143. [doi: 10.1111/j.1468-2850.2007.00072.x]
33. Ahmadpour N, Keep M, Janssen A, Rouf AS, Marthick M. Design strategies for virtual reality interventions for managing pain and anxiety in children and adolescents: scoping review. JMIR Serious Games 2020 Jan 31;8(1):e14565 [FREE Full text] [doi: 10.2196/14565] [Medline: 32012042]
34. A Digital NHS? An Introduction to the Digital Agenda and Plans for Implementation. The King's Fund. 2016. URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/A_digital_NHS_Kings_Fund_Sep_2016.pdf [accessed 2020-07-17]

Abbreviations

DTx: digital therapeutic
VAS: visual analog scale

Edited by G Eysenbach; submitted 10.12.19; peer-reviewed by J Salisbury, S Thiebes; comments to author 27.01.20; revised version received 16.03.20; accepted 13.06.20; published 11.08.20.

Please cite as:
Bray L, Sharpe A, Gichuru P, Fortune PM, Blake L, Appleton V
The Acceptability and Impact of the Xploro Digital Therapeutic Platform to Inform and Prepare Children for Planned Procedures in a Hospital: Before and After Evaluation Study
J Med Internet Res 2020;22(8):e17367
URL: http://www.jmir.org/2020/8/e17367/
doi:10.2196/17367
PMID:32780025

Original Paper

Converting Visitors of Physicians' Personal Websites to Customers in Online Health Communities: Longitudinal Study

Qin Chen[1], MSc; Xiangbin Yan[1], PhD; Tingting Zhang[1], PhD

School of Economics and Management, University of Science and Technology Beijing, Beijing, China

Corresponding Author:
Tingting Zhang, PhD
School of Economics and Management
University of Science and Technology Beijing
No 30 Xueyuan Road
Haidian District
Beijing, 100083
China
Phone: 86 18612627077
Email: tzhang@ustb.edu.cn

Abstract

Background: With the dramatic development of Web 2.0, increasing numbers of patients and physicians are actively involved in online health communities. Despite extensive research on online health communities, the conversion rate from visitor to customer and its driving factors have not been discussed.

Objective: The aim of this study was to analyze the conversion rate of online health communities and to explore the effects of multisource online health community information, including physician-generated information, patient-generated information, and system-generated information.

Methods: An empirical study was conducted to examine the effects of physician-generated, patient-generated, and system-generated information on the conversion rate of physicians' personal websites by analyzing short panel data from 2112 physicians over five time periods in a Chinese online health community.

Results: Multisource online health community information (ie, physician-generated, patient-generated, and system-generated information) positively affected the conversion rate. Physician-generated and patient-generated information showed a substitute relationship rather than a complementary relationship. In addition, the usage time of a personal website positively moderated patient-generated information, but negatively moderated physician-generated information.

Conclusions: This study contributes to the electronic health literature by investigating the conversion rate of online health communities and the effect of multisource online health community information. This study also contributes to understanding the drivers of conversion rate on service websites, which can help to successfully improve the efficiency of online health communities.

(J Med Internet Res 2020;22(8):e20623) doi:10.2196/20623

KEYWORDS

online health communities; conversion rate; multisource information; physician-generated information; patient-generated information; system-generated information; usage time

Introduction

Background

Medical resources are both limited and irrationally distributed [1]. Large cities have the most medical services, causing many patients in small cities and countries to struggle for basic health care [2]. With the support of Health 2.0 technologies, an increasing number of internet users have realized that the internet is useful for acquiring knowledge and information on diseases and treatments [3], which has stimulated the development of online health communities (OHCs). OHCs can be divided into online physician-patient communities and online patient-patient communities according to the user composition and the communication model [4]. In this study, we focused on an online physician-patient community (hereafter referred to generally as OHC). An OHC is a platform on which patients are now able to consult physicians on health issues and the

treatment of their illness anytime and anywhere. Current evidence shows that OHCs have considerably significant effects in terms of reducing medical costs, improving operational efficiency and effectiveness, enhancing the equity of medical resources, and meeting patient satisfaction [5].

Customers increasingly prefer the internet to purchase products and services, and a digital channel has become an important platform for sellers of different sectors [6]. Numerous technologies and strategies are utilized to increase website traffic; however, an increase in website traffic alone does not guarantee an increase in sales. It is also important to convert visitors into customers [7,8]. Despite the importance of the conversion rate, research on understanding and analyzing the conversion rates of service websites is limited, especially in OHCs. Service has three well-documented features: intangibility, heterogeneity, and inseparability [9]. First, most services cannot be counted, measured, inventoried, tested, and verified in advance of the sale to assure quality. Second, performance often varies from producer to producer, from customer to customer, and from day to day. Third, production and consumption of services are inseparable. Unlike the conversion rate of other types of websites, the conversion rate for a service website represents a proportion of customers to total visitors who successfully locate information [10]. In the process of choosing a physician in an OHC, the patient decides whether or not to consult online or offline after visiting a physician's personal website [11,12], which converts the visitor of a personal website to a customer. Therefore, the conversion rate of an OHC can be defined as the proportion of customers to visitors who successfully locate information on physicians' personal websites and decide to consult. Therefore, how to convert visitors of a personal website to customers is an essential issue for physicians and the manager of the OHC.

Customers rely on several different sources to gather information on a variety of products or services prior to purchase [13-15]. Similarly, patients rely on OHC information from multiple sources to select a physician, and the physician's personal website gathers this OHC information (eg, online and offline information [4,5], individual and organizational information [16], and system-generated and patient-generated information [12]). Among them, system-generated information is created by the platform where physicians work, and reveals the service provider's contribution, grade, popularity, and other characteristics. This information is independent of physicians and patients in OHCs [12]. Patient-generated information is generated by patients who have experienced medical consultation (eg, reviews and ratings), which is typically used as a clue to reflect the outcomes of the physician's service [17-19]. System-generated information and patient-generated information have been shown to be relevant to patients' choices [12]. However, it is unclear whether this information will also affect the conversion rate of physicians' personal websites.

In addition, the information presented in a personal website also includes information generated by the physician's own online behaviors and activities (ie, physician-generated information) [18]. For example, the number of articles published by a physician can be used to measure their activeness [18], knowledge contribution [5], and professional capital [20]. In addition, physicians' online behaviors and activities are important cues to reflect their delivery process quality [12]. Patients tend to judge a physician's service quality from two dimensions: the outcomes and delivery process of the service [12], which correspond to the definitions of patient-generated information and physician-generated information that we applied in this study. Therefore, determining whether physician-generated information is related to the conversion rate of a personal website and its relationship with patient-generated information are worth studying.

System-generated information is automatically calculated by the system using an algorithm [12] and is updated at intervals, whereas patient-generated information and physician-generated information are recorded in the physician's personal website as of its launch. Thus, the amount of patient-generated and physician-generated information is related to the physician's usage time of the personal website. Moreover, the longer the physician uses the website, the more familiar they become with online medical services. Similar to patients who prefer experienced physicians in traditional medical services, patients may prefer to choose a physician who is more familiar with online medical services. Thus, a physician's usage time may affect the relationship between patient-generated information, physician-generated information, and the conversion rate of the personal website.

The main objective of this study was to investigate the impact of multisource OHC information (physician-generated information, patient-generated information, and system-generated information) on the conversion rate of physicians' personal websites, using data from a Chinese OHC. The main research questions were: (1) Could physician-generated information, patient-generated information, and system-generated information affect the conversion rate of physicians' personal websites in OHCs? (2) Could physician-generated and patient-generated information be used to complement the conversion rate of physicians' personal websites in OHCs? and (3) Does the usage time of a personal website moderate the effects of physician-generated and patient-generated information on the conversion rate? If so, to what degree (strengthen or weaken)?

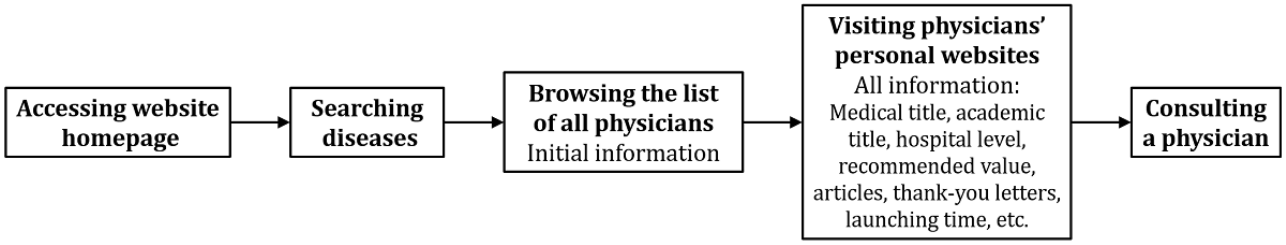
To answer these three research questions, we developed a research model, which was tested using data collected from 2112 physicians' personal websites over five time periods from the Chinese OHC Haodf.com [21].

OHCs

As Web 2.0 technologies are increasingly being used within the health care industry, patients and physicians have begun to participate more actively in health management [22]. Numerous studies have started to explore the benefits of OHCs from different perspectives. For example, Li et al [5] explored the economic success of physicians in the online consultation market. Guo et al [20] analyzed how physicians gained social and economic returns in OHCs. In contrast to traditional medical service delivery, OHCs give patients the opportunity to review the abundant amount of various types of information, and then use this information to choose the physician they wish to consult [23].

Specifically, the patient decision-making process in OHCs includes several stages and is also a process of gradually selecting physicians based on the information obtained, as shown in Figure 1. According to their diseases, patients can obtain a list of physicians that specialize in that disease and initial information to select alternatives, and then determine whether they should visit the personal website of the physician to get more information. When patients enter the physician's personal website, they can obtain more comprehensive information (eg, online behaviors, number of patients and visitors, articles) and decide whether or not to consult and how to consult [12]. Thus, a physician's personal website plays an important role in the patient decision-making process, and the OHC information presented determines whether a patient is only a visitor or can be a customer. Although previous research has explored the effects on patients' choices of different stages [12,18,24], there is currently no research available that focused on why some physicians are not selected for consultation even if their personal websites are visited by patients.

Figure 1. Patient decision-making process in online health communities.



Conversion Rate

In electronic commerce (e-commerce), conversion rates represent the proportion of orders to website visitors. For instance, if 100 people visit a website and 3 of them purchase the product, the conversion rate is 3.0% [25]. Prior research has shown that website factors, including website features [26,27] and functionalities [25], and customer factors, including customer reviews and sentiments [28], browsing, and purchasing experiences [29,30], have effects on the conversion rates of websites.

Compared with the conversion rates of other types of websites, research on the conversion rate of a service website is still in a preliminary stage. As service is intangible, inseparable, and heterogeneous [9], Jackson [10] defined the conversion rate of a service website as the proportion of the number of customers to the total number of visitors, and customers were defined as those who successfully locate the information that the service provider wanted them to find (eg, downloadable FAQ, emails the correct support address, or find the answer they were looking for). Cezar and Ögüt [7] analyzed the conversion rates in online hotel booking sites from the reviews of customers, recommendations, and ranking order of the system. Zhou et al [31] investigated the effect of posts generated by other members' activities in transforming visitors into members in online brand communities. However, there is no research on the conversion rate of OHCs.

Considering the great demand for medical services and limited medical resources [32], the efficiency of an OHC is crucial. Visiting a physician's personal website is a vital stage in the process of patient decision making, which is the prerequisite for converting visitors to physicians' personal websites into customers. The number of customers indicates the number of patients a physician has provided medical services for; that is, the number of visitors that have successfully converted. A higher conversion rate of a personal website means that more patients choose this physician for consultation, and this physician can also obtain more social and economic returns [5,20]. The increased conversion rate of all physicians' personal websites means that the efficiency of the OHC has increased. Therefore, the present study focused on the conversion rate of OHCs; that is, the proportion of customers to the total visitors of a physician's personal website.

Multisource Information in OHCs

When customers search for information to make critical purchase decisions, they often use multiple information sources [13-15], including electronic word-of-mouth sources, neutral/third party sources, and manufacturer/retailer sources. Similarly, patients selecting a physician in an OHC also need to search for information from multiple sources. Previous studies have investigated the link between multisource OHC information and patients' choices. For example, Li el al [18] explored the effects of patients' rating and physicians' activeness on the number of patients. Liu et al [16] found that the number of physicians' appointments was positively associated with their individual offline and online reputations, as well as their organizational offline and online reputations. Yang et al [12] categorized OHC information into system-generated and patient-generated information, and explored their effects on patients' online searches, evaluations, and decisions.

System-generated information represents pieces of information that are system- or machine-rendered [33]. This online information about products and services from neutral/third parties is considered to be more useful and objective [34], and provides references for customers (eg, ranking orders and recommendations) [35,36]. Matching a customer's preferences against the most similar customers' preferences facilitates the purchasing process, as it narrows alternatives down to a few and hence decreases the time spent searching [7]. System-generated information in an OHC is created by the system based on the service provider's contribution and popularity, which is independent of physicians and patients [12]. Patient-generated information is derived from the patient feedback channel of OHCs and is generated by patients who have experienced a medical consultation (eg, reviews and ratings). Generally, patient-generated information reflects the

outcomes of a physician's service [12], and is related to the perception of the physician's skill [17] and patients' choices at different stages [12,18,19].

However, an OHC is an online community that uses information technology to present a medical ecosystem including patients and physicians [4]. OHC information also includes information generated by physicians' online behaviors and activities [18] (ie, physician-generated information). In an OHC, physicians can provide additional types of consultation than offline hospitals or clinics [37] (eg, network consultation, telephone consultation, appointment registration). In addition, physicians can update personal information, publish articles, reply to consultations, and manage patients in their personal websites. These online behaviors and activities can generate considerable information. Previous studies have shown that the number of articles published by a physician indicates the degree of activeness, knowledge contribution, and professional capital, which affect patients' choices [18]. Therefore, this study focused on information from three different sources: physician-generated, patient-generated information, and system-generated information.

Research Model and Hypotheses

The customer decision-making process consists of five stages: needs recognition, information search, pre-evaluation, purchase decision, and postevaluation [38]. This is a process of continually eliminating options relying on multisource information, which is similar to the process of a patient selecting a physician in OHCs. As service is intangible, inseparable, and heterogeneous [9], patients narrow down the alternatives and finally choose one physician to consult by relying on multisource OHC information. Considering the importance of the physician's personal website for the patient decision-making process, in this study, we focused on the roles of the multisource OHC information (physician-generated information, patient-generated information, and system-generated information) in the process of converting visitors to customers.

Physician-Generated Information and Conversion Rate

As the delivery of physicians' services is based on the platform, the information of physicians' behaviors and activities is considerable. The more information the physician generates, the more active the physician is on the platform [18]. Conversely, less available physician-generated information represents the behavior of inactive physicians. This kind of information can therefore reveal the delivery process quality of the physician's service [12]. Patients prefer physicians who are actively involved online and have a high-quality service delivery process [12,18], and physicians' online behaviors and activities have been shown to affect outpatient visits in hospitals [39]. When patients see a large amount of physician-generated information on a physician's personal website, they will be more willing to convert as a customer. Therefore, it is hypothesized that *physician-generated information positively affects the conversion rate of a physician's personal website (H1).*

Patient-Generated Information and Conversion Rate

A physician's personal website also serves as a feedback channel for patients. Feedback from other people who have experienced a certain quality is a useful signal to help receivers understand quality, and is of vital importance in affecting information for receivers' decision making [40]. Patient-generated information (eg, ratings and reviews) comes from other patients who have consulted the physician. Therefore, a greater amount of patient-generated information in a physician's personal website indicates a higher number of patients who have selected the physician for consultation. Moreover, patient-generated information reflects a physician's service outcomes [12]. Such information is more objective and credible than traditional information from friends and family members [41]. Patients prefer to choose a physician who has been consulted by a large number of patients [12] and delivers high-quality service outcomes. Therefore, it is hypothesized that: *Patient-generated information positively affects the conversion rate of a physician's personal website (H2).*

The coexistence of physician-generated and patient-generated information may cause both items to complement each other in driving patients to choose a consultation after visiting a physician's personal website. In the e-commerce context, service quality has a significant positive influence on customers' purchase decisions. Because service is created by the interaction between service provider and receiver, service quality evaluations should concentrate on two dimensions: the service delivery process and service outcomes [42]. Patients require multisource information to judge the quality of the service delivery process and outcomes [12]. Per the preceding discussion, physician-generated information reflects the service delivery process and patient-generated information represents the service outcomes, and the two should complement each other. When patients obtain more physician-generated information or patient-generated information on a physician's personal website, the possibility of choosing the physician for consultation (ie, the conversion rate of the physician's personal website) will be enhanced. Therefore, it is hypothesized that: *Physician-generated information and patient-generated information have a complementary relationship in affecting the conversion rate of a physician's personal website (H3).*

System-Generated Information and Conversion Rate

System-generated information is generated by the third-party platform (eg, ranking orders and recommendations) [35,36]. Online product recommendations from third parties are perceived as being more useful in terms of providing accurate information. Many consumers are willing to search for this information to satiate the uncertainty they feel toward information from manufacturers/retailers [34]. The system of an OHC also generates recommendations for physicians. Although this recommendation does not directly reflect a physician's service quality [12], it is more objective and accurate. In an OHC, patients will form an overall impression through systematic recommendations. Compared with physicians with low recommended value, patients are more willing to assume that physicians who have highly recommended value are credible. There is a reason to believe that the conversion

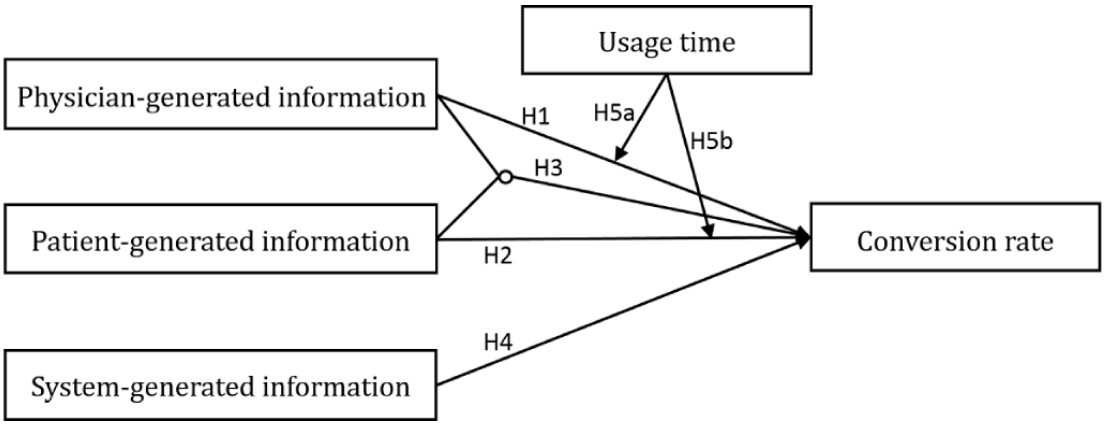
rate of the personal website of a highly recommended physician will be high. Therefore, it is hypothesized that: *System-generated information positively affects the conversion rate of a physician's personal website (H4).*

Moderating Effect of Usage Time

The physician's personal website is a platform for presenting OHC information, including physician-generated and patient-generated information. Studies have shown that the launching of personal websites by physicians can significantly increase the amount of patient reviews [43] (ie, patient-generated information). The longer the website is used, the more physician-generated and patient-generated information should be presented on the personal website. Moreover, the time the physician uses the personal website also indicates the degree of familiarity with online medical services, and patients prefer to choose a more experienced physician. Therefore, it is hypothesized that: *The usage time of a personal website positively moderates the relationship between physician-generated information and the conversion rate (H5a); The usage time of a personal website positively moderates the relationship between patient-generated information and the conversion rate (H5b).*

The overall research model is schematically shown in Figure 2.

Figure 2. Research model.



Methods

For this study, we used data from Haodf.com [21], a leading OHC in China, focusing specifically on physicians treating coronary heart disease as an example. A physician's personal website in Haodf.com presents multisource OHC information, including the number of people who visit that physician's website, the number of patients who consult that physician, physician-generated information (eg, articles), patient-generated information (eg, thank-you letters), and system-generated information (eg, recommended value). Figure 3 shows an example of a physician's personal website.

By means of web crawler technology, the data were collected in five time periods: from March 2019 to July 2019 (for 1 month each over these 5 months), covering 2112 physicians. To investigate whether physician-generated, patient-generated, and system-generated information affect the conversion rate of a physician's personal website, a longitudinal study was designed.

Table 1 summarizes the main variables considered in our study. We used the proportion of the number of patients relative to the number of visits before time *t* as a proxy for the conversion rate of a physician's personal website. The number of patients included those who only consult online, and those who consult again after offline consultation, which can be considered to comprise patients with two kinds of consultations. The mean and maximum number of patients and visits were not an order of magnitude, and their variances were large. Thus, to avoid the calculated conversion rate values being too small, we conducted natural logarithm transformation, ln(*X*+1), as shown in equation (1). Figure 4 shows the frequency statistics of conversion rates over the five time periods, clearly demonstrating variation in conversion rates, which is worth studying.

$$\text{Conversion rate}_{i,t}\ (\%)=[\ln(\text{Patients}_{i,t}+1)/\ln(\text{Visits}_{i,t}+1)]\times 100 \quad \textbf{(1)}$$

The independent variables included physician-generated information, patient-generated information, and system-generated information. Physician-generated information was measured by the number of articles published by the physician before time *t*. Patient-generated information was measured by the number of thank-you letters written by patients after consulting before time *t*. Since the distributions of articles and thank-you letters were nonnormal and some values were 0, natural logarithm transformations were also applied prior to analysis. System-generated information was measured by the recommended value of the system at time *t*, ranging from 1 to 5. The physician's order in the list is also in accordance with the recommended value. Moreover, usage time was used as a moderator, measured by the difference between the launch time of a physician's personal website and time *t*.

Control variables included medical title, academic title, and hospital level by the physician at time *t*. These factors represent objective characteristics of a physician, which are reviewed by the system when the physician enters the OHC. A physician's medical title has four levels: resident physician, attending physician, deputy chief physician, and chief physician, coded from 1 to 4, respectively. The three academic titles for a physician are lecturer, associate professor, and professor, coded from 1 to 3, respectively. Hospitals are also ranked according to three levels coded 1 to 3.

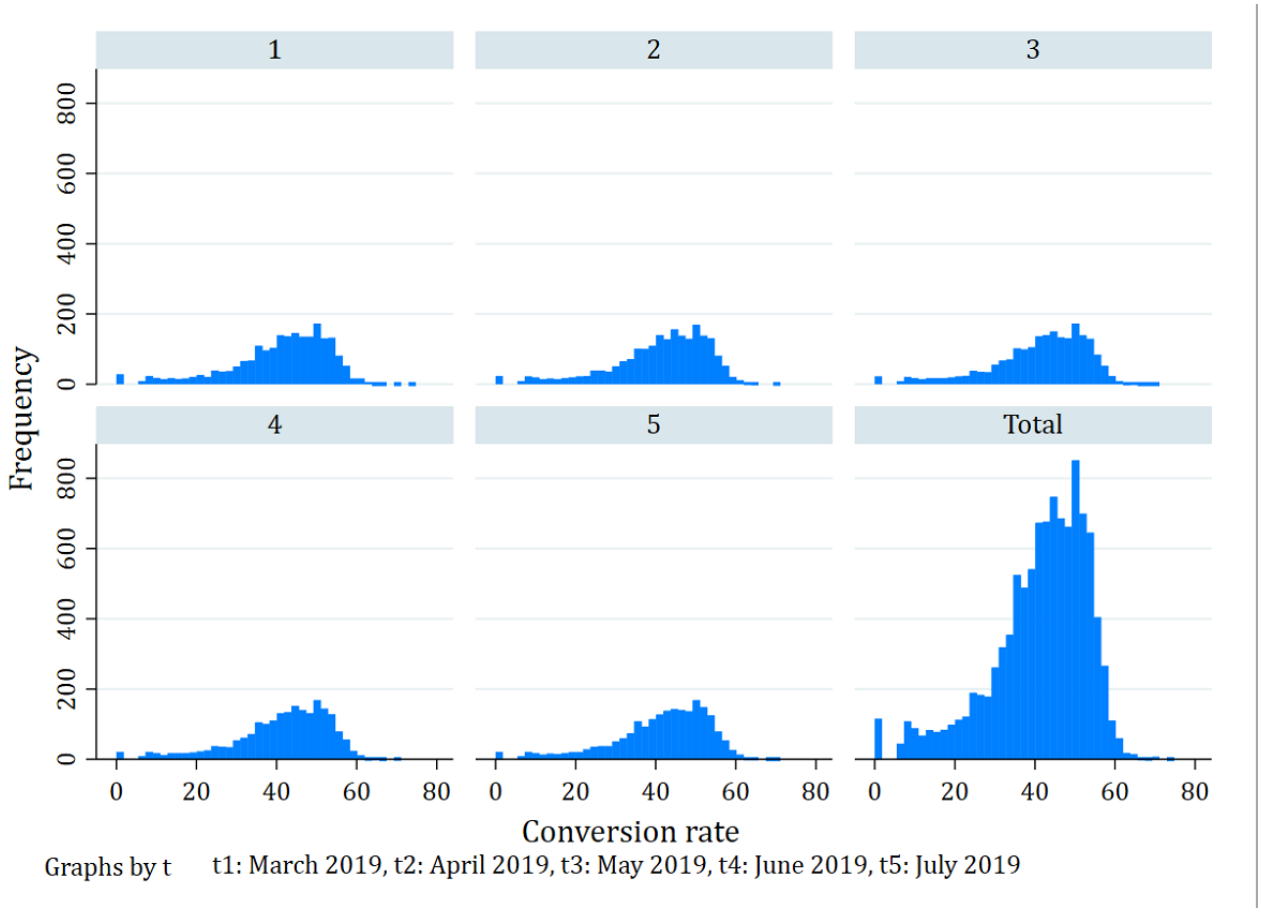
Figure 3. An example of a physician's personal website.

Table 1. Variables description.

Variable	Description	Proxy
Conversion rate (dependent variable)	Proportion of customers to visitors who successfully locate information in the physicians' personal websites and decide to consult.	Patient visits
Independent variables		
Physician-generated information	The information generated by physicians' behaviors and activities in the OHC[a].	Articles
Patient-generated information	The information generated by patients who have experienced medical consultation in the OHC.	Thank-you letters
System-generated information	The information generated by the OHC system.	Recommended value
Usage time (moderator)	The time since the physician has launched their personal website in the OHC.	Usage time
Control variables		
Medical title	The physician's administrative position in the hospital, which is a manifestation of the physician's ranking.	Medical title
Academic title	The physician's professional position, which is also a manifestation of the physician's ranking.	Academic title
Hospital title	The level of the physical hospital where a physician works.	Hospital level

[a]OHC: online health community.

Figure 4. Frequencies of conversion rates in the online health community over five time periods.



Results

Descriptive Statistics and Correlations

Table 2 and Table 3 show the descriptive statistics and correlations among variables, respectively. The numbers of articles and letters showed a positive correlation with the conversion rate, and the number of articles was positively correlated with the number of letters. The recommended value was positively correlated with the conversion rate. Moreover, the number of articles and letters was each positively correlated with usage time.

Table 2. Descriptive statistics (N=10,560).

Variable	Mean (SD)	Minimum	Maximum
Medical title	2.455 (1.463)	0	4
Academic title	0.866 (1.201)	0	3
Hospital level	2.927 (0.378)	0	3
Articles	11.246 (53.232)	0	1315
Thank-you letters	20.738 (66.354)	0	1632
Recommended value	3.816 (0.287)	3.100	5
Usage time	5.058 (3.150)	0	11.420
Patients	529.189 (1390.845)	0	29,613
Visits	439,980 (1302.529)	17	1.81e+07

Table 3. Correlations among variables (N=10,560).

Variable	Medical title	Academic title	Hospital level	Ln(Articles+1)	Ln(Thank-you letters+1)	Recommended value	Usage time	Conversion rate
Medical title								
β	1	.583	.008	.029	.099	.138	.185	–.014
P value	—[a]	<.001	.40	.003	<.001	<.001	<.001	.15
Academic title								
β	.583	1	.015	.137	.254	.307	.392	.065
P value	<.001	—	.14	<.001	<.001	<.001	<.001	<.001
Hospital level								
β	.008	.015	1	–.047	.064	.084	.056	–.045
P value	.40	.14	—	<.001	<.001	<.001	<.001	<.001
ln(Articles+1)								
β	.029	.137	–.047	1	.391	.281	.252	.410
P value	.003	<.001	<.001	—	<.001	<.001	<.001	<.001
ln(Thank-you letters+1)								
β	.099	.254	.064	.391	1	.758	.365	.581
P value	<.001	<.001	<.001	<.001	—	<.001	<.001	<.001
Recommended value								
β	.138	.307	.084	.281	.758	1	.292	.431
P value	<.001	<.001	<.001	<.001	<.001	—	<.001	<.001
Usage time								
β	0.185	.392	.056	.252	.365	.292	1	.009
P value	<.001	<.001	<.001	<.001	<.001	<.001	—	.34
Conversion rate								
β	–.014	.065	–.045	0.410	.581	.431	.009	1
P value	.15	<.001	<.001	<.001	<.001	<.001	.34	—

[a]Not applicable.

Estimation Model

To test the hypotheses, we formulated equation (2). Since the distribution of some variables (*X*) may not be normal and some values were 0, the natural logarithm transformation was used for these variables prior to analysis.

$$\text{Conversion rate}_{i,t}(\%)=[\ln(\text{Patients}_{i,t}+1)/\ln(\text{Visits}_{i,t}+1)]\times 100=\alpha_0+\alpha_1\text{Medical title}_{i,t}+\alpha_2\text{Academic title}_{i,t}+\alpha_3\text{Hospital level}_{i,t}+\alpha_4\text{Physician-generated information}_{i,t}+\alpha_5\text{Patient-generated information}_{i,t}+\alpha_6\text{System-generated information}_{i,t}+\alpha_7\text{Physician-generated information}_{i,t}*\text{Patient-generated information}_{i,t}+\alpha_8\text{Usage time}_{i,t}+\alpha_9\text{Physician-generated information}_{i,t}*\text{Usage time}_{i,t}+\alpha_{10}\text{Patient-generated information}_{i,t}*\text{Usage time}_{i,t}+\alpha_{i,t} \quad \textbf{(2)}$$

Regression Results

The results of the Hausman test ($\chi^2(11)=238.49$, $P<.001$) showed that a fixed-effects model was more suitable for this study. The fixed-effects model assists in controlling for unobserved heterogeneity (ie, heterogeneity among individual physicians) when this heterogeneity is constant over time. Therefore, the results of the fixed-effects model are reported accordingly in the main analysis, as shown in Table 4.

In addition, the results are presented hierarchically. Model 1 contained only control variables, whereas model 2 to model 4 included the independent variables and interaction terms. Model 5 represents the full model that included all of the independent variables as well as the interaction terms.

From model 2, the coefficient for physician-generated information was positive and statistically significant, indicating that physician-generated information positively affects the conversion rate of a physician's personal website. Therefore, H1 is supported. The results of model 2 also showed that the coefficient of patient-generated information was positive and

statistically significant, suggesting that patient-generated information positively affects the conversion rate of a physician's personal website. Therefore, H2 is supported.

Model 3 showed that the interaction between physician-generated information and patient-generated information was negative and significant. This means that the effect of physician-generated and patient-generated information on the conversion rate is a substitute relationship rather than a complementary relationship, which is in contrast to H3.

Model 2 showed that the coefficient of system-generated information was positive and statistically significant, suggesting that system-generated information positively affects the conversion rate of a physician's personal website. Therefore, H4 is supported.

Model 4 showed that the interaction between physician-generated information and usage time was negative and significant. This finding suggests that the relationship between physician-generated information and the conversion rate can be negatively moderated by the usage time of a personal website. That is, when the usage time of a personal website is lower, the relationship between physician-generated information and the conversion rate is stronger, as shown in Figure 5. Therefore, H5a is not supported.

By contrast, model 4 also showed that the interaction between patient-generated information and usage time was positive and significant. This finding suggests that the effect of patient-generated information on the conversion rate is stronger for physicians with high usage time of their personal websites, as shown in Figure 6. Therefore, H5b is supported.

Table 4. Regression results (fixed-effects models).

Variable	Model 1[a]		Model 2[b]		Model 3[c]		Model 4[d]		Model 5[e]	
	β[a] (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value
Constant	41.206 (1.300)	<.001	21.759 (1.867)	<.001	33.068 (1.251)	<.001	35.119 (1.427)	<.001	21.156 (2.052)	<.001
Medical title	–.048 (.014)	.001	.006 (.014)	.68	.015 (.014)	.27	.010 (.017)	.52	.007 (.016)	.64
Academic title	.012 (.028)	.66	–.046 (.026)	.08	–.050 (.026)	.06	–.032 (.026)	.23	–.026 (.026)	.33
Hospital ranking	.023 (.444)	.96	–.046 (.420)	.91	.025 (.417)	.95	.064 (.419)	.88	.010 (.413)	.98
PHI[f]	N/A[g]	N/A	3.009 (.127)	<.001	4.436 (.165)	<.001	3.993 (.167)	<.001	5.070 (.185)	<.001
PI[h]	N/A	N/A	1.694 (.131)	<.001	2.843 (.145)	<.001	.757 (.196)	<.001	1.315 (.202)	<.001
SI[i]	N/A	N/A	3.336 (.381)	<.001	N/A	N/A	N/A	N/A	3.465 (.382)	<.001
PHI×PI	N/A	N/A	N/A	N/A	–.904 (.067)	<.001	N/A	N/A	–.970 (.071)	<.001
Time[j]	N/A	N/A	N/A	N/A	N/A	N/A	–.346 (.160)	.03	–.429 (.160)	.008
PHI×Time	N/A	N/A	N/A	N/A	N/A	N/A	–.333 (.040)	<.001	–.147 (.041)	<.001
PI×Time	N/A	N/A	N/A	N/A	N/A	N/A	.325 (.041)	<.001	.382 (.041)	<.001

[a]R^2 within=0.003; R^2 (between)=0.002; R^2 (overall)=0.001.

[b]R^2 (within)=0.110; R^2 (between)=0.331; R^2 (overall)=0.329.

[c]R^2 (within)=0.121; R^2 (between)=0.334; R^2 (overall)=0.332.

[d]R^2 (within)=0.113; R^2 (between)=0.355; R^2 (overall)=0.353.

[e]R^2 (within)=0.139; R^2 (between)=0.344; R^2 (overall)=0.343.

[f]PHI: Physician-generated information.

[g]N/A: not applicable.

[h]PI: Patient-generated information.

[i]SI: System-generated information.

[j]Time: Usage time.

Figure 5. The moderating effect of usage time on physician-generated information and the conversion rate.

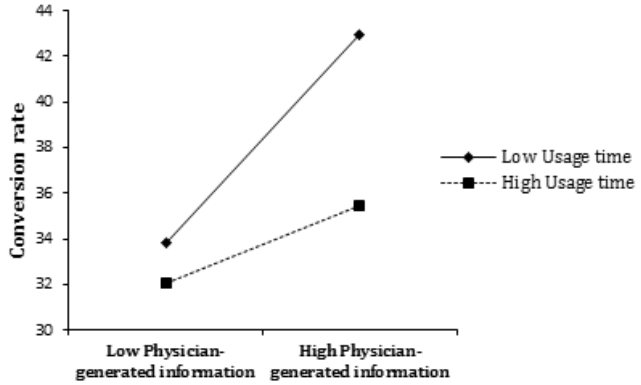
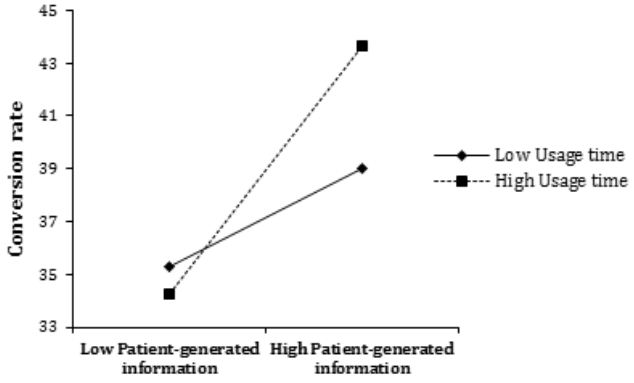


Figure 6. The moderating effect of usage time on patient-generated information and the conversion rate.



Robustness Check

To check the robustness of the results, we considered the time effect and used the two-way fixed-effects model to rerun the estimation model (Equation 2). In this case, time was defined as a dummy variable and *t*1 (March 2019) was used as the baseline period. The new estimation model is shown in equation (3), and the results are presented in Table 5, which were consistent with the results of the previous model (Table 4). In addition, the joint significance of the time dummy variable was tested, which confirmed that the time effect should be included in the new estimation model. The robustness check results suggested that H1, H2, H4 and H5b were supported.

$$\text{Conversion rate}_{i,t}(\%)=[\ln(\text{Patients}_{i,t}+1)/\ln(\text{Visits}_{i,t}+1)]\times 100 = \beta_0 + \beta_1\text{Medical title}_{i,t} + \beta_2\text{Academic title}_{i,t} + \beta_3\text{Hospital level}_{i,t} + \beta_4\text{Physician-generated information}_{i,t} + \beta_5\text{Patient-generated information}_{i,t} + \beta_6\text{System-generated information}_{i,t} + \beta_7\text{Physician-generated information}_{i,t}*\text{Patient-generated information}_{i,t} + \beta_8\text{Usage time}_{i,t} + \beta_9\text{Physician-generated information}_{i,t}*\text{Usage time}_{i,t} + \beta_{10}\text{Patient-generated information}_{i,t}*\text{Usage time}_{i,t} + \beta_{11}t2 + \beta_{12}t3 + \beta_{13}t4 + \beta_{14}t5 + \beta_{i,t} \quad \textbf{(3)}$$

Table 5. Robustness check (fixed-effects models).

Variable	Model 1[a]		Model 2[b]		Model 3[c]		Model 4[d]		Model 5[e]	
	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value
Constant	40.801 (1.301)	<.001	21.470 (1.883)	<.001	33.180 (1.255)	<.001	158.795 (9.053)	<.001	141.657 (9.053)	<.001
Medical title	.006 (.014)	.89	–.013 (.014)	.73	–.021 (.038)	.58	–.044 (.038)	.25	–.038 (.038)	.31
Academic title	–.001 (.029)	.98	–.041 (.028)	.14	–.040 (.027)	.14	–.009 (.027)	.75	–.005 (.027)	.85
Hospital rank	.060 (.443)	.89	–.038 (.420)	.93	.029 (.417)	.95	.074 (.414)	.86	.020 (.409)	.96
PHI[f]	N/A[g]	N/A	2.986 (.128)	<.001	4.443 (.166)	<.001	4.034 (.165)	<.001	5.081 (.183)	<.001
PI[h]	N/A	N/A	1.637 (.135)	<.001	2.823 (.148)	<.001	.628 (.194)	.001	1.168 (.201)	<.001
SI[i]	N/A	N/A	3.446 (.387)	<.001	N/A	N/A	N/A	N/A	3.446 (.378)	<.001
PHI*PI	N/A	N/A	N/A	N/A	–.905 (.067)	<.001	N/A	N/A	–.945 (.070)	<.001
Time[j]	N/A	N/A	N/A	N/A	N/A	N/A	–25.554 (1.831)	<.001	–24.984 (1.805)	<.001
PHI×Time	N/A	N/A	N/A	N/A	N/A	N/A	–.345 (.039)	<.001	–.164 (.041)	<.001
PI×Time	N/A	N/A	N/A	N/A	N/A	N/A	.335 (.041)	<.001	.391 (.041)	<.001
t2[k]	.121 (.039)	.002	.035 (.037)	.34	.015 (.037)	.67	1.944 (.144)	<.001	1.894 (.142)	<.001
t3[l]	.197 (.039)	<.001	.057 (.037)	.12	.031 (.037)	.40	4.108 (.298)	<.001	3.997 (.293)	<.001
t4[m]	.254 (.039)	<.001	.072 (.038)	.06	.031 (.037)	.40	6.386 (.462)	<.001	6.222 (.456)	<.001
t5[n]	.307 (.117)	.009	–.011 (.112)	.92	–.089 (.111)	.42	8.299 (.612)	<.001	8.108 (.604)	<.001

[a]R^2 within=0.009; R^2 (between)=0.003; R^2 (overall)=0.000.
[b]R^2 (within)=0.110; R^2 (between)=0.330; R^2 (overall)=0.328.
[c]R^2 (within)=0.121; R^2 (between)=0.334; R^2 (overall)=0.332.
[d]R^2 (within)=0.132; R^2 (between)=0.001; R^2 (overall)=0.001.
[e]R^2 (within)=0.157; R^2 (between)=0.001; R^2 (overall)=0.001.
[f] PHI: Physician-generated information.
[g]N/A: not applicable.
[h]PI: Patient-generated information.
[i]SI: System-generated information.
[j]Time: Usage time.
[k]*t2*: April 2019.
[l]*t3*: May 2019.
[m]*t4*: June 2019.
[n]*t5*: July 2019.

Discussion

Principal Findings

The aim of this study was to explore the conversion rate of OHCs. Considering the importance of the physician's personal website in the patient decision-making process, the conversion rate in this study was considered to be the proportion of customers among visitors of a physician's personal website. The results showed that physician-generated information had a positive effect on the conversion rate. Corresponding to previous research, the number of articles published by a physician reflects their activeness and affects patients' choices [18], thereby affecting social and economic returns [20].

Just as reviews and posts generated by customers or members affect the conversion rate of websites [7,28,31], patient-generated information also positively affects the conversion rate of physicians' personal websites. Physician-generated and patient-generated information serve as two clues to reflect two dimensions of a physician's service quality: the service delivery process and service outcomes. The results showed that there was no complementary relationship but rather a substitute relationship between physician-generated and patient-generated information. A possible explanation for this finding is that both physician-generated and patient-generated information contain information about a physician's quality, and these two types of online quality information may have some overlap. More patient-generated information also means that there are many patients who have chosen to consult this physician. In general, patients believe that a good physician with more patients should be very busy and should not have sufficient time to contribute much online. As a result, patients may tend to believe that the signals are manipulated by the physician or the system [5].

This study also found that system-generated information positively affected the conversion rate of a physician's personal website. This result is consistent with a prior study suggesting that ranking orders and recommendations generated by the system affect the conversion rates of online hotel booking websites [7].

Although physician-generated and patient-generated information on a physician's personal website are cumulative since the launch of the website, the moderating effects of usage time on the two types of information differed. That is, the usage time of a personal website had a positive moderating effect on the relationship between patient-generated information and the conversion rate, but had a negative moderating effect on the relationship between physician-generated information and the conversion rate. This may reflect that patient-generated information grows in proportion to the time a physician uses the personal website, whereas physician-generated information grows more slowly. It may also be related to the fact that compared with physician-generated information, patient-generated information is more relevant to the physician's familiarity with online medical services.

Theoretical Implications

This study offers theoretical contributions in the following ways. First, conversion rates of other types of websites (ie, retail or e-commerce websites, potential customer generation websites, and content websites) and their influencing factors have been extensively studied. This study focused on the conversion rate of a service website (OHC). The conversion rate in this study was considered as the proportion of customers among patients who visited a physician's personal website. Paying attention to the conversion rate of physicians' personal websites and analyzing its influencing factors can provide clearer understanding of the patient decision-making process in OHCs.

Second, previous studies divided OHC information into patient-generated information and system-generated information according to the source. By contrast, this study also added the component of physician-generated information and thus extended research on multisource OHC information. The results showed that multisource OHC information (physician-generated, patient-generated, and system-generated information) positively affected the conversion rate of a physician's personal website.

Third, this study employed usage time as OHC information related to the familiarity of physicians with online medical services, which was analyzed together with physician-generated and patient-generated information. The results demonstrated that the moderating effects of usage time on the two kinds of OHC information were different. From this perspective, this study extends the understanding of OHC information from the time dimension.

Implications for Practice

This study provides several relevant practical implications. First, for the manager, analyzing and improving the conversion rate of a physician's personal website is conducive to improving the efficiency of the entire platform. Our results further indicate that system-generated information (eg, recommended value) positively affects the conversion rate of a physician's personal website. The website manager should improve the accuracy of system-generated information and update it in time.

Second, the results showed that physician-generated information affected the conversion rate of a physician's personal website. Therefore, physicians who want to improve the conversion rate of their websites must be active and hardworking, and pay attention to the information left by their own behaviors and activities.

Third, although both physician-generated and patient-generated information positively affected the conversion rate, they showed a substitute relationship rather than a complementary relationship. In other words, these two types of OHC information have distinct roles in affecting the conversion rate of a physician's personal website. Physicians can mainly focus on either physician-generated information or patient-generated information, taking into account their specific objectives or resource constraints in different situations. Additionally, physicians should distinguish between the service delivery process and service outcomes in patients' assessments of service quality.

Fourth, the results showed that physician-generated and patient-generated information had a time effect on the conversion rate, and the two were different. Therefore, physicians should not only pay attention to the cumulative amount of physician-generated and patient-generated information in personal websites but also the directly proportional effects to their usage time.

Limitations and Future Research

This study has certain limitations. First, the interpretation of the findings is limited by using data from only one Chinese OHC, Haodf.com, and one type of physician, specialists in coronary heart disease. Therefore, collecting data from physicians with various types of expertise on different platforms simultaneously is necessary to further verify the research model. Second, it is difficult to distinguish between two types of consultations: patients who browse and consult online, and

patients who browse online but consult offline. In this study, we used the number of patients and visits collected by the platform, which might have resulted in the loss of some information. Third, it is difficult to obtain patient-level data from this platform, since the platform tends to mask the users' names to protect patient privacy. Therefore, we were not able to empirically examine how different patients make choices based on available OHC information. Consequently, this exclusion may be a potential limitation of this study.

Conclusions

In this study, the conversion rate of OHCs was analyzed (ie, how to convert visitors into customers). We hypothesized that multisource OHC information (physician-generated, patient-generated, and system-generated information) would affect the conversion rate, and that usage time would moderate the relationships between physician-generated information, patient-generated information, and the conversion rate. Short-term panel data over five time periods (months) were used to test these hypotheses. The results indicate that physician-generated, patient-generated, and system-generated information positively affect the conversion rate. In addition, physician-generated and patient-generated information have a substitute relationship rather than a complementary relationship in affecting the conversion rate. Moreover, the usage time of a personal website positively moderates patient-generated information, but negatively moderates physician-generated information.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (71531013, 71729001) and by the Ministry of Education Youth Program in Humanities and Social Sciences (20YJC630214).

Conflicts of Interest

None declared.

References

1. Patel V, Vaidya R, Naik D, Borker P. Irrational drug use in India: a prescription survey from Goa. J Postgrad Med 2005;51(1):9-12 [FREE Full text] [Medline: 15793331]
2. Luo P, Chen K, Wu C, Li Y. Exploring the social influence of multichannel access in an online health community. J Assoc Inf Sci Technol 2017 Sep 19;69(1):98-109. [doi: 10.1002/asi.23928]
3. Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the internet affects patients' experience of cancer: a qualitative study. BMJ 2004 Mar 06;328(7439):564 [FREE Full text] [doi: 10.1136/bmj.328.7439.564] [Medline: 15001506]
4. Guo X, Guo S, Vogel D, Li Y. Online Healthcare Community Interaction Dynamics. J Manage Sci Eng 2016 Dec;1(1):58-74. [doi: 10.3724/SP.J.1383.101004]
5. Li J, Tang J, Jiang L, Yen DC, Liu X. Economic Success of Physicians in the Online Consultation Market: A Signaling Theory Perspective. Int J Electron Comm 2019 Mar 26;23(2):244-271. [doi: 10.1080/10864415.2018.1564552]
6. Gudigantala N, Bicen P, Eom M. An examination of antecedents of conversion rates of e-commerce retailers. Manage Res Rev 2016 Jan 18;39(1):82-114. [doi: 10.1108/mrr-05-2014-0112]
7. Cezar A, Ögüt H. Analyzing conversion rates in online hotel booking. Int J Contemp Hospitality Mngt 2016 Feb 08;28(2):286-304. [doi: 10.1108/ijchm-05-2014-0249]
8. Perdikaki O, Kesavan S, Swaminathan JM. Effect of Traffic on Sales and Conversion Rates of Retail Stores. Manuf Serv Oper Manag 2012 Jan;14(1):145-162. [doi: 10.1287/msom.1110.0356]
9. Parasuraman A, Zeithaml VA, Berry LL. A Conceptual Model of Service Quality and Its Implications for Future Research. J Marketing 2018 Nov 19;49(4):41-50. [doi: 10.1177/002224298504900403]
10. Jackson S. About website conversion rate. Website Gear. 2004. URL: http://content.websitegear.com/article/conversion_rate.htm [accessed 2020-05-24]
11. Li Y, Song Y, Zhao W, Guo X, Ju X, Vogel D. Exploring the Role of Online Health Community Information in Patients' Decisions to Switch from Online to Offline Medical Services. Int J Med Inform 2019 Oct;130:103951. [doi: 10.1016/j.ijmedinf.2019.08.011] [Medline: 31473534]
12. Yang H, Guo X, Wu T, Ju X. Exploring the effects of patient-generated and system-generated information on patients' online search, evaluation and decision. Electron Comm Res Appl 2015 May;14(3):192-203. [doi: 10.1016/j.elerap.2015.04.001]
13. Beritelli P, Bieger T, Laesser C. The Impact of the Internet on Information Sources Portfolios. J Travel Tourism Market 2007 Sep 19;22(1):63-80. [doi: 10.1300/j073v22n01_05]
14. Jang S, Prasad A, Ratchford BT. Consumer Search of Multiple Information Sources and its Impact on Consumer Price Satisfaction. J Interact Market 2017 Nov;40:24-40. [doi: 10.1016/j.intmar.2017.06.004]
15. Kim JS, Ratchford BT. Consumer Choice and Use of Multiple Information Sources for Automobile Purchases. International Journal of Electronic Commerce 2014 Dec 08;16(3):7-40. [doi: 10.2753/jec1086-4415160301]

16. Liu X, Guo X, Wu H, Wu T. The Impact of Individual and Organizational Reputation on Physicians' Appointments Online. Int J Electron Comm 2016 Jun 20;20(4):551-577. [doi: 10.1080/10864415.2016.1171977]
17. Li S, Lee-Won RJ, McKnight J. Effects of Online Physician Reviews and Physician Gender on Perceptions of Physician Skills and Primary Care Physician (PCP) Selection. Health Commun 2019 Oct;34(11):1250-1258. [doi: 10.1080/10410236.2018.1475192] [Medline: 29792519]
18. Li Y, Ma X, Song J, Yang Y, Ju X. Exploring the Effects of Online Rating and the Activeness of Physicians on the Number of Patients in an Online Health Community. Telemed J E Health 2019 Nov;25(11):1090-1098. [doi: 10.1089/tmj.2018.0192] [Medline: 30676279]
19. Lu N, Wu H. Exploring the impact of word-of-mouth about Physicians' service quality on patient choice based on online health communities. BMC Med Inform Decis Mak 2016 Nov 26;16(1):151 [FREE Full text] [doi: 10.1186/s12911-016-0386-0] [Medline: 27888834]
20. Guo S, Guo X, Fang Y, Vogel D. How Doctors Gain Social and Economic Returns in Online Health-Care Communities: A Professional Capital Perspective. J Mgnt Inform Syst 2017 Aug 17;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
21. Haodf.com. URL: http://www.haodf.com/ [accessed 2020-08-10]
22. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009 Feb;6(2):492-525 [FREE Full text] [doi: 10.3390/ijerph6020492] [Medline: 19440396]
23. Cao X, Liu Y, Zhu Z, Hu J, Chen X. Online selection of a physician by patients: Empirical study from elaboration likelihood perspective. Comput Human Behav 2017 Aug;73:403-412. [doi: 10.1016/j.chb.2017.03.060]
24. Liu F, Li Y, Ju X. Exploring Patients' Consultation Behaviors in the Online Health Community: The Role of Disease Risk. Telemed J E Health 2019 Mar;25(3):213-220. [doi: 10.1089/tmj.2018.0033] [Medline: 29927721]
25. Ayanso A, Yoogalingam R. Profiling Retail Web Site Functionalities and Conversion Rates: A Cluster Analysis. Int J Electron Comm 2014 Dec 08;14(1):79-114. [doi: 10.2753/jec1086-4415140103]
26. McDowell WC, Wilson RC, Kile CO. An examination of retail website design and conversion rate. J Bus Res 2016 Nov;69(11):4837-4842. [doi: 10.1016/j.jbusres.2016.04.040]
27. Soonsawad P. Developing a New Model for Conversion Rate Optimization: A Case Study. Int J Bus Manag 2013 Apr 21;8(10):41. [doi: 10.5539/ijbm.v8n10p41]
28. Ludwig S, de Ruyter K, Friedman M, Brüggen EC, Wetzels M, Pfann G. More than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. J Marketing 2013 Jan;77(1):87-103. [doi: 10.1509/jm.11.0560]
29. Moe WW, Fader PS. Dynamic Conversion Behavior at E-Commerce Sites. Manag Sci 2004 Mar;50(3):326-335. [doi: 10.1287/mnsc.1040.0153]
30. Sismeiro C, Bucklin RE. Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach. J Market Res 2018 Oct 10;41(3):306-323. [doi: 10.1509/jmkr.41.3.306.35985]
31. Zhou Z, Wu JP, Zhang Q, Xu S. Transforming visitors into members in online brand communities: Evidence from China. J Bus Res 2013 Dec;66(12):2438-2443. [doi: 10.1016/j.jbusres.2013.05.032]
32. Yang H, Ju X. Investigating the influences of motivators on physician contribution behaviors in online health community: Offline status as a moderator. 2016 Presented at: Pacific Asia Conference on Information Systems (PACIS); 2016; Taiwan p. 11.
33. Westerman D, Spence PR, Van Der Heide B. A social network as information: The effect of system generated reports of connectedness on credibility on Twitter. Comput Human Behav 2012 Jan;28(1):199-206. [doi: 10.1016/j.chb.2011.09.001]
34. Chen J, Teng L, Yu Y, Yu X. The effect of online information sources on purchase intentions between consumers with high and low susceptibility to informational influence. J Bus Res 2016 Feb;69(2):467-475. [doi: 10.1016/j.jbusres.2015.05.003]
35. Goolsbee A, Chevalier JA. Measuring Prices and Price Competition Online: Amazon and Barnes and Noble. SSRN J 2002;1(2):203-222. [doi: 10.2139/ssrn.319701]
36. Senecal S, Nantel J. The influence of online product recommendations on consumers' online choices. J Retail 2004 Jan;80(2):159-169. [doi: 10.1016/j.jretai.2004.04.001]
37. Wu H, Lu N. Service provision, pricing, and patient satisfaction in online health communities. Int J Med Inform 2018 Feb;110:77-89. [doi: 10.1016/j.ijmedinf.2017.11.009] [Medline: 29331257]
38. Haines GH, Howard JA, Sheth JN. The Theory of Buyer Behavior. J Am Stat Assoc 1970 Sep;65(331):1406. [doi: 10.2307/2284311]
39. Wu H, Deng Z. Do physicians' online activities impact outpatient visits? An examination of online health communities. 2019 Presented at: Pacific Asia Conference on Information Systems (PACIS) 2019; 2019; Xi'an, China.
40. Gupta AK, Govindarajan V, Malhotra A. Feedback-seeking behavior within multinational corporations. Strat Mgmt J 1999 Mar;20(3):205-222. [doi: 10.1002/(sici)1097-0266(199903)20:3<205::aid-smj17>3.0.co;2-h]
41. Gao G, Greenwood BN, McCullough J, Agarwal R. A Digital Soapbox? The Information Value of Online Physician Ratings. 2012 Jul Presented at: Academy of Management Proceedings; 2012; Briarcliff Manor, NY p. 16354. [doi: 10.5465/AMBPP.2012.16354abstract]

42. Grönroos C. A Service Quality Model and its Marketing Implications. Eur J Market 1984 Apr;18(4):36-44. [doi: 10.1108/eum0000000004784]
43. Wang H, Liu C, Gao B. A study on the influence of doctor personal website opening on patients' review volume. Chinese J Manag 2018;15(6):901-907. [doi: 10.3969/j.issn.1672?884x.2018.06.013]

Abbreviations

e-commerce: electronic commerce
OHC: online health community

Edited by G Eysenbach; submitted 23.05.20; peer-reviewed by T Muto, X Song; comments to author 12.06.20; revised version received 23.06.20; accepted 25.06.20; published 26.08.20.

Please cite as:
Chen Q, Yan X, Zhang T
Converting Visitors of Physicians' Personal Websites to Customers in Online Health Communities: Longitudinal Study
J Med Internet Res 2020;22(8):e20623
URL: http://www.jmir.org/2020/8/e20623/
doi:10.2196/20623
PMID:32845248

Original Paper

Machine Learning Model for Risk Prediction of Community-Acquired Acute Kidney Injury Hospitalization From Electronic Health Records: Development and Validation Study

Chien-Ning Hsu[1,2], PhD; Chien-Liang Liu[3], PhD; You-Lin Tain[4], MD, PhD; Chin-Yu Kuo[3], MSc; Yun-Chun Lin[3], MSc

[1]Department of Pharmacy, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung, Taiwan

[2]School of Pharmacy, Kaohsiung Medical University, Kaohsiung, Taiwan

[3]Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu, Taiwan

[4]Division of Pediatric Nephrology, Kaohsiung Chang Gung Memorial Hospital and Chang Gung Medical University, Kaohsiung, Taiwan

Corresponding Author:
Chien-Liang Liu, PhD
Department of Industrial Engineering and Management
National Chiao Tung University
1001 University Road
Hsinchu, 300
Taiwan
Phone: 886 +886975368975
Email: jacky168@gmail.com

Abstract

Background: Community-acquired acute kidney injury (CA-AKI)-associated hospitalizations impose significant health care needs and contribute to in-hospital mortality. However, most risk prediction models developed to date have focused on AKI in a specific group of patients during hospitalization, and there is limited knowledge on the baseline risk in the general population for preventing CA-AKI-associated hospitalization.

Objective: To gain further insight into risk exploration, the aim of this study was to develop, validate, and establish a scoring system to facilitate health professionals in enabling early recognition and intervention of CA-AKI to prevent permanent kidney damage using different machine-learning techniques.

Methods: A nested case-control study design was employed using electronic health records derived from a group of Chang Gung Memorial Hospitals in Taiwan from 2010 to 2017 to identify 234,867 adults with at least two measures of serum creatinine at hospital admission. Patients were classified into a derivation cohort (2010-2016) and a temporal validation cohort (2017). Patients with the first episode of CA-AKI at hospital admission were classified into the case group and those without CA-AKI were classified in the control group. A total of 47 potential candidate variables, including age, gender, prior use of nephrotoxic medications, Charlson comorbid conditions, commonly measured laboratory results, and recent use of health services, were tested to develop a CA-AKI hospitalization risk model. Permutation-based selection with both the extreme gradient boost (XGBoost) and least absolute shrinkage and selection operator (LASSO) algorithms was performed to determine the top 10 important features for scoring function development.

Results: The discriminative ability of the risk model was assessed by the area under the receiver operating characteristic curve (AUC), and the predictive CA-AKI risk model derived by the logistic regression algorithm achieved an AUC of 0.767 (95% CI 0.764-0.770) on derivation and 0.761 on validation for any stage of AKI, with positive and negative predictive values of 19.2% and 96.1%, respectively. The risk model for prediction of CA-AKI stages 2 and 3 had an AUC value of 0.818 for the validation cohort with positive and negative predictive values of 13.3% and 98.4%, respectively. These metrics were evaluated at a cut-off value of 7.993, which was determined as the threshold to discriminate the risk of AKI.

Conclusions: A machine learning–generated risk score model can identify patients at risk of developing CA-AKI-related hospitalization through a routine care data-driven approach. The validated multivariate risk assessment tool could help clinicians to stratify patients in primary care, and to provide monitoring and early intervention for preventing AKI while improving the quality of AKI care in the general population.

(J Med Internet Res 2020;22(8):e16903) doi:10.2196/16903

KEYWORDS

community-acquired acute kidney injury (CA-AKI); hospitalization; treatment decision making; clinical decision support system; machine learning; feature selection with extreme gradient boost (XGBoost); least absolute shrinkage and selection operator (LASSO); risk prediction

Introduction

Acute kidney injury (AKI) is defined as an acute increase in serum creatinine (SCr) or reduction in urine volume [1]. Most AKI cases (67%-80%) develop in the community (ie, community-acquired AKI [CA-AKI]), and despite substantial hospitalization care [2-4], CA-AKI is associated with an increased risk of in-hospital mortality compared with that of hospitalized patients without AKI (65%-90%) [2-5]. The financial burden associated with AKI, including the need for dialysis and intensive unit care during hospitalization and the lack of kidney recovery following discharge care, poses significant strain on the health care system [6,7].

The 22nd Acute Disease Quality Initiative Consensus Conference suggested that high-quality care for patients with AKI or those at risk of AKI should start at the community level and continue in the emergency department, hospital setting, and after discharge from inpatient care [8]. AKI is often reversible. The diagnosis of early-stage AKI is difficult as it depends on SCr measurements and urine outputs that are difficult to routinely monitor in outpatient practice. Existing evidence has highlighted the need for clinical tools to provide early and accurate predictions for diagnosing CA-AKI and to deliver preventive management that may prevent irreversible nephron loss in the general population. However, most of the AKI prediction models developed to date focus on a specific group in the hospital setting, such as the risk of hospital-acquired AKI developed following operations [9,10], cardiac procedures [11,12], liver transplantation [13], or intensive care unit admission [14-16]; the few studies assessing risk for the general population are prone to external validity bias.

With significant advances in the application of machine-learning techniques, the extreme gradient boost (XGBoost) [13,14], least absolute shrinkage and selection operator (LASSO) [11,12,15], and random forest [13,16] models have been employed for predicting AKI risk in different clinical scenarios and have shown promising advantages based on the aggregation of data from electronic health records (EHRs). However, current methodological approaches for AKI risk prediction for highly selective groups of patients have limited impact in terms of the rapid integration of such applications into real-world clinical decision support systems. In addition, some AKI prediction models focus on biomarkers that are not widely available for assessment in practice [15]. AKI could be associated with a variety of causes such as nephrotoxins, existing disease status, and volume status. Therefore, a diagnostic tool with routinely measured characteristics and laboratory tests can easily identify patients with a high probability of developing CA-AKI and inform physicians on the possibility of its occurrence. Subsequent attention and action to hemodynamic monitoring and avoidance of nephrotoxins may ultimately enhance care and improve patient outcomes.

The primary aim of this study was to develop and validate a risk prediction model of CA-AKI hospitalization that can be used to identify patients at higher risk of developing CA-AKI and requiring hospital care. Because the XGBoost and LASSO algorithms have been widely and successfully used for predicting the risk of AKI development from EHR data [11-15], both algorithms were employed in this study to further explore machine-learning models for CA-AKI hospitalization risk prediction and gain insights into improving such prediction models. The secondary aim of the study was to transform the prediction model into a scoring function to quantify the risk of CA-AKI hospitalization. This scoring function can facilitate risk assessment and guide treatment decisions for modifiable risk management and prevention in an outpatient setting.

Methods

Study Cohort

A nested case-control study was performed on hospitalized patients aged 20 years or older admitted to the emergency department or outpatient clinic and requiring hospitalization between 2010 and 2017 (Multimedia Appendix 1) from a group of Chang Gung Memorial Hospitals (CGMHs) located in different cities from the north to south of Taiwan. The EHR data from CGMHs included 6.1% outpatient and 10.2% inpatient encounters of the Taiwan population in 2015 [17].

Adults hospitalized from 2010 to 2016 served as the training dataset and data from patients hospitalized during 2017 were used in the internal validation model. To estimate the robust probability of CA-AKI associated with hospitalization, patients' reference (≤3 months) and index (at the admission date) values of SCr were required to determine an acute episode of kidney injury. This study was approved by the Institutional Review and Ethics Board of CGMH, Taoyuan in Taiwan (permit number: 201801461B0). All datasets used in this study were deidentified prior to being transferred to the study investigators. The study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement for reporting multivariable prediction model development and validation [18].

Predictors

Based on a literature review of factors shown to increase the risk of AKI and expert opinions [19,20], we identified 55 candidate predictor variables (Multimedia Appendix 2), including comorbid conditions, outpatient nephrotoxic medicines, recent emergency department visits, outpatient or hospital admissions, and potential laboratory results during the baseline period within 90 days prior to the index hospital

admission (medicines, visits, and laboratory results) and within 365 days preceding the index admission (comorbid conditions).

These diagnosis codes were then classified into 17 binary comorbidity groupings as defined by the Charlson comorbidity index (CCI) [21]. Patients' concomitant outpatient medicine records were classified into 16 therapeutic binary indicator groupings and counted as the total number of active classes of medicines. Laboratory results included 10 test items in continuous data. These prespecified variables in the initial set of candidate variables for training and validation model prediction algorithms were not limited to strong statistical assumptions of causality or correlation and presumably provided an opportunity to discover new knowledge from machine-learning methods. The full list of candidate variables with corresponding variable names in the data form can be found in Multimedia Appendix 2.

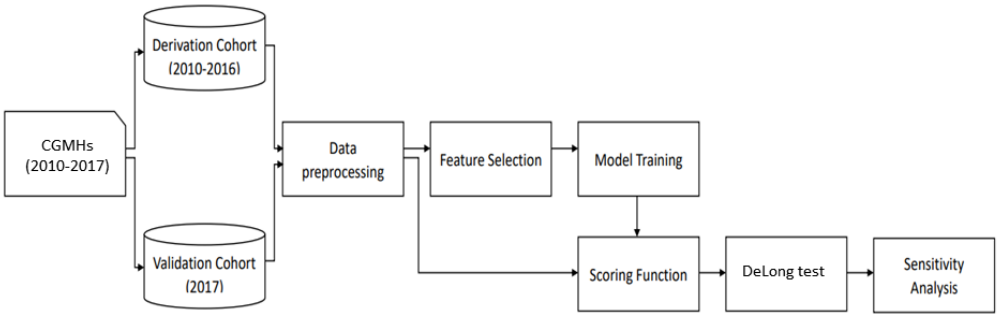
Outcomes

The outcome was CA-AKI-associated hospitalization defined according to the Kidney Disease: Improving Global Outcomes (KDIGO) criteria as an increase in SCr within 48 hours of 0.3 mg/dl from the reference value, ≥1.5 of the reference value (or increase to 4 mg/dl) within 7 days or up to 90 days prior to the patient's first admission (index date) in the study period [1]. The reference SCr value was first retrieved based on the availability of measured SCr within 2 days prior to the index date, within 7 days for patients without a recent 2-day SCr, and within 30 days or up to 90 days for patients without any SCr measurements for 7 days. Two approaches were used to determine reference SCr. First, for patients who had multiple SCr measures in the 2- and 7-day time windows, the latest SCr (ie, closest to the index date) was chosen as the reference SCr. Second, the mean SCr value within 8-90 days before the index date was defined as the reference SCr for patients who had multiple SCr measurements in the period [22]. Only patients who had reference and index SCr measures were included for analysis in the study (Multimedia Appendix 1). Patients who fulfilled the KDIGO AKI criteria were classified into the case group and patients without AKI at admission were included in the control group. Stage 1 AKI was defined as an SCr increase of ≥0.3 mg/dl from the reference value within 2 days or an increase of 1.5 to 1.9 of reference SCr within 7 days; stage 2 was defined as an increase of 2.0 to 2.9 of reference SCr; and stage 3 AKI was defined as an increase of ≥3 of reference SCr, an increase to ≥4 mg/dl, or when the patient required dialysis or kidney transplantation for AKI [1].

Model Development and Validation

The analytical process included four major stages: preprocessing, feature selection, prediction model construction, and scoring function. The first stage involved the selection of prespecified variables and imputation of missing values. The purpose of the second stage, feature selection, was to select important variables associated with the prediction outcome using state-of-the-art algorithms, XGBoost and LASSO. The third stage involved the construction of a prediction model according to data-driven technology. Finally, we used the model coefficients to build a scoring function, which outputs risk scores based on the prediction results. The whole process is schematically presented in Figure 1.

Figure 1. Flowchart for prediction and risk scoring. CGMHs: Chang Gung Memorial Hospitals.



Preprocessing

To estimate the risk of hospitalized patients with CA-AKI at admission for establishing a prevention strategy in an outpatient setting, the model only included predicted variables that were available before hospitalization. Notably, SCr measured on the index hospital admission and the physiological measurements after hospitalization were not included in the model.

Missing values are commonly present in medical records, and dropping medical records or variables with incomplete data would lead to small sample sizes. To develop a more precise model, we selected and discarded the variables with an original missing rate of more than 90% in a step-by-step manner to confirm which variable would significantly contribute to the model even though it had a high missing rate. We considered two approaches to impute the missing continuous values stratified by sex by replacing any missing value with the median or mean of the corresponding group. The experimental results indicated that imputation by the median stratified by sex yields better performance than that of imputation by the mean; thus, we applied imputation by the median stratified by sex to address the problem of missing values. There were no missing data for categorical variables in the dataset. Once the preprocessing step was completed, 47 variables remained for further processing (Table 1).

Table 1. Patient characteristics between the derivation and temporal validation cohorts.

Predictor candidates	Derivation cohort (n=204,064)				Temporal validation cohort (n=30,803)			
	n	CA-AKI[a] (n=17,230)	No CA-AKI (n=186,834)	*P* value[b]	n	CA-AKI (n=2218)	No CA-AKI (n=28,585)	*P* value[b]
Age at index hospitalization (years), mean (SD)		65.26 (15.49)	60.55 (16.23)	<.0001		65.69 (15.98)	60.12 (16.07)	<.0001
Sex, n (%)				.003				.21
Male	93,026	8041 (46.67)	84,985 (45.49)		14,644	1083 (48.83)	13,561 (47.44)	
Female	111,038	9189 (53.33)	101,849 (54.51)		16,159	1135 (51.17)	15,024 (52.56)	
Charlson comorbid condition at baseline, n (%)								
Acute myocardial infarction	4689	600 (3.48)	4089 (2.19)	<.0001	445	56 (2.52)	389 (1.36)	<.0001
Congestive heart failure	11,507	1903 (11.04)	9604 (5.14)	<.0001	1368	209 (9.42)	1159 (4.05)	<.0001
Peripheral vascular diseases	3976	609 (3.53)	3367 (1.80)	<.0001	342	43 (1.94)	299 (1.05)	.0001
Cerebral vascular accident	22,833	2622 (15.22)	20,211 (10.82)	<.0001	2991	286 (12.89)	2705 (9.46)	<.0001
Dementia	5101	693 (4.02)	4408 (2.36)	<.0001	382	54 (2.43)	328 (1.15)	<.0001
Pulmonary disease	18,930	1978 (11.48)	16,952 (9.07)	<.0001	2372	213 (9.60)	2159 (7.55)	<.001
Rheumatic disease	2387	253 (1.47)	2134 (1.14)	0.0001	437	37 (1.67)	400 (1.40)	.30
Peptic ulcer	27,709	2966 (17.21)	24,743 (13.24)	<.0001	3597	338 (15.24)	3259 (11.40)	<.0001
Mild liver diseases	30,682	3217 (18.67)	27,465 (14.70)	<.0001	2382	197 (8.88)	2185 (7.64)	.04
Diabetes without complication	45,795	6260 (36.33)	39,535 (21.16)	<.0001	5636	696 (31.38)	4940 (17.28)	<.0001
Diabetes with complications	12,017	2218 (12.87)	9799 (5.24)	<.0001	1863	330 (14.88)	1533 (5.36)	<.0001
Paraplegia	2484	248 (1.44)	2236 (1.20)	.006	290	19 (0.86)	271 (0.95)	.67
Renal disease	19,620	5603 (32.52)	14017 (7.50)	<.0001	2867	684 (30.84)	2183 (7.64)	<.0001
Any malignancy	50,927	4650 (26.99)	46,277 (24.77)	<.0001	6957	603 (27.19)	6354 (22.23)	<.0001
Severe liver diseases	3439	687 (3.99)	2752 (1.47)	<.0001	209	56 (2.52)	153 (0.54)	<.0001
Metastatic solid tumor	13,638	1459 (8.47)	12,179 (6.52)	<.0001	1788	199 (8.97)	1589 (5.56)	<.0001
Prior use of nephrotoxic medicine, n (%)								
NSAIDs[c] or COX II[d] inhibitors	61,320	4664 (27.07)	56,656 (30.32)	<.0001	8153	531 (23.94)	7622 (26.66)	.005
Opioid analgesics	14,511	1825 (10.59)	12,686 (6.79)	<.0001	1697	215 (9.69)	1482 (5.18)	<.0001
Any analgesics	67,782	5587 (32.43)	62,195 (33.29)	.02	8983	642 (28.94)	8341 (29.18)	.81
Antimicrobials[e]	53,454	5185 (30.09)	48,269 (25.84)	<.0001	7485	641 (28.90)	6844 (23.94)	<.0001
Antiepileptics (gabapentin or phenytoin)	1441	144 (0.84)	1297 (0.69)	.03	137	12 (0.54)	125 (0.44)	.48

Predictor candidates	Derivation cohort (n=204,064)				Temporal validation cohort (n=30,803)			
	n	CA-AKI[a] (n=17,230)	No CA-AKI (n=186,834)	*P* value[b]	n	CA-AKI (n=2218)	No CA-AKI (n=28,585)	*P* value[b]
Renin-angiotensin system inhibitors or potassium-sparing diuretics	50,879	6686 (38.80)	44,193 (23.65)	<.0001	6827	791 (35.66)	6036 (21.12)	<.0001
Contrast media	14,115	515 (2.99)	13,600 (7.28)	<.0001	2477	76 (3.43)	2401 (8.40)	<.0001
Nonmetformin OHA[f]	27,476	3780 (21.94)	23,696 (12.68)	<.0001	3678	433 (19.52)	3245 (11.35)	<.0001
Metformin OHA	14,059	1234 (7.16)	12,825 (6.86)	.14	1774	148 (6.67)	1626 (5.69)	.06
Any OHA	32,887	4178 (24.25)	28,709 (15.37)	<.0001	4488	505 (22.77)	3983 (13.93)	<.0001
Immunosuppressants	8100	900 (5.22)	7200 (3.85)	<.0001	1156	137 (6.18)	1019 (3.56)	<.0001
Antihyperuricemia	9588	1810 (10.50)	7778 (4.16)	<.0001	1232	266 (11.99)	966 (3.38)	<.0001
Antiinflammation/intestine	1167	75 (0.44)	1092 (0.58)	.01	153	9 (0.41)	144 (0.50)	.53
Antihistamines, antipsychotics, antispasmodics	37,791	4402 (25.55)	33,389 (17.87)	<.0001	5179	545 (24.57)	4634 (16.21)	<.0001
Bisphosphonates	822	68 (0.39)	754 (0.40)	.86	107	6 (0.27)	101 (0.35)	.52
Digoxin	2851	374 (2.17)	2477 (1.33)	<.0001	242	28 (1.26)	214 (0.75)	.008
Statins	24,818	2660 (15.44)	22,158 (11.86)	<.0001	4158	369 (16.64)	3789 (13.26)	<.0001
Fibrates	4024	478 (2.77)	3546 (1.90)	<.0001	468	50 (2.25)	418 (1.46)	.003
Lithium	145	9 (0.05)	136 (0.07)	.33	22	0 (0.00)	22 (0.08)	.19
Nitrates	12,339	1832 (10.63)	10,507 (5.62)	<.0001	1276	155 (6.99)	1121 (3.92)	<.0001
Anticoagulants	11,341	1392 (8.08)	9949 (5.33)	<.0001	2244	245 (11.05)	1999 (6.99)	<.0001
Baseline laboratory result, mean (SD)								
SCr[g]	204,064	2.43 (2.61)	1.02 (0.68)	<.0001	30,803	2.15 (2.32)	0.98 (0.59)	<.0001
eGFR[h]	204,064	66.68 (54.96)	82.47 (34.06)	<.0001	30,803	68.67 (52.78)	83.59 (31.80)	<.0001
BUN[i]	100,474	36.1 (28.06)	19.06 (14.00)	<.0001	14,674	34.79 (27.56)	18.5 (12.91)	<.0001
Total cholesterol	17,570	173.6 (39.41)	179.24 (36.74)	<.0001	2499	176.8 (43.40)	177.83 (37.17)	.76
LDL[j]-cholesterol	57,784	99.46 (31.45)	103.27 (30.43)	<.0001	9452	97.42 (31.78)	102.38 (30.43)	<.001
Triglyceride	63,600	141.2 (80.91)	134.87 (77.48)	<.0001	9612	140.76 (82.69)	135.1 (78.73)	.05
Serum uric acid	59,096	7.04 (2.33)	6.3 (1.95)	<.0001	8729	6.52 (2.26)	5.98 (1.86)	<.0001
Calcium	56182	8.66 (0.76)	8.86 (0.64)	<.0001	7988	8.64 (0.74)	8.92 (0.65)	<.0001

Predictor candidates	Derivation cohort (n=204,064)				Temporal validation cohort (n=30,803)			
	n	CA-AKI[a] (n=17,230)	No CA-AKI (n=186,834)	*P* value[b]	n	CA-AKI (n=2218)	No CA-AKI (n=28,585)	*P* value[b]
Phosphorus	36181	4.32 (1.20)	3.59 (0.75)	<.0001	5053	4.28 (1.21)	3.61 (0.73)	<.0001

[a]CA-AKI: community-acquired acute kidney injury.

[b]Independent *t* tests were performed for continuous data, and Pearson Chi-square tests were performed for categorical data in between-groups comparisons.

[c]NSAIDs: nonsteroidal anti-inflammatory drug.

[d]COX II: cyclooxygenase 2.

[e]Antimicrobials include aminoglycosides, penicillins, antivirals, trimoxazole/trimethoprim, fluconazole, teicoplanin/vancomycin, or tetracycline.

[f]OHA: oral hypoglycemic agent.

[g]SCr: serum creatinine.

[h]eGFR: estimated glomerular filtration rate ($175 \times SCr^{-1.154} \times age^{-0.203}$ [12] $\times$ [0.742,female]).

[i]BUN: blood urea nitrogen.

[j]LDL: low-density lipoprotein.

Feature Selection

To build a scoring function to help clinicians effectively assess the risk of CA-AKI hospitalization, the number of variables (10) involved in the scoring function was considered based on the commonly used scoring functions in health care and our expert opinion. For instance, there are 14 parameters in the APACHE II score for mortality prediction in a critical care setting (ie, age, temperature, mean atrial pressure, pH, heart rate/pulse, respiratory rate, sodium, potassium, creatinine, acute kidney failure, hematocrit, white blood cell count, Glasgow Coma Scale, and FiO_2) [23] and 8 parameters in the CHA2DS2VASc score for thromboembolism risk in atrial fibrillation (ie, congestive heart failure, hypertension, age≥75, diabetes, stroke/transient ischemic attack/thrombo-embolism, vascular disease, age 65-74, and sex) [24]. We used the feature selection technique to determine the most important features. An exhaustive search of the best combination of features can conceivably be performed on problems with few features. However, the problem is known to be a nondeterministic polynomial time (NP)-hard problem [25], meaning that the search quickly becomes computationally intractable. Therefore, we used XGBoost and LASSO to perform feature selection.

LASSO is a regression analysis method that uses L1 constraint to perform variable selection and regularization, providing a base to select a subset of the available covariates for use in the final model. XGBoost is an improved algorithm based on the gradient boosting decision tree, which can help avoid model overfitting [26] by considering L1 and L2 constraints in the objective function. The XGBoost model always involves many classification and regression trees, each of which comprises splitting nodes during model learning. Each splitting node corresponds to a variable or feature. This study considered the average gain of the feature when it is used in the trees. The average gain is obtained by calculating the average improvement in accuracy brought about by a feature to the branches it belongs.

Prediction Model Construction

After completing feature selection, we used the top 10 important features to build a prediction model for CA-AKI hospitalization. As mentioned above, feature selection is an NP-hard problem. Different methods resulted in the same result for feature selection, demonstrating that the selected features are important from two different perspectives (Multimedia Appendix 3). The performance of prediction models with all, 10, and 5 features was examined to ensure the appropriateness of a 10-features predictive CA-AKI hospitalization risk model (Multimedia Appendix 4).

Next, based on the 10 selected features, we used a logistic regression model as the prediction model because it could reveal the coefficients of the 10 features, which facilitates interpretation for medical personnel to assess the magnitude of the relationship between the individual and outcome variables, and to understand how the outcomes are induced from the model. The ability of model discrimination was determined with the area under the receiver operating characteristic (ROC) curve (AUC). We applied 5-fold crossvalidation to ensure that all data points were used for model training and evaluation so that the obtained model could be generalized to unseen data. Moreover, the 95% CIs for the metrics, including AUC, sensitivity, and specificity, were also determined.

Scoring Function

We used the coefficients of the logistic regression model to build a scoring function, which could be used in clinical settings and provide more explanatory power. The outcome of the logistic regression model was the probability of CA-AKI hospitalization. We transformed the probability into a score by multiplying the probability by 100. To distinguish between CA-AKI and non-CA-AKI patients from the scores, it is necessary to determine a threshold value.

We considered two methods to determine the cut-off point for distinguishing between CA-AKI and non-CA-AKI patients. The first method was based on the Youden index [27] to determine the point at which the summation of sensitivity and specificity is maximal, and the second method determined the point that yields the highest sensitivity with a minimum specificity of 0.7 [16]; here, we designate the former as the regular threshold and the latter as the special threshold. With these thresholds, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were

calculated to analyze the diagnostic ability of the proposed scoring function.

After rescaling the output of the scoring function, we designed an easy-to-use app with an Excel worksheet so that clinicians can estimate the risk score of a patient based on the patient's relevant data. For missing data entries for a patient, the app uses a default value to perform the estimation, which is obtained from the median of the continuous variable or the mode of the discrete variable stratified by sex.

Validation

The temporal validation cohort included hospitalized adults identified during 2017, which was used as the validation set. The scoring function developed in the derivation cohort was applied to the validation cohort to ensure its performance and generalization ability. The AUC and sensitivity values in the validation set were similar to those in the training set, indicating that the proposed method did not suffer from the overfitting problem.

Statistical Analysis

The DeLong test was applied to assess the generalization ability of the proposed scoring function to ensure the prediction capability to unseen data. We tested the ROC curves of the scoring function on the derivation and validation cohorts, and set the significance level at .05.

Sensitivity Analysis

Sensitivity analysis was performed to assess the proposed scoring function regarding the discrimination of patients with severe AKI (AKI stage 2-3). In the experimental setting, we relabeled the outcome of the same dataset based on the information on AKI stage. Following the same setting, we divided the dataset into the derivation cohort (2010-2016) and the temporal validation cohort (2017). In the previous setting, the goal was to develop a model to predict whether a patient would suffer from CA-AKI, and therefore separated the whole population into two groups: CA-AKI patients (case) and non-CA-AKI patients (control). In contrast to the previous experiment, the sensitivity analysis included patients without CA-AKI and those with stage 1 CA-AKI in the control group, whereas patients with stage 2 and 3 CA-AKI were included in the case group. We subsequently used the developed scoring function and the same cut-off thresholds to predict the relabeled dataset to investigate whether the scoring function could better distinguish between severe CA-AKI and non-CA-AKI groups.

Results

Characteristics of the Study Cohort

A total of 234,867 patients with hospital admission were analyzed in the final CA-AKI cohort; there were 48% admissions from the emergency department and 52% admissions from the outpatient setting between January 1, 2010 and December 31, 2017. The rate of CA-AKI was 8.44% (17,230/204,064) in the derivation cohort and 7.20% (2218/30,803) in the temporal validation cohort (Figure 1). The mean age of patients in the CA-AKI group was higher than that of patients in the non-CA-AKI group in the derivation cohort, which were similar to the temporal validation cohort (Table 1). The frequency of patients with an estimated glomerular filtration rate (eGFR) below 60 ml/min/1.73m^2 at baseline was higher in the CA-AKI group than that in the non-CA-AKI group (50.67%, 8732/17,230 vs 24.86%, 46,445/186,834) in the derivation cohort, as well as for patients in the validation cohort (46.75%, 1037/2218 vs 21.72%, 6207/28,585). The mean levels of SCr and eGFR at baseline in both the derivation and validation cohorts are presented in Table 1. Compared to the non-CA-AKI group, patients with CA-AKI had a higher mean CCI score (3.08, SD 2.33 vs 1.83, SD 2.02) and had more frequent use of renin-angiotensin system (RAS) inhibitors or diuretics (38.8% vs 23.65%) and antimicrobials (30.09% vs 25.84%) in the baseline period in both the derivation and validation cohorts (Table 1).

Model Performance

The LASSO and XGBoost models selected the same top 10 important variables among the 47 variables in the derivation cohort, and the AUC (0.789, 95% CI 0.785-0.793) was slightly higher for XGBoost than for the LASSO model (0.7671, 95% CI 0.7621-0.7721) (Multimedia Appendix 3). Table 2 shows the top 10 variables. In the training of the derivation cohort, the logistic regression model had an AUC of 0.7670 (95% CI 0.7608-0.7732), sensitivity of 0.6142 (95% CI 0.5855-0.6431), and specificity of 0.7848 (95% CI 0.7529-0.8167). In addition to the predictive model, the coefficients for the 10 variables used to develop our proposed scoring function are listed in Table 3.

Table 2. Top 10 features selected by the extreme gradient boost (XGBoost) and least absolute shrinkage and selection operator (LASSO) algorithms.

Type	Important features
Basic information	Age at index hospitalization
Charlson comorbid condition	Diabetes without complication, Chronic kidney disease, Severe liver diseases
Prior use of nephrotoxic medicine	RAS[a] inhibitors/K-sparing diuretics
Baseline laboratory result	Serum creatinine, eGFR[b], BUN[c], Calcium, Phosphorus

[a]RAS: renin-angiotensin system.

[b]eGFR, estimated glomerular filtration rate (175 × SCr − 1.154 × age − 0.203 × [0.742, female]).

[c]BUN: blood urea nitrogen.

Table 3. Community-acquired acute kidney injury risk coefficients in the final model.[a]

Variable (Xi)	Coefficient (βi)
SCr[b]	0.7244
Age	0.0207
eGFR[c]	0.0169
BUN[d]	0.0072
Calcium	–0.4669
Phosphorus	0.3542
DM[e]	0.3065
CKD[f]	0.6235
SLD[g]	0.9647
RAS[h] inhibitors/ K-sparing diuretics	0.4099

[a]Intercept of the model: –3.6838.

[b]SCr: serum creatinine.

[c]eGFR: estimated glomerular filtration rate.

[d]BUN: blood urea nitrogen.

[e]DM: diabetes without complication.

[f]CKD: chronic kidney disease.

[g]SLD: severe liver disease.

[h]RAS: renin-angiotensin system.

Scoring Function for CA-AKI Hospitalization

The scoring function was established based on the coefficients obtained from the logistic regression model. Equation (1) shows the formula of the scoring function *z*, in which βi is the coefficient for the *i*th feature *Xi*. Detailed definitions of the variables and their corresponding coefficients are presented in Table 3. The final score was obtained by transforming the value of *z* into a probability with a sigmoid function, and then multiplying it by 100 to make the risk score range from 0 to 100.

To verify the generalization ability of the proposed scoring function, the DeLong test was applied to the ROC curves of the derivation and validation cohorts. As shown in Figure 2, the *P* value of the test was approximately .30, indicating no statistically significant difference between the ROC curves of the derivation and validation cohorts and that the scoring function does not suffer from the overfitting problem. A higher score indicates a higher the risk of CA-AKI hospitalization. Figure 3 shows the risk score distributions between the case (CA-AKI) and control (non-CA-AKI) groups for the derivation and validation cohorts. The experimental results indicated that the risk score of the case group is normally higher than that of the control group in the derivation and validation cohorts, meaning that the scoring function could stratify the two groups well.

Figure 2. DeLong test for the receiver operating characteristic curves of derivation and validation cohorts.

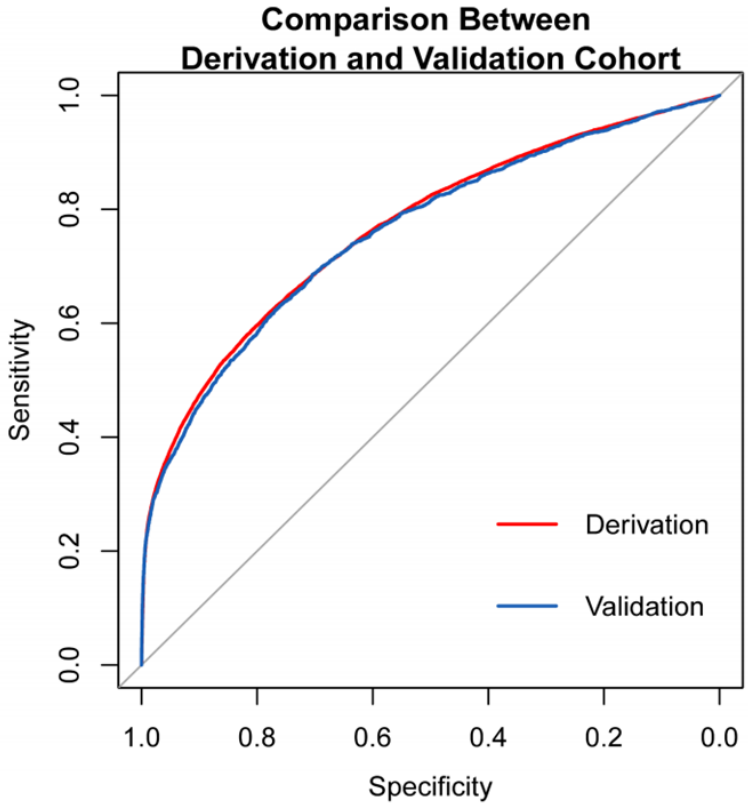


Figure 3. Risk score distribution. Left: Derivation cohort with CA-AKI stages 1-3 (case). Right: Validation cohort with CA-AKI stages 1-3 (case).

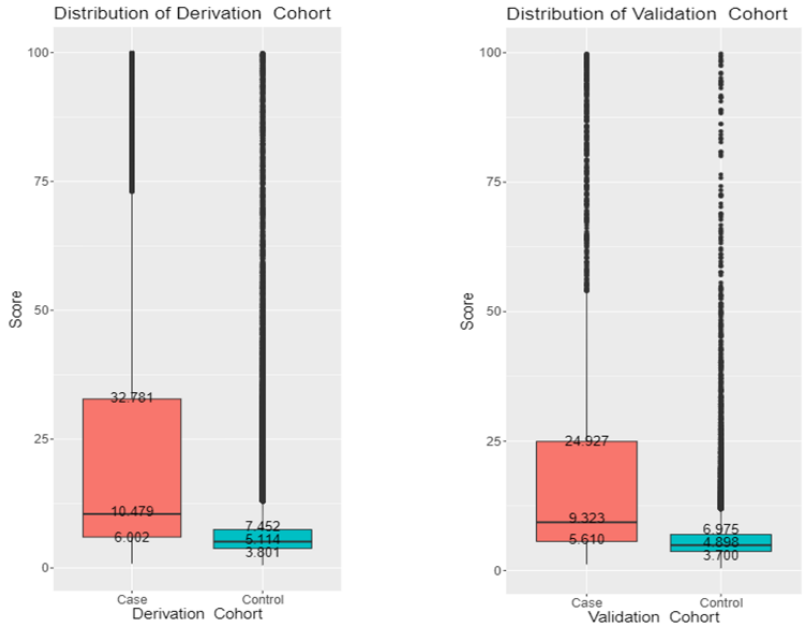


Table 4 shows the results of model performance, demonstrating that the scoring function could achieve better results in sensitivity and NPV by setting a special threshold as the cut-off point. The ROC curves for the cut-off thresholds determined by the two methods are presented in Figure 4, in which the values in parentheses are specificity and sensitivity. Finally, the risk equation for CA-AKI hospitalization risk was established in an Excel worksheet (Multimedia Appendix 5) to allow for automatic computation by importing patient information in the clinical decision support system.

Table 4. Model performance in the derivation and validation cohorts.

Performance metric[a]	Cut-off point with regular threshold (7.993)			Cut-off point with special threshold (6.804)		
	CA-AKI[b] stages 1-3		CA-AKI stages 2 and 3	CA-AKI stages 1-3		CA-AKI stages 2 and 3
	Derivation cohort	Validation cohort	Validation cohort	Derivation cohort	Validation cohort	Validation cohort
AUC[c]	0.767 (0.758-0.777)[d]	0.761	0.818	0.767 (0.758-0.777)	0.761	0.818
Sensitivity	0.612 (0.591-0.634)	0.569	0.689	0.687 (0.665-0.708)	0.651	0.75
Specificity	0.785 (0.782-0.788)	0.814	0.807	0.700 (0.694-0.706)	0.736	0.728
PPV[e]	0.208 (0.201-0.215)	0.192	0.133	0.174 (0.169-0.180)	0.161	0.106
NPV[f]	0.956 (0.954-0.959)	0.961	0.984	0.960 (0.958-0.963)	0.964	0.985

[a]The performance for each cohort was evaluated based on disease severity and cut-off threshold values.

[b]CA-AKI: community-acquired acute kidney injury.

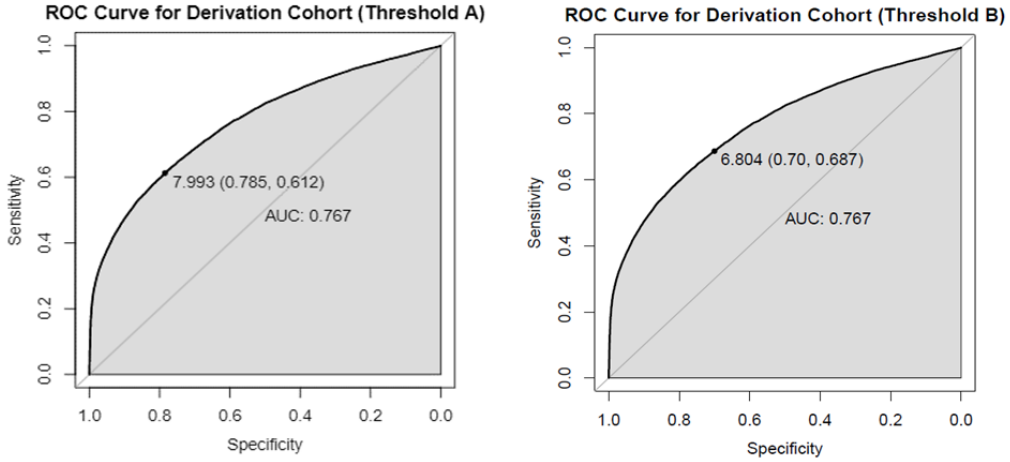
[c]AUC: area under the receiver operating characteristic curve.

[d]The values in the parentheses are 95% CIs calculated through 5-fold crossvalidation.

[e]PPV: positive predictive value.

[f]NPV: negative predictive value.

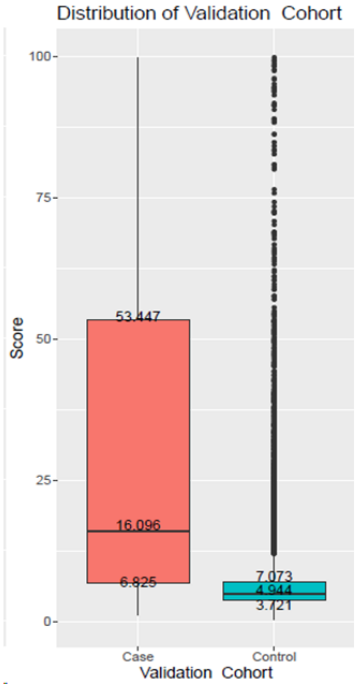
Figure 4. Receiver operating characteristic (ROC) curve for the derivation cohort. Threshold A: Cut-off regular threshold value of 7.993; Threshold B: Cut-off special threshold value of 6.804.



Sensitivity Analysis

The sensitivity analysis was conducted with the original model using the relabeled validation data to evaluate model performance in distinguishing between severe AKI and less severe AKI cases. The risk score distribution between the case and control groups is depicted in Figure 5, which is more distinguishable. Moreover, sensitivity and specificity were more balanced compared to the original values, regardless of using the regular threshold (7.993) or special threshold (6.804) as the cut-off value in Figure 4. The experimental results are shown in Table 4. Using the cut-off special threshold, the AUC value was 0.818, sensitivity was 0.75, and NPV was 0.985 in the temporal validation cohort, which were all better than the original values, except for PPV, indicating that the proposed scoring function performed better in distinguishing severe CA-AKI.

Figure 5. Risk score distribution of the validation cohort with stage 2-3 community-acquired-acute kidney injury (case).



Discussion

Principal Findings

To the best of our knowledge, this is the first study to use a machine-learning model to develop a 10-variable scoring function for assessing the risk of CA-AKI hospitalization. The advantages of the predictive risk model include prediction based on routinely available EHR data in practice and full applicability to general patients in an outpatient setting. Most importantly, the quantified risk score can serve as an assessment tool to support preventive management for the risks of CA-AKI hospitalization that can be modified. The two proposed methods, XGBoost and LASSO, selected identical top 10 features from different perspectives to support the importance of these predictors in the scoring function.

LASSO is a statistical method that can improve prediction accuracy and model interpretation by imposing an L1 penalty, resulting in a sparse model. Notably, LASSO shrinks some model coefficients to zero, providing a base to eliminate variables whose coefficients are not statistically different from zero. In contrast, XGBoost is an ensemble machine-learning model involving multiple decision trees, and it can estimate feature importance by considering the contribution of a specific feature for each tree during the learning process.

Importantly, XGBoost reflects the diverse etiologies of CA-AKI (eg, physiological status of the patient, coexisting medical problems, and underlying causes) affecting the general population. The experimental results indicated that using the proposed scoring function for assessing the risk of CA-AKI hospitalization showed fairly good to good performance with respect to the AUC (0.76-0.82) to detect any stage or moderate to severe stages of CA-AKI by decision thresholds on the validation models.

Similar to previous AKI investigations, underlying comorbidities (diabetes mellitus, severe liver diseases, chronic kidney disease) [28,29] and recent use of RAS inhibitors or potassium-sparing diuretics [19] were identified in the present CA-AKI hospitalization risk score. The risk model also indicated that low calcium levels and high phosphorus levels were potentially modifiable predictors and could be targeted for correction [30]. This feature has substantial clinical implications because it demonstrates that the model can be applied to prospectively support clinical decision systems in real time for rapid screening and recognition of patients with a predicted risk of CA-AKI in outpatient settings.

Early-stage AKI is generally asymptomatic, and therefore SCr monitoring is required. Previous findings have suggested that even small changes are common and are associated with increased mortality and length of hospital stay [31]. However, because a baseline SCr measurement is not always available in practice settings, risk assessment for mild CA-AKI hospitalization in a diverse population can be a challenge. The present study is one of the few CA-AKI studies that included patients with an SCr measurement in the community as the baseline level of renal function, which was compared to another SCr measurement performed at hospital admission or requested by the general practitioner to better define the nature of CA-AKI [3,4,32,33]. In this study, patients with CA-AKI in the final cohort were older (65.26, SD 15.49 years) than those in the non-CA-AKI group; moreover, 54.4% of the patients were women and 50.67% had preexisting chronic kidney disease (eGFR<60 ml/min/1.73m^2 at baseline). In comparison, a study with a British population reported a mean age of 74.4 (SD 15.4) years, 50%-52% female patients, and 31.9%-34.6% of patients with preexisting chronic kidney disease [4,30], whereas another study with a US population reported a mean age of 67.8 (SD 12.2) years and 43.2% of patients with preexisting chronic kidney disease [3]; these populations were considered to be

comparable with the CA-AKI patients identified using KDIGO SCr-based criteria in different populations. The present study used large-scale clinical data with a representative and adequate sample size to develop a diagnostic tool for CA-AKI risk evaluation.

Because of a lack of data regarding the risk modeling of CA-AKI that included up to a few hundred cases, we summarized the major findings in large, TRIPOD-adherent studies (published after 2010) that have analyzed the risk of AKI development in the short term (within 3 or 7 days) following general admission (see Multimedia Appendix 6) [34-36]. In those studies, variable candidates were first selected according to significance (usually $P<.05$) in the univariate analysis and determined in the multivariate logistic regression analysis with stepwise selection. The handling of missing data in the variables was usually not addressed in these previous risk models. The AUC value was calculated in both training and validation data to indicate the capability of discrimination. Hosmer Lemeshow analysis and *P* values were nonsignificant, suggesting acceptable calibration.

A prediction model for the risk of AKI 72 hours following admission was established using data from 3 centers in the United Kingdom [34]. Based on 35 variables collected within 24 hours after admission in 2011, 12 predictors were selected, including age, primary diagnosis, previous admissions, CCI score, HbA1C, troponin, proteinuria, eGFR, potassium, magnesium, C-reactive protein, and white blood cell count. The AUC in the derivation cohort (n=6626) was 0.67 (95% CI 0.64-0.71) in the internal validation cohort for any AKI risk and was 0.71 (95% CI 0.67-0.76) in the external validation model (n=1585) [34].

Another prospective study using prospectively collected AKI screening data in a single center in the United Kingdom in 2011 predicted the risk of AKI less than 7 days after admission [35]. The baseline SCr was retrieved between 1 and 6 months prior to hospitalization. Of the 25 collected variables, the following 7 variables were selected in the multivariate logistic model: age (60-79, ≥80 years), congestive cardiac failure, chronic kidney disease, diabetes, liver disease, respiratory rate ≥20/min, alert, verbal, pain, unresponsive status. The AUC value was 0.72 (95% CI 0.66-0.77) and 0.76 (95% CI 0.71-0.82) for the derivation and internal validation cohort, respectively. A recent external validation study in a single UK nonspecialist acute hospital (2013-2015) reported that the AUC of the prediction model was 0.65 (95%CI 0.62-0.67) in the medical setting and was 0.66 (95%CI 0.62-0.70) in the surgical setting [36]. In addition, the sensitivity analysis showed that for patients without baseline SCr information across the medical and surgical cohorts, the AUC was 0.71 (95%CI 0.67-0.75) and 0.68 (0.58-0.75), respectively, indicating poor to fair predictive capability (range 0.65-0.71) [36].

Strengths and Limitations

The current risk prediction model included patients with CA-AKI at hospital admission with modifiable and nonmodifiable predictors commonly measured in routine care. Excluding SCr on the index hospital admission, XGBoost and LASSO demonstrated the feasibility of using machine learning for predicting CA-AKI hospitalization risk in the general population. The present model, which used large-scale clinical data with a representative and adequate sample size and top 10 important predictors having clinical significance on prevention of CA-AKI requiring inpatient care, can be considered as a benchmark for further evaluations.

Another strength of the present risk model is that the continuous risk score of the model can be incorporated into clinical decision support systems, facilitating their usability. Because the CA-AKI hospitalization rate was considerably low in the present study cohort, a risk score over 7.993 was associated with a low PPV (20%) but high NPV (96%), suggesting an ability to correctly identify low-risk patients (ruled out). For instance, the CA-AKI hospitalization risk equation can be easily fitted with the most recent real-time clinical data for automatic screening and implementation of preventive strategies (ie, to stop nephrotoxic medication or ordering nephrologists referred care) in general outpatient and emergency department settings. Recently, a digital-based AKI care pathway incorporating mobile phone detection with a multidisciplinary care response team and care protocol was proposed for the UK National Health System, which showed significant practical value in the field [37]. Furthermore, the results of implementing the CA-AKI risk score in practical settings can help to prospectively evaluate its impact on the quality of patient care, such as reducing AKI risk exposure, preventive measures, and management to avoid renal insults.

This study has several limitations. First, it is based on data obtained from a large hospital cohort in Taiwan. This could affect the generalizability of the study findings, although based on previous data, patient characteristics included in the current study are not different from those of other populations. Similar to previous retrospective studies, medical histories relied on previously coded events in EHRs, and thus possible residual risk may have been underestimated in the baseline period, thereby increasing the uncertainty of the risk prediction performance. Biomarkers of kidney injury such as cystatin C and neutrophil gelatinase-associated lipocalin, which are important to provide diagnostic information but are not routinely tested in practice settings, were absent from the developed model in the present study. Additionally, the study did not address external validation. Although it is clear that the integration of different EHR systems remains a challenge to the medical community in many health systems, including Taiwan, it is still necessary to conduct further studies to leverage the risk scoring system to available EHRs. External validations using this prediction model outside of a study setting and in different geographic populations are envisioned as future work. Lastly, the current study focused on the feasibility of using a machine-learning model for CA-AKI hospitalization risk prediction, but did not examine the incidence of CA-AKI, cause, and management of CA-AKI or its implementation on patient care. The prediction model and risk scoring function developed in the present study can nevertheless serve as a risk assessment tool and link clinical decision support systems for prospective validation.

Conclusion

This study demonstrated that the selected variables were truly crucial for predicting the risk of CA-AKI hospitalization, as both XGBoost and LASSO identified the same top 10 important variables. The discrimination ability of the CA-AKI risk scoring function was good with an AUC value of 0.76 and 0.82 in the detection of CA-AKI hospitalization at any stage and for moderate to severe stages, respectively, according to decision thresholds on the validation cohort, and suggested the feasibility of AKI detection and prevention in wider populations in the community. In addition, the easy-to-use risk calculator can facilitate its widespread implementation in daily routines and workflows for patient-centered care and prospective validation of machine-learning applications.

Acknowledgments

The authors thank Mrs. Chun-Hua Liao and Ying-Jen Hsu from the Department of Management of Information Systems, Chang Gung Memorial Foundation, for their technical assistance with electronic health record acquisition and data retrieval. The authors also appreciate Miss Pei-Chi Liu and Miss Hsiao-Ching Kuo from the Department of Pharmacy for their assistance in the data cleaning process. This work is supported by Kaohsiung Chang Gung Memorial Hospital (CMRPG8H0961). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

CH and YT conceived of the study. CH performed data curation. CL, CK, and YL conducted the formal analysis. CH, YT, CL, CK, and YL performed the investigation. CH and CL contributed to the methodology. CL, CK, and YL developed the software. CH, CL, CK, and YL performed validation analyses. YT was responsible for results visualization. CH and CL wrote the first draft of the manuscript, and CH, YT, CL, CK, and YL contributed to writing, editing, and review of the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient selection process.
[PDF File (Adobe PDF File), 40 KB - jmir_v22i8e16903_app1.pdf]

Multimedia Appendix 2
Data variables used in the machine learning models.
[DOCX File , 22 KB - jmir_v22i8e16903_app2.docx]

Multimedia Appendix 3
Number of predictors and AUC of XGBoost models.
[DOCX File , 18 KB - jmir_v22i8e16903_app3.docx]

Multimedia Appendix 4
The top 10 important variables and model performance between XGBoost and LASSO algorithms.
[DOCX File , 25 KB - jmir_v22i8e16903_app4.docx]

Multimedia Appendix 5
CA-AKI hospitalization risk calculator.
[PDF File (Adobe PDF File), 81 KB - jmir_v22i8e16903_app5.pdf]

Multimedia Appendix 6
Traditional prediction models for AKI developed in an acute medical setting or shortly after general admission.
[DOCX File , 22 KB - jmir_v22i8e16903_app6.docx]

References

1. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. Nephron Clin Pract 2012;120(4):c179-c184 [FREE Full text] [doi: 10.1159/000339789] [Medline: 22890468]
2. Hsu C, Lee C, Su C, Wang YL, Chen H, Chuang J, et al. Incidence, Outcomes, and Risk Factors of Community-Acquired and Hospital-Acquired Acute Kidney Injury: A Retrospective Cohort Study. Medicine (Baltimore) 2016 May;95(19):e3674. [doi: 10.1097/MD.0000000000003674] [Medline: 27175701]

3. Schissler MM, Zaidi S, Kumar H, Deo D, Brier ME, McLeish KR. Characteristics and outcomes in community-acquired versus hospital-acquired acute kidney injury. Nephrology (Carlton) 2013 Mar;18(3):183-187. [doi: 10.1111/nep.12036] [Medline: 23336108]
4. Wonnacott A, Meran S, Amphlett B, Talabani B, Phillips A. Epidemiology and outcomes in community-acquired versus hospital-acquired AKI. Clin J Am Soc Nephrol 2014 Jun 06;9(6):1007-1014 [FREE Full text] [doi: 10.2215/CJN.07920713] [Medline: 24677557]
5. Soto K, Campos P, Pinto I, Rodrigues B, Frade F, Papoila AL, et al. The risk of chronic kidney disease and mortality are increased after community-acquired acute kidney injury. Kidney Int 2016 Nov;90(5):1090-1099 [FREE Full text] [doi: 10.1016/j.kint.2016.07.018] [Medline: 27658322]
6. Collister D, Pannu N, Ye F, James M, Hemmelgarn B, Chui B, Alberta Kidney Disease Network. Health Care Costs Associated with AKI. Clin J Am Soc Nephrol 2017 Nov 07;12(11):1733-1743 [FREE Full text] [doi: 10.2215/CJN.00950117] [Medline: 29051143]
7. Silver SA, Chertow GM. The Economic Consequences of Acute Kidney Injury. Nephron 2017;137(4):297-301 [FREE Full text] [doi: 10.1159/000475607] [Medline: 28595193]
8. Kashani K, Rosner MH, Haase M, Lewington AJP, O'Donoghue DJ, Wilson FP, et al. Quality Improvement Goals for Acute Kidney Injury. Clin J Am Soc Nephrol 2019 Jun 07;14(6):941-953 [FREE Full text] [doi: 10.2215/CJN.01250119] [Medline: 31101671]
9. Thottakkara P, Ozrazgat-Baslanti T, Hupf BB, Rashidi P, Pardalos P, Momcilovic P, et al. Application of Machine Learning Techniques to High-Dimensional Clinical Data to Forecast Postoperative Complications. PLoS One 2016;11(5):e0155705 [FREE Full text] [doi: 10.1371/journal.pone.0155705] [Medline: 27232332]
10. Fritz BA, Chen Y, Murray-Torres TM, Gregory S, Ben Abdallah A, Kronzer A, et al. Using machine learning techniques to develop forecasting algorithms for postoperative complications: protocol for a retrospective study. BMJ Open 2018 Apr 10;8(4):e020124 [FREE Full text] [doi: 10.1136/bmjopen-2017-020124] [Medline: 29643160]
11. Huang C, Murugiah K, Mahajan S, Li S, Dhruva SS, Haimovich JS, et al. Enhancing the prediction of acute kidney injury risk after percutaneous coronary intervention using machine learning techniques: A retrospective cohort study. PLoS Med 2018 Nov;15(11):e1002703 [FREE Full text] [doi: 10.1371/journal.pmed.1002703] [Medline: 30481186]
12. Ibrahim NE, McCarthy CP, Shrestha S, Gaggin HK, Mukai R, Magaret CA, et al. A clinical, proteomics, and artificial intelligence-driven model to predict acute kidney injury in patients undergoing coronary angiography. Clin Cardiol 2019 Feb;42(2):292-298. [doi: 10.1002/clc.23143] [Medline: 30582197]
13. Lee H, Yoon SB, Yang S, Kim WH, Ryu H, Jung C, et al. Prediction of Acute Kidney Injury after Liver Transplantation: Machine Learning Approaches vs. Logistic Regression Model. J Clin Med 2018 Nov 08;7(11):428 [FREE Full text] [doi: 10.3390/jcm7110428] [Medline: 30413107]
14. Koyner JL, Carey KA, Edelson DP, Churpek MM. The Development of a Machine Learning Inpatient Acute Kidney Injury Prediction Model. Crit Care Med 2018 Jul;46(7):1070-1077. [doi: 10.1097/CCM.0000000000003123] [Medline: 29596073]
15. Bhatraju PK, Zelnick LR, Katz R, Mikacenic C, Kosamo S, Hahn WO, et al. A Prediction Model for Severe AKI in Critically Ill Adults That Incorporates Clinical and Biomarker Data. Clin J Am Soc Nephrol 2019 Apr 05;14(4):506-514 [FREE Full text] [doi: 10.2215/CJN.04100318] [Medline: 30917991]
16. Chiofolo C, Chbat N, Ghosh E, Eshelman L, Kashani K. Automated Continuous Acute Kidney Injury Prediction and Surveillance: A Random Forest Model. Mayo Clin Proc 2019 May;94(5):783-792. [doi: 10.1016/j.mayocp.2019.02.009] [Medline: 31054606]
17. Shao SC, Chan YY, Kao Yang YH, Lin SJ, Hung MJ, Chien RN, et al. The Chang Gung Research Database-A multi-institutional electronic medical records database for real-world epidemiological studies in Taiwan. Pharmacoepidemiol Drug Saf 2019 May;28(5):593-600. [doi: 10.1002/pds.4713] [Medline: 30648314]
18. Collins GS, Reitsma JB, Altman DG, Moons KGM, TRIPOD Group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. Circulation 2015 Jan 13;131(2):211-219 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.014508] [Medline: 25561516]
19. Stucker F, Ponte B, De la Fuente V, Alves C, Rutschmann O, Carballo S, et al. Risk factors for community-acquired acute kidney injury in patients with and without chronic kidney injury and impact of its initial management on prognosis: a prospective observational study. BMC Nephrol 2017 Dec 29;18(1):380 [FREE Full text] [doi: 10.1186/s12882-017-0792-2] [Medline: 29287584]
20. Murugan R, Weissfeld L, Yende S, Singbartl K, Angus DC, Kellum JA, GeneticInflammatory Markers of Sepsis (GenIMS) Investigators. Association of statin use with risk and outcome of acute kidney injury in community-acquired pneumonia. Clin J Am Soc Nephrol 2012 Jun;7(6):895-905 [FREE Full text] [doi: 10.2215/CJN.07100711] [Medline: 22461537]
21. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. J Clin Epidemiol 2004 Dec;57(12):1288-1294. [doi: 10.1016/j.jclinepi.2004.03.012] [Medline: 15617955]
22. Thomas ME, Blaine C, Dawnay A, Devonald MAJ, Ftouh S, Laing C, et al. The definition of acute kidney injury and its use in practice. Kidney Int 2015 Jan;87(1):62-73 [FREE Full text] [doi: 10.1038/ki.2014.328] [Medline: 25317932]

23. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II. Critical Care Medicine 1985;13(10):818-829. [doi: 10.1097/00003246-198510000-00009]
24. Lip GYH, Nieuwlaat R, Pisters R, Lane DA, Crijns HJGM. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest 2010 Feb;137(2):263-272. [doi: 10.1378/chest.09-1584] [Medline: 19762550]
25. Amaldi E, Kann V. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. Theor Comput Sci 1998 Dec;209(1-2):237-260. [doi: 10.1016/s0304-3975(97)00115-1]
26. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. : ACM; 2016 Presented at: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco p. 785-794 URL: https://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf [doi: 10.1145/2939672.2939785]
27. Youden WJ. Index for rating diagnostic tests. Cancer 1950 Jan;3(1):32-35. [doi: 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3] [Medline: 15405679]
28. Jain A, McDonald HI, Nitsch D, Tomlinson L, Thomas SL. Risk factors for developing acute kidney injury in older people with diabetes and community-acquired pneumonia: a population-based UK cohort study. BMC Nephrol 2017 May 01;18(1):142 [FREE Full text] [doi: 10.1186/s12882-017-0566-x] [Medline: 28460637]
29. Mesropian PD, Othersen J, Mason D, Wang J, Asif A, Mathew RO. Community-acquired acute kidney injury: A challenge and opportunity for primary care in kidney health. Nephrology (Carlton) 2016 Sep;21(9):729-735. [doi: 10.1111/nep.12751] [Medline: 26890822]
30. Leaf DE, Christov M. Dysregulated Mineral Metabolism in AKI. Semin Nephrol 2019 Jan;39(1):41-56. [doi: 10.1016/j.semnephrol.2018.10.004] [Medline: 30606407]
31. Cruz DN, Ferrer-Nadal A, Piccinni P, Goldstein SL, Chawla LS, Alessandri E, NEFROINT Investigators. Utilization of small changes in serum creatinine with clinical risk factors to assess the risk of AKI in critically lll adults. Clin J Am Soc Nephrol 2014 Apr;9(4):663-672 [FREE Full text] [doi: 10.2215/CJN.05190513] [Medline: 24677553]
32. Sawhney S, Fluck N, Fraser SD, Marks A, Prescott GJ, Roderick PJ, et al. KDIGO-based acute kidney injury criteria operate differently in hospitals and the community-findings from a large population cohort. Nephrol Dial Transplant 2016 Jun;31(6):922-929 [FREE Full text] [doi: 10.1093/ndt/gfw052] [Medline: 27190340]
33. Talabani B, Zouwail S, Pyart RD, Meran S, Riley SG, Phillips AO. Epidemiology and outcome of community-acquired acute kidney injury. Nephrology (Carlton) 2014 May;19(5):282-287. [doi: 10.1111/nep.12221] [Medline: 24571827]
34. Bedford M, Stevens P, Coulton S, Billings J, Farr M, Wheeler T. Development of risk models for the prediction of new or worsening acute kidney injury on or during hospital admission: a cohort and nested study. In: Health Services and Delivery Research. Southampton: NIHR Journals Library; 2016.
35. Forni LG, Dawes T, Sinclair H, Cheek E, Bewick V, Dennis M, et al. Identifying the patient at risk of acute kidney injury: a predictive scoring system for the development of acute kidney injury in acute medical patients. Nephron Clin Pract 2013;123(3-4):143-150. [doi: 10.1159/000351509] [Medline: 23887252]
36. Hodgson LE, Dimitrov BD, Roderick PJ, Venn R, Forni LG. Predicting AKI in emergency admissions: an external validation study of the acute kidney injury prediction score (APS). BMJ Open 2017 Mar 08;7(3):e013511 [FREE Full text] [doi: 10.1136/bmjopen-2016-013511] [Medline: 28274964]
37. Connell A, Montgomery H, Martin P, Nightingale C, Sadeghi-Alavijeh O, King D, et al. Evaluation of a digitally-enabled care pathway for acute kidney injury management in hospital emergency admissions. NPJ Digit Med 2019;2:67 [FREE Full text] [doi: 10.1038/s41746-019-0100-6] [Medline: 31396561]

Abbreviations

AKI: acute kidney injury
AUC: area under the receiver operating characteristic curve
CA-AKI: Community-acquired acute kidney injury
CCI: Charlson comorbidity index
CGMHs: Chang Gung Memorial Hospitals
eGFR: estimated glomerular filtration rate
EHR: electronic health records
KDIGO: Kidney Disease: Improving Global Outcomes
LASSO: least absolute shrinkage and selection operator
NP: nondeterministic polynomial time
NPV: negative predictive value
PPV: positive predictive value
RAS: renin-angiotensin system
ROC: receiver operating characteristic
SCr: serum creatinine
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

XGBoost: extreme gradient boost

Edited by G Eysenbach; submitted 04.11.19; peer-reviewed by B Mortazavi, A Connell; comments to author 28.04.20; revised version received 12.06.20; accepted 07.07.20; published 04.08.20.

Please cite as:
Hsu CN, Liu CL, Tain YL, Kuo CY, Lin YC
Machine Learning Model for Risk Prediction of Community-Acquired Acute Kidney Injury Hospitalization From Electronic Health Records: Development and Validation Study
J Med Internet Res 2020;22(8):e16903
URL: https://www.jmir.org/2020/8/e16903
doi:10.2196/16903
PMID:32749223

Original Paper

Detection of Bacteremia in Surgical In-Patients Using Recurrent Neural Network Based on Time Series Records: Development and Validation Study

Hyung Jun Park[1], MD; Dae Yon Jung[2], PhD; Wonjun Ji[1], MD; Chang-Min Choi[1], MD, PhD

[1]Department of Pulmonary and Critical Care Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[2]Big Data & AI Lab, Hana Institute of Technology, Hana TI, Seoul, Republic of Korea

Corresponding Author:
Chang-Min Choi, MD, PhD
Department of Pulmonary and Critical Care Medicine
Asan Medical Center
University of Ulsan College of Medicine
Seoul
Republic of Korea
Phone: 82 01097926907
Email: ccm@amc.seoul.kr

Abstract

Background: Detecting bacteremia among surgical in-patients is more obscure than other patients due to the inflammatory condition caused by the surgery. The previous criteria such as systemic inflammatory response syndrome or Sepsis-3 are not available for use in general wards, and thus, many clinicians usually rely on practical senses to diagnose postoperative infection.

Objective: This study aims to evaluate the performance of continuous monitoring with a deep learning model for early detection of bacteremia for surgical in-patients in the general ward and the intensive care unit (ICU).

Methods: In this retrospective cohort study, we included 36,023 consecutive patients who underwent general surgery between October and December 2017 at a tertiary referral hospital in South Korea. The primary outcome was the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC) for detecting bacteremia by the deep learning model, and the secondary outcome was the feature explainability of the model by occlusion analysis.

Results: Out of the 36,023 patients in the data set, 720 cases of bacteremia were included. Our deep learning–based model showed an AUROC of 0.97 (95% CI 0.974-0.981) and an AUPRC of 0.17 (95% CI 0.147-0.203) for detecting bacteremia in surgical in-patients. For predicting bacteremia within the previous 24-hour period, the AUROC and AUPRC values were 0.93 and 0.15, respectively. Occlusion analysis showed that vital signs and laboratory measurements (eg, kidney function test and white blood cell group) were the most important variables for detecting bacteremia.

Conclusions: A deep learning model based on time series electronic health records data had a high detective ability for bacteremia for surgical in-patients in the general ward and the ICU. The model may be able to assist clinicians in evaluating infection among in-patients, ordering blood cultures, and prescribing antibiotics with real-time monitoring.

(J Med Internet Res 2020;22(8):e19512) doi:10.2196/19512

KEYWORDS

deep learning; bacteremia; early detection; time series; recurrent neural network; neural network; informatics; surgery; sepsis; modeling

Introduction

Bacteremia is associated with increased morbidity and mortality [1]. As bacteria grow in the bloodstream, subsequent immune response can cause sepsis, a life-threatening organ dysfunction [2]. Early administration of antibiotics is important for reducing the mortality associated with this infectious condition [3]. As the observable features of infection are symptoms or laboratory data, the systemic inflammatory response syndrome (SIRS) criteria have been used to detect sepsis in bedside medicine [4], even though it harbors issues such as underscoring and inaccuracy [2]. Variations of rule-based scoring systems such as the Modified Early Warning Score (MEWS) [5] and the

Sequential Organ Failure Assessment (SOFA) [6] have also been developed; however, these systems are not commonly used outside the setting of critical care [2].

More recently, machine learning or deep learning–based models that use data from electronic health records (EHRs) were developed for early prediction of sepsis [7-9]. However, labeling the start time of sepsis is a delicate matter. Different scholars have used different criteria to define sepsis, such as the increase of SOFA score [8] or the detection of any two SIRS criteria in in-patients with the International Classification of Diseases, Ninth Revision (ICD-9) codes of sepsis [7,9,10]. The SIRS criteria should only be used when an infection is suspected [11]. This is why the ICD-9 code for sepsis is used with SIRS criteria; however, using the ICD-9 code does not guarantee that the patients are indeed suspected with sepsis, especially if other interventions such as surgery are performed. Another gold standard of defining sepsis is the Sepsis-3 criteria, which refers to the increase of the SOFA score, that is also used in suspected infection. Considering the infectious condition, some studies used a time stamp of antibiotics and blood culture as the suspected time of infection [12,13]. However, the SOFA score is continuously measured only for patients in the intensive care unit (ICU), as in the case of the Glasgow coma scale.

For these reasons, although existing models have shown significant results, they cannot be used in the general ward especially when factors other than infection, such as surgery, may affect the SIRS criteria through inflammation or when the SOFA score cannot be measured. To overcome the limitations of indirect measurements of infection, direct measurements of blood culture could be helpful for defining infection. In the EHR data set, the results of a blood culture are recorded along with the reception time and the isolated species; therefore, clinicians can define the period of bacteremia so that the labeling of bacteremia only represents the infection. Once a clinician identifies the risk for patients, the clinician could investigate the source of infection and prescribe antibiotics.

Recently, the time series models based on the long short-term memory model [14] and the gated recurrent unit [15] have gained popularity due to their end-to-end modeling, ease of incorporating exogenous variables, and ability for extracting features [16]. The models use a time window to characterize the trend of features. Time series data such as vital signs or laboratory measurements could have different features whether the body temperature increases slowly or quickly. Moreover, if the model uses a longer time window, the model could imply a longer trend of data. With the ability to characterize the trend of time series data, previous studies showed significant performance using those models for predicting sepsis or acute kidney injury [8,17].

This paper presents a model based on a recurrent neural network (RNN) that continuously detects and predicts bacteremia for surgical in-patients in the general ward and the ICU. We compared the performance of this model with the traditional models used in sepsis detection (ie, SIRS criteria, SOFA score, and MEWS). To enhance the reader's understanding, the paper also presents figures depicting continuous detection alongside vital signs and laboratory measurements.

Methods

Study Population

We retrospectively included all patients who had undergone general surgeries at Asan Medical Center (Seoul, South Korea) between October 2007 and December 2017. The following surgical procedures, coded by the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), were included: lung lobectomy (32.0-32.4), gastrectomy (43.4-43.9), hepatectomy (50.0-50.4), and pancreaticoduodenectomy (51.5-51.7, 52.7). We excluded patients who did not undergo spirometry within 3 months before the surgery as well as those who received operations other than the previously mentioned types according to the manually written operation records (first e-table in Multimedia Appendix 1). For querying the EHRs, we used the Asan Biomedical Research Environment system at Asan Medical Center [18,19].

The data can be categorized into time-invariant and time-variant data depending on whether the data changed over the admission period. The time-invariant data included demographic data (ie, age, sex, height, weight, body mass index), underlying disease as coded with the ICD-10-CM code, type of operation, disease for operation, amount of transfusion during operation, spirometry (forced expiratory volume in 1 second [FEV_1], forced vital capacity [FVC], FEV_1/FVC, each with raw and predicted values), and the department of surgery. The time-variant data were vital signs (systolic blood pressure [sBP], diastolic blood pressure [dBP], pulse rate, respiratory rate, temperature) and laboratory data (groups as white blood cell [WBC], red blood cell, liver function, electrolyte, kidney function, arterial blood gas analysis, inflammation).

Definition of Bacteremia

Bacteremia was defined as a laboratory-confirmed bloodstream infection that meets at least one of the following criteria [1]. First, the patients must have a recognized pathogen cultured from ≥1 blood specimen. Second, patients must have a fever (>38.0 °C) or hypotension (sBP<90 mmHg) in case of common skin contaminant (eg, diphtheroids, *Bacillus* species, *Propionibacterium* species, coagulase-negative staphylococci, or micrococci) that should be isolated from more than 2 blood cultures.

Ground Truth for Bacteremia Periods

Two time points of bacteremia were present in the EHR data set: time ordered by a clinician and time of reception at the department of laboratory. Often there existed a time discrepancy between a clinician's order and the actual sampling time. Therefore, the reception time (which was a little later than sampling time) was used as the ground truth time point of the bacteremia. In a previous study, bacteremia episodes identified in more than one blood culture within 24 hours were considered as a single episode [20]. In general, within the 24-hour period, vital signs are not readily stabilized even with the appropriate use of antibiotics. Thus, we defined the "bacteremia period" as 24 hours after the time of bacteremia and labeled it as the ground truth.

The prediction target at each point in time prior to the blood culture was a binary variable that was deemed positive if the bacteremia occurred within a predetermined time window. To determine the effect of the length of the time window on the detection ability of the model, we trained three different models (8, 16, and 24 hours prior) for predicting future time points.

Models for Detecting and Predicting Bacteremia

All models function across the entire time period of admission. In the beginning of an admission period, there were not enough data to fill each time window. In these cases, only the existing data were used for predicting bacteremia. For example, to predict bacteremia 24 hours after admission, with a 96 hour (4 days) time window, only data from the first 24 hours of admission were used, and the vacant 72 hours of data, which was prior to the admission, were masked (Keras Masking layer). We tested time-variant variables with different time steps (1, 2, 4, 6, 8, 10, and 12 days) to find the optimal length of time steps that results in superior detection performance.

When constructing the model, different approaches were taken depending on the nature of the variables: time-variant variables were treated with an RNN-based model and time-invariant variables were treated with dense neural networks (first e-Figure in Multimedia Appendix 1). The outputs of both arms were concatenated, and the probability of bacteremia was calculated by a dense neural network. The code used to train and evaluate the model is available on GitHub [21].

We used the area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), sensitivity, specificity, and positive predictive value as comparative measures. The 95% CI was calculated by using a bootstrap approach in which we resampled the data at each time point 1000 times. All analyses were conducted in Python, version 3.7.5 (Python Software Foundation).

Details on data preprocessing, feature embedding, hyperparameter optimizing, ensemble of each batch for overcoming imbalance of data, and model testing are described in the method section of Multimedia Appendix 1.

Significance of Features Analysis

The performance of the model changes when some variables are masked. This method is known as occlusion analysis, which is frequently applied in the field of image analysis [17,22]. We investigated the relative importance of the variables in our trained models through the occlusion analysis, in which variable groups were occluded one by one to determine their respective influence on the prediction of bacteremia. For example, if the occluded variables held higher importance in bacteremia detection, the resulting model would have a lower performance. Each group of variables was independently embedded with separate autoencoders, and the groups are described in the second e-Table in Multimedia Appendix 1.

Results

Baseline Characteristics

A total of 56,339 patients and 58,223 admission records were included when applying ICD-10-CM codes and spirometric results. After applying the exclusion criteria, 35,256 patients and 36,023 admission records were left in the final data set. The baseline demographic characteristics, operation types, and hours of operation are described in the third e-Table in Multimedia Appendix 1. The incidence of bacteremia was 1.9% (720/36,023 cases) and the period of bacteremia was 0.22% (1006/1,362,865 person-time). Furthermore, the annual incidence of bacteremia was rather stable during the study period, ranging from 1.6% (76/4484 cases) to 2.3% (45/1882 cases; second e-Figure in Multimedia Appendix 1).

The median time of bacteremia was 9 (IQR 2-18) days since admission, and the median time of postoperative bacteremia was 9 (IQR 6-15) days after surgery. If bacteremia occurred prior to surgery, patients underwent surgeries at a median of 10 (IQR 5-18) days after the occurrence of bacteremia (third e-Figure in Multimedia Appendix 1).

Predicted Probability in Laboratory Sheets

Figure 1 depicts the predicted probability of bacteremia along with the vital signs and the laboratory data during hospital admission. As shown in Figure 1, the probability of bacteremia notably increased when the pulse rate and body temperature were elevated. In contrast, laboratory data such as WBC, hemoglobin, platelet, and c-reactive protein (CRP) did not show such changes in accordance with the increase in bacteremia probability (Figure 1). More examples of good, bad, and obscure prediction results are presented in the fourth e-Figure in Multimedia Appendix 1.

Figure 2 shows the actual time of negative blood cultures as green bars. Even though the model was only trained based on the red bars (ie, bacteremia periods), the green bars indicate high probabilities for bacteremia. As the results of a blood culture could be false negative, the green bar might represent bacteremia at which clinicians should inspect patients and prescribe antibiotics. Other examples related to this figure are shown in the fourth e-Figure in Multimedia Appendix 1.

Our model had an AUROC of 0.978 (95% CI 0.974-0.981) and an AUPRC of 0.17 (95% CI 0.147-0.203; Figure 3). The AUROC of previous models are as follows: SIRS 0.778 (95% CI 0.768-0.786), SOFA 0.738 (95% CI 0.728-0.748), and MEWS 0.673 (95% CI 0.662-0.682). The AUPRC of previous models are as follows: SIRS 0.011 (95% CI 0.010-0.013), SOFA 0.010 (95% CI 0.009-0.011), and MEWS 0.010 (95% CI 0.008-0.011).

Figure 1. Patterns of the probability of bacteremia along with vital signs and laboratory data. Data from a 76-year-old female patient admitted for pancreatic cancer who underwent pylorus-preserving pancreaticoduodenectomy on hospital day 8. The graph in the bottom shows the probability of bacteremia at each time step. Red bars represent the actual period of bacteremia during which bacteria was isolated in the blood culture. The name of the pathogen is written in a small box. On hospital day 21, fever was noted and the probability of bacteremia was elevated. Lab data did not show a notable correlation with bacteremia probabilities. CRP: c-reactive protein; WBC: white blood cell.

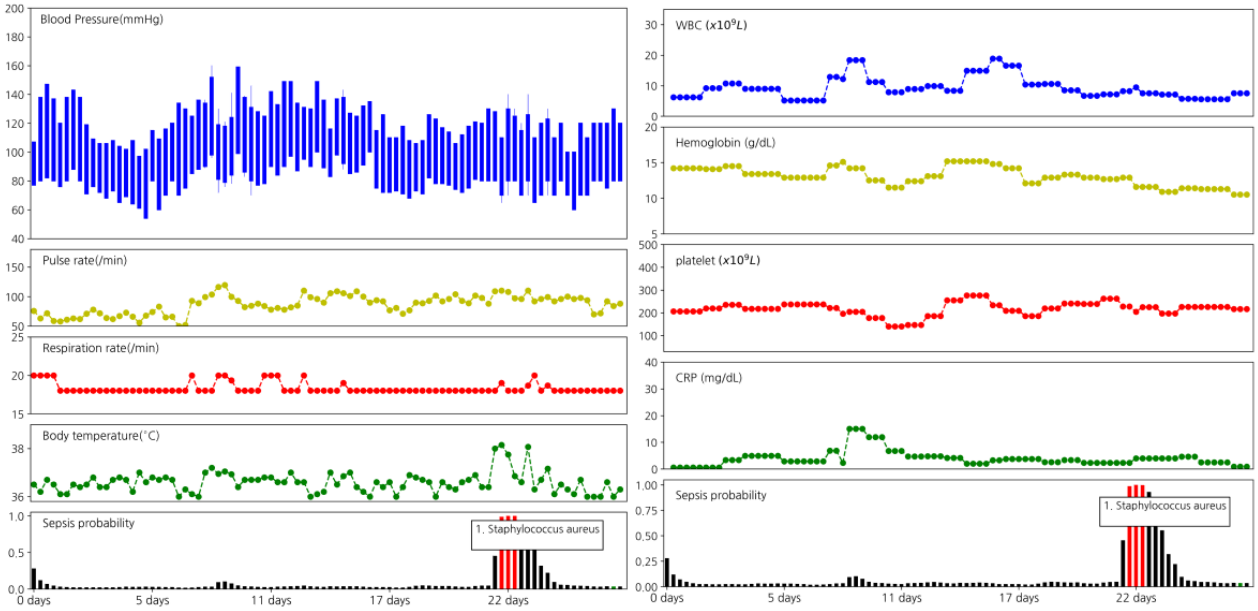


Figure 2. Time of negative blood culture could represent high likelihood of bacteremia. Data from a 77-year-old male patient admitted for intrahepatic duct stone. The lobectomy of the liver was carried out on hospital day 3. On hospital day 15, high fever was noted, and the blood culture was performed; however, no bacterial species were isolated. On hospital day 20, the second high fever was identified, and the blood culture was performed again. Candida Albicans was isolated, and the vital sign was subsequently stabilized. The green bar is the blood culture with no isolation. CRP: c-reactive protein; WBC: white blood cell.

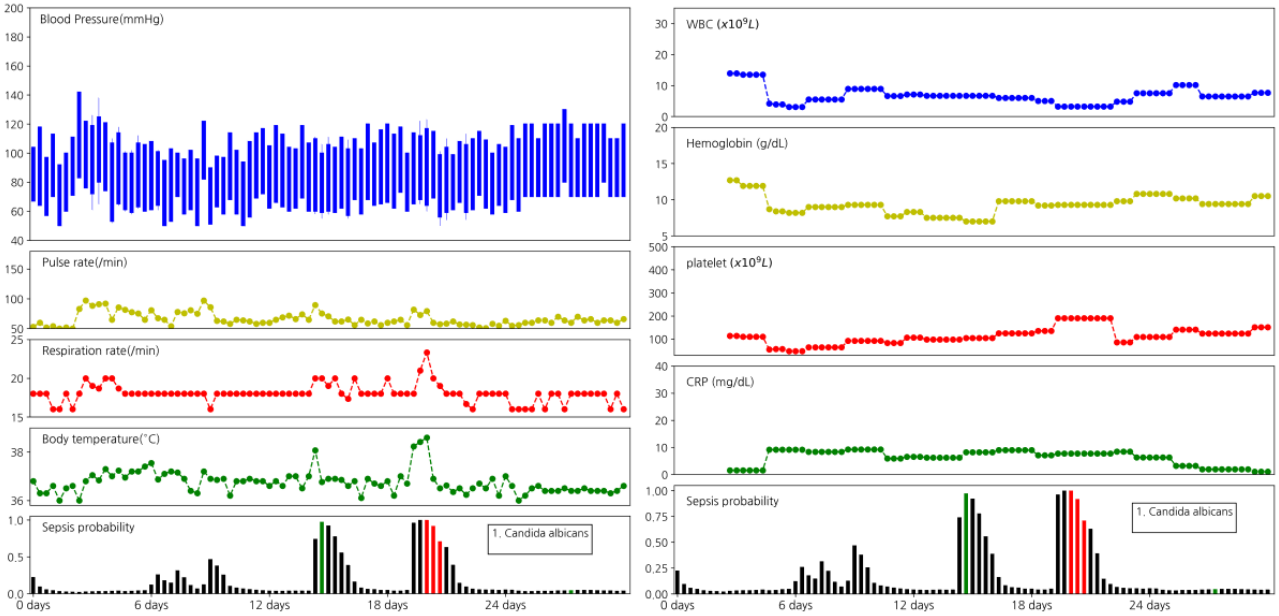


Figure 3. Receiver operating characteristics and precision-recall curve of the proposed model. The AUROC of the model was 97%, and the area under the precision-recall curve was 17%, which were higher compared with those of previous models. Each circle of previous criteria is the metric of the cut-off value of the models. AUROC: area under the receiver operating characteristic curve; MEWS: Modified Early Warning Score; SIRS: systemic inflammatory response syndrome; SOFA: Sequential Organ Failure Assessment.

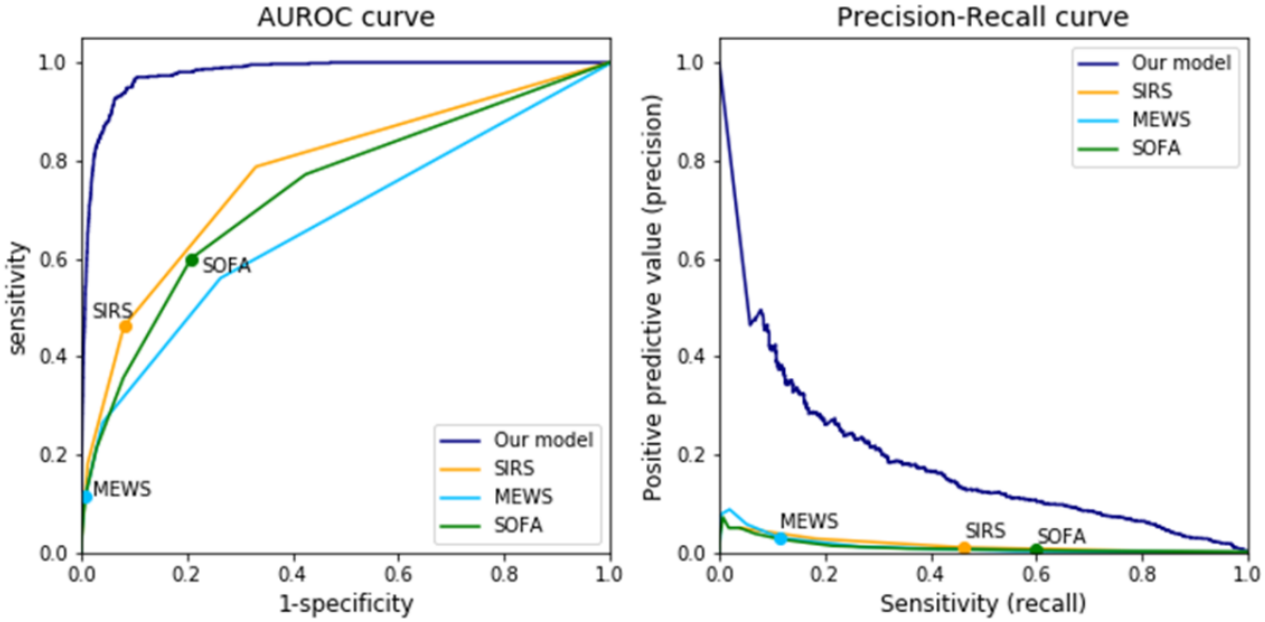


Table 1 shows the trend of sensitivity and specificity of the proposed model according to different thresholds, along with the performances of other models (ie, SIRS, SOFA, and MEWS). The positive predicted value was relatively low because of the low incidence of bacteremia in our data set. Overall, our model showed superior performance to the SIRS criteria [23], SOFA score [6], and MEWS score [24] in terms of both AUROC and AUPRC for detecting bacteremia.

Table 1. Performance of the model compared with previous criteria.

Model and threshold	Sensitivity	Specificity	PPV[a]
Recurrent neural network model			
0.1	0.94	0.92	0.023
0.2	0.88	0.95	0.034
0.3	0.86	0.96	0.044
0.4	0.83	0.97	0.054
0.5	0.79	0.98	0.065
0.6	0.72	0.98	0.079
0.7	0.65	0.99	0.099
0.8	0.53	0.99	0.122
0.9	0.41	0.99	0.165
SIRS[b] criteria (>2 score)	0.46	0.92	0.011
SOFA[c] score (>2 increase from baseline)	0.60	0.79	0.006
MEWS[d] score (>4 score)	0.12	0.99	0.031

[a]PPV: positive predictive value.

[b]SIRS: systemic inflammatory response syndrome.

[c]SOFA: Sequential Organ Failure Assessment.

[d]MEWS: Modified Early Warning Score.

Various Experiments for Bacteremia Prediction

We developed models predicting 8, 16, and 24 hours before the event of bacteremia (Table 2). Both the AUROC and AUPRC values decreased as the time to predict bacteremia increased from 8 to 24 hours (AUROC 0.963 to 0.929; AUPRC 0.176 to 0.154).

When we made the model using various time step lengths (1, 2, 4, 6, 8, 10, and 12 days), the AUROC values did not show

notable decreases (0.977 to 0.971) as the time windows increased from 1 to 12 days; in contrast, the AUPRC value increased from 0.139 to 0.174 (Table 2). Performance for predicting bacteremia events 8, 16, and 24 hours prior to the event are shown. Different time steps and performances in our data set are shown. The time steps indicate the length of time period used in the RNN model to predict bacteremia.

Table 2. Model performance for predicting bacteremia according to forecasting time to event and time steps of the recurrent neural network model.

Variables	AUROC[a]	AUPRC[b]
Prior time to event (hour)		
0 (at event)	0.98	0.17
8 prior	0.96	0.18
16 prior	0.95	0.17
24 prior	0.93	0.15
Time steps (days)		
1	0.98	0.14
2	0.98	0.15
4	0.98	0.15
6	0.98	0.15
8	0.97	0.16
10	0.97	0.174
12	0.98	0.165

[a]AUROC: area under the receiver operating characteristic curve.

[b]AUPRC: area under the precision-recall curve.

Relative Importance of Variables

Table 3 shows the results of the occlusion analysis. We found that occlusion of vital signs resulted in the largest decrease in both the AUROC and AUPRC values, followed by kidney-related values and WBC. In contrast, the occlusion of time-invariant data, which were reported to be important in predicting postoperative complications [25-27], showed little effects in decreasing AUROC or AUPRC. For comparison with previous studies that only used time-invariant data [25-27], we also trained a model only with time-invariant data; as a result, we observed that this model had a similar performance with those in previous studies (AUROC 0.84, AUPRC 0.15; fourth e-Table in Multimedia Appendix 1). Vital signs and lab data were more important than time-invariant data, even though the latter could also somewhat assess the risk of bacteremia. Particularly, body temperature was the most important vital sign in detecting bacteremia, followed by dBP and pulse rate (fifth e-Table in Multimedia Appendix 1).

The model performance was described as when the important variables were occluded.

Table 3. Detecting performance of the proposed model in occlusion analysis.

Methods	AUROC[a]	AUPRC[b]
Original model	0.98	0.17
Occluding method		
Occluding vital sign	0.85	0.05
Occluding kidney-related values	0.95	0.06
Occluding WBC[c]	0.96	0.07
Occluding electrolyte	0.96	0.10
Occluding RBC[d]-related lab	0.97	0.11
Occluding ABGA[e]	0.97	0.12
Occluding inflammatory markers	0.98	0.14
Occluding time-invariant data	0.97	0.14
Occluding liver function test	0.98	0.15

[a]AUROC: area under the receiver operating characteristic curve.

[b]AUPRC: area under the precision-recall curve.

[c]WBC: white blood cell.

[d]RBC: red blood cell.

[e]ABGA: arterial blood gas analysis.

Discussion

Summary of the Principal Finding

By using the deep learning method, we devised a model that generates a real-time probability of detecting and predicting bacteremia. The proposed model had an AUROC of 0.978 in detecting bacteremia every 8 hours, which is a notably superior performance compared with other existing criteria [2,5,6,24]. In predicting bacteremia 24 hours in advance, the model showed a relatively lower performance (AUROC 0.929) than the detecting model. Occlusion analysis showed that vital signs were the most important variables in bacteremia detection. Confirming our expectation that patterns of time-variant variables such as vital signs could be used to characterize the risk type of a patient, the model using long time steps showed more accurate results (time steps: 10 days vs 1 days, AUPRC: 0.174 vs 0.139).

This study is also important for the continuous monitoring of bacteremia so that clinicians can get advice on the risk of uncontrolled infection. Other studies on predicting sepsis rely on assessments about the clinical state (based on the SOFA or SIRS criteria) [8,9,20]. However, if the outcome label included systemic inflammations as well as infections, the probability results would be difficult for interpretation by clinicians [2]. To overcome the unclear labeling problem, we used the direct results from the blood culture and suggested the “bacteremia period,” which can be used as an indication of infection. Since the model derives the results solely from the blood cultures, the resulting predicted probabilities directly indicate the infection so that physicians can get advice when they search for the source of infection, start new antibiotics, or monitor the appropriate response to antibiotics.

Importance of Variables in the Deep Learning Model

Deep neural networks are often questioned for being nontransparent and because the basis of the prediction results is hard to explain [28]. By depicting the probability with vital signs and major laboratory findings, we were able to determine whether the prediction results were proper (fourth e-Figure in Multimedia Appendix 1). Specifically, we observed that the bacteremia probability was elevated in accordance with increases in body temperature, respiratory rate, and pulse rate, whereas laboratory data such as WBC and CRP did not show notable correlations with the bacteremia probability.

To explain what variables drive the model, we used occlusion analysis, a method used in image analyses [28]. If the model is driven by an important location within the image, the result must not be changed after occluding the surrounding of the image [22]. We found that the most important variable of our model was the vital sign, followed by kidney-related values and WBC. The result resembles the SIRS criteria, which consists of three vital sign categories [2], and supports our expectation that our model would focus on relevant variables for predicting bacteremia. Underlying diseases are also known as important predictors of postoperative complications [25-27]. However, time-invariant variables such as the underlying diseases and the type of surgery did not have significant effects in the detection of bacteremia. In clinical practice, patients with a high likelihood of postoperative complications are not always suspected of having an infection, unless they show features of infection such as fever. As previous risk factors suggest that only high-risk patients acquired infection during the whole admission period, the detection of bacteremia should be based on clinical clues such as vital signs or laboratory data. This is in line with the routine practice of clinicians suspecting infection based on vital signs and clinical features rather than underlying disease.

Validation of Predicted Probability Compared With Medical Chart

Blood cultures are usually performed for diagnosing infections or monitoring the bacteremia, and some negative blood cultures could be false negatives [1]. Investigating negative blood cultures (green bars in the fourth e-Figure in Multimedia Appendix 1), the predicted probability of the green bar is high among negative cases (Figure 2). However, a green bar following 2 days after a positive blood culture showed a lower probability of bacteremia (4.2.2 e-Figure in Multimedia Appendix 1). Although the negative blood culture was not trained in our model, the probability at the green bar represents the time of the blood culture following a clinician's suspicion. Usually, the probability for bacteremia is high at the green bar, meaning that both clinicians and our model suspected high risk of infection at similar time points. In other words, our model is not just trained for the labeling but also trained against general features of bacteremia.

Additionally, we examined the performance of our model using different time steps. We assumed that additional information exists when there are spikes in body temperature or lab data, or when the data exhibits recognizable patterns throughout the time. Using the RNN-based model, such information can be considered in the hidden state and be used to predict bacteremia. When we increased the time steps per bacteremia prediction, the performance of the model in bacteremia detection was increased, indicating that the model was able to further learn the patterns of vital signs and lab data. For example, when the changes in value were not steep, the predicted probability was relatively low despite the elevated pulse rate and body temperature (4.1.3 e-Figure in Multimedia Appendix 1). These patterns of vital signs and lab data could also be used for differentiating different species of bacteremia if they have distinct disease patterns.

Limitations

Our study has the following limitations. First, we used the outcome defined by positive blood culture. In the Prehospital Antibiotics Against Sepsis Trial, only 42.6% of the cases were culture-positive sepsis [29]. Therefore, this labeling may have affected our model and the results. However, because we assumed that the vital signs and laboratory results are similar between culture-positive sepsis and culture-negative sepsis, the model could predict higher probability even when the blood culture produced negative results. Further prospective study is needed to validate the proper prediction about culture-negative sepsis. Second, our model was trained on data from a single tertiary hospital in Korea and may, thus, have limited generalizability. Nevertheless, our data set does not seem to significantly deviate from the country-wide data, as the incidence rates of bacteremia in our data set and the general Korean cohort data were 1.9% and 2.2%, respectively [30]. As our data set included all consecutive patients who underwent surgeries, it could represent the global population of surgical patients undergoing major upper abdominal surgery and thoracic surgery. In addition, as our model used the data set in a retrospective manner, a prospective study is needed to determine whether our proposed model confers real-time values in helping clinicians predict bacteremia at an earlier stage.

Conclusions

In conclusion, we have applied the deep learning algorithm to develop a model for detecting and predicting bacteremia with in-hospital data. Our model may help clinicians to make appropriate decisions regarding early responses to bacteremia. In the future, clinicians may be able to improve the clinical outcomes of patients with bacteremia using this algorithm in the EHR system.

Acknowledgments

We thank Seung Hoon Chae for editing for English grammar.

Authors' Contributions

The conceptualization was done by HP, DYJ, and CMC. The methodology was designed by HP. The formal analysis and investigation was done by HP and DYJ. The software was used by HP and DYJ. Writing of the original draft preparation was done by HP. Review and editing was done by HP, DYJ, WJ, and CMC. The study was supervised by WJ and CMC.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 1436 KB - jmir_v22i8e19512_app1.docx]

References

1. Seifert H. The clinical importance of microbiological findings in the diagnosis and management of bloodstream infections. Clin Infect Dis 2009 May 15;48 Suppl 4:S238-S245. [doi: 10.1086/598188] [Medline: 19374579]

2. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 2016 Mar 23;315(8):801-810 [FREE Full text] [doi: 10.1001/jama.2016.0287] [Medline: 26903338]
3. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, Surviving Sepsis Campaign Guidelines Committee including the Pediatric Subgroup. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. Crit Care Med 2013 Mar;41(2):580-637. [doi: 10.1097/CCM.0b013e31827e83af] [Medline: 23353941]
4. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Crit Care Med 1992 Jun;20(6):864-874. [Medline: 1597042]
5. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. QJM 2001 Oct;94(10):521-526. [doi: 10.1093/qjmed/94.10.521] [Medline: 11588210]
6. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996 Jul;22(7):707-710. [doi: 10.1007/BF01709751] [Medline: 8844239]
7. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med 2015 Aug 05;7(299):299ra122. [doi: 10.1126/scitranslmed.aab3719] [Medline: 26246167]
8. Liu R, Greenstein JL, Granite SJ, Fackler JC, Bembea MM, Sarma SV, et al. Data-driven discovery of a novel sepsis pre-shock state predicts impending septic shock in the ICU. Sci Rep 2019 Apr 16;9(1):6145. [doi: 10.1038/s41598-019-42637-5] [Medline: 30992534]
9. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open 2018 Jan 26;8(1):e017833 [FREE Full text] [doi: 10.1136/bmjopen-2017-017833] [Medline: 29374661]
10. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. Comput Biol Med 2017 Oct 01;89:248-255. [doi: 10.1016/j.compbiomed.2017.08.015] [Medline: 28843829]
11. Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. In: Chest. 1992 Jun Presented at: The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine; 1992; Chicago p. 1644-1655. [doi: 10.1378/chest.101.6.1644]
12. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR Med Inform 2016 Sep 30;4(3):e28 [FREE Full text] [doi: 10.2196/medinform.5909] [Medline: 27694098]
13. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med 2018 Apr;46(4):547-553 [FREE Full text] [doi: 10.1097/CCM.0000000000002936] [Medline: 29286945]
14. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
15. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv 2014;1:9.
16. Assaad M, Boné R, Cardot H. A new boosting algorithm for improved time-series forecasting with recurrent neural networks. Inf Fusion 2008 Jan;9(1):41-55. [doi: 10.1016/j.inffus.2006.10.009]
17. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 2019 Aug;572(7767):116-119 [FREE Full text] [doi: 10.1038/s41586-019-1390-1] [Medline: 31367026]
18. Shin S, Lyu Y, Shin Y, Choi HJ, Park J, Kim W, et al. Lessons learned from development of de-identification system for biomedical research in a Korean tertiary hospital. Healthc Inform Res 2013 Jun;19(2):102-109 [FREE Full text] [doi: 10.4258/hir.2013.19.2.102] [Medline: 23882415]
19. Shin S, Park YR, Shin Y, Choi HJ, Park J, Lyu Y, et al. A De-identification method for bilingual clinical texts of various note types. J Korean Med Sci 2015 Jan;30(1):7-15 [FREE Full text] [doi: 10.3346/jkms.2015.30.1.7] [Medline: 25552878]
20. Lee KH, Dong JJ, Jeong SJ, Chae M, Lee BS, Kim HJ, et al. Early detection of bacteraemia using ten clinical variables with an artificial neural network approach. J Clin Med 2019 Oct 02;8(10):1592 [FREE Full text] [doi: 10.3390/jcm8101592] [Medline: 31581716]
21. Park HJ. Early detection of of bacteremia in general ward. GitHub. URL: https://github.com/podkd87/bacteremia [accessed 2020-07-27]
22. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. arXiv. 2013 Nov 12. URL: https://arxiv.org/abs/1311.2901 [accessed 2020-07-21]

23. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, SCCM/ESICM/ACCP/ATS/SIS. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Crit Care Med 2003 Apr;31(4):1250-1256. [doi: 10.1097/01.CCM.0000050454.01978.3B] [Medline: 12682500]
24. Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The value of Modified Early Warning Score (MEWS) in surgical in-patients: a prospective observational study. Ann R Coll Surg Engl 2006 Oct;88(6):571-575. [doi: 10.1308/003588406x130615]
25. Kocabas A, Kara K, Özgür G, Sönmez H, Burgut R. Value of preoperative spirometry to predict postoperative pulmonary complications. Respir Med 1996 Jan;90(1):25-33. [doi: 10.1016/s0954-6111(96)90241-3]
26. Kim TH, Lee JS, Lee SW, Oh YM. Pulmonary complications after abdominal surgery in patients with mild-to-moderate chronic obstructive pulmonary disease. COPD 2016 Nov;Volume 11:2785-2796. [doi: 10.2147/copd.s119372]
27. Nijbroek SG, Schultz MJ, Hemmes SN. Prediction of postoperative pulmonary complications. Curr Opinion Anaesthesiology 2019;32(3):443-451. [doi: 10.1097/aco.0000000000000730]
28. Buhrmester V, Münch D, Arens M. Analysis of explainers of black box deep neural networks for computer vision: a survey. arXiv 2019:22.
29. Alam N, Oskam E, Stassen PM, Exter PV, van de Ven PM, Haak HR, PHANTASi Trial Investigatorsthe ORCA (Onderzoeks Consortium Acute Geneeskunde) Research Consortium the Netherlands. Prehospital antibiotics in the ambulance for sepsis: a multicentre, open label, randomised trial. Lancet Respir Med 2018 Jan;6(1):40-50. [doi: 10.1016/S2213-2600(17)30469-1] [Medline: 29196046]
30. Kim J, Kim K, Lee H, Ahn S. Epidemiology of sepsis in Korea: a population-based study of incidence, mortality, cost and risk factors for death in sepsis. Clin Exp Emerg Med 2019 Mar;6(1):49-63. [doi: 10.15441/ceem.18.007] [Medline: 30781941]

Abbreviations

AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
CRP: c-reactive protein
dBP: diastolic blood pressure
EHR: electronic health record
FEV1: forced expiratory volume in 1 second
FVC: forced vital capacity
ICD-9: International Classification of Diseases, Ninth Revision
ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification
ICU: intensive care unit
MEWS: Modified Early Warning Score
RNN: recurrent neural network
sBP: systolic blood pressure
SIRS: systemic inflammatory response syndrome
SOFA: Sequential Organ Failure Assessment
WBC: white blood cell

Edited by G Eysenbach; submitted 22.04.20; peer-reviewed by S Azimi, CT Cheng, AS Poncette; comments to author 14.05.20; revised version received 23.06.20; accepted 14.07.20; published 04.08.20.

Please cite as:
Park HJ, Jung DY, Ji W, Choi CM
Detection of Bacteremia in Surgical In-Patients Using Recurrent Neural Network Based on Time Series Records: Development and Validation Study
J Med Internet Res 2020;22(8):e19512
URL: https://www.jmir.org/2020/8/e19512
doi:10.2196/19512
PMID:32669261

Original Paper

Reproducible Machine Learning Methods for Lung Cancer Detection Using Computed Tomography Images: Algorithm Development and Validation

Kun-Hsing Yu[1,2,3], MD, PhD; Tsung-Lu Michael Lee[4], PhD; Ming-Hsuan Yen[5,6], BSc; S C Kou[2], PhD; Bruce Rosen[7,8], MD, PhD; Jung-Hsien Chiang[6], PhD; Isaac S Kohane[1,8], MD, PhD

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States
[2]Department of Statistics, Harvard University, Cambridge, MA, United States
[3]Department of Pathology, Brigham and Women's Hospital, Boston, MA, United States
[4]Department of Information Engineering, Kun Shan University, Tainan, Taiwan
[5]Graduate Program of Multimedia Systems and Intelligent Computing, National Cheng Kung University and Academia Sinica, Tainan, Taiwan
[6]Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan
[7]Department of Radiology, Athinoula A Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA, United States
[8]Division of Health Sciences and Technology, Harvard–Massachusetts Institute of Technology, Boston, MA, United States

Corresponding Author:
Jung-Hsien Chiang, PhD
Department of Computer Science and Information Engineering
National Cheng Kung University
No 1 University Road
Tainan
Taiwan
Phone: 886 6 2757575 ext 62534
Email: jchiang@mail.ncku.edu.tw

Abstract

Background: Chest computed tomography (CT) is crucial for the detection of lung cancer, and many automated CT evaluation methods have been proposed. Due to the divergent software dependencies of the reported approaches, the developed methods are rarely compared or reproduced.

Objective: The goal of the research was to generate reproducible machine learning modules for lung cancer detection and compare the approaches and performances of the award-winning algorithms developed in the Kaggle Data Science Bowl.

Methods: We obtained the source codes of all award-winning solutions of the Kaggle Data Science Bowl Challenge, where participants developed automated CT evaluation methods to detect lung cancer (training set n=1397, public test set n=198, final test set n=506). The performance of the algorithms was evaluated by the log-loss function, and the Spearman correlation coefficient of the performance in the public and final test sets was computed.

Results: Most solutions implemented distinct image preprocessing, segmentation, and classification modules. Variants of U-Net, VGGNet, and residual net were commonly used in nodule segmentation, and transfer learning was used in most of the classification algorithms. Substantial performance variations in the public and final test sets were observed (Spearman correlation coefficient = .39 among the top 10 teams). To ensure the reproducibility of results, we generated a Docker container for each of the top solutions.

Conclusions: We compared the award-winning algorithms for lung cancer detection and generated reproducible Docker images for the top solutions. Although convolutional neural networks achieved decent accuracy, there is plenty of room for improvement regarding model generalizability.

(J Med Internet Res 2020;22(8):e16709) doi:10.2196/16709

KEYWORDS

computed tomography, spiral; lung cancer; machine learning; early detection of cancer; reproducibility of results

Introduction

Lung cancer is one of the most prevalent cancers worldwide, causing 1.76 million deaths per year [1,2]. Chest computed tomography (CT) scans play an essential role in the screening for [3] and diagnosis of lung cancer [4]. A randomized controlled trial demonstrated that low-dose CT screening reduced mortality from lung cancer among high-risk patients [3], and recent studies showed the benefit of CT screening in community settings [5]. The wide adoption of lung cancer screening is expected to benefit millions of people [6]. However, millions of CT scan images obtained from patients constitute a heavy workload for radiologists [7]. In addition, interrater disagreement has been documented [8]. Previous studies suggested that computer-aided diagnostic systems could improve the detection of pulmonary nodules in CT examination [9-12]. To stimulate the development of machine learning models for automated CT diagnosis, the Kaggle Data Science Bowl provided labeled chest CT images from 1397 patients and awarded $1 million in prizes to the best algorithms for automated lung cancer diagnosis, which is the largest machine learning challenge on medical imaging to date. In response, 1972 teams worldwide have participated and 394 teams have completed all phases of the competition [13], making it the largest health care–related Kaggle contest [14]. This provides a unique opportunity to study the robustness of medical machine learning models and compare the performance of various strategies for processing and classifying chest CT images at scale.

Due to the improved performance of machine learning algorithms for radiology diagnosis, some developers have sought commercialization of their models. However, given the divergent software platforms, packages, and patches employed by different teams, their results were not easily reproducible. The difficulty in reusing the state-of-the-art models and reproducing the diagnostic performance markedly hindered further validation and applications.

To address this gap, we reimplemented, examined, and systematically compared the algorithms and software codes developed by the best-performing teams of the Kaggle Data Science Bowl. Specifically, we investigated all modules developed by the 10 award-winning teams, including their image preprocessing, segmentation, and classification algorithms. To ensure the reproducibility of results and the reusability of the developed modules, we generated a Docker image for each solution using the Docker Community Edition, a popular open-source software development platform that allows users to create self-contained systems with the desired version of software packages, patches, and environmental settings. According to Docker, there are over 6 million Dockerized applications, with 130 billion total downloads [15]. The Docker images are easily transferrable from one server to another, which ensures the reproducibility of scientific computing [16]. Our Dockerized modules will facilitate further development of computer-aided diagnostic algorithms for chest CT images and contribute to precision oncology.

Methods

Data and Classification Models

We obtained the low-dose chest CT datasets in Digital Imaging and Communications in Medicine format from the Kaggle Data Science Bowl website [13]. The dataset was acquired from patients with high risks for developing lung cancers. In this Kaggle challenge, a training set (n=1397) with ground truth labels (362 with lung cancer; 1035 without) and a public test set (n=198) without labels were provided to the participants. The ground truth label is 1 if the patient developed lung cancer within 1 year of the date the CT scan was performed and 0 otherwise. The diagnosis was confirmed by pathology evaluation as a part of the National Lung Screening Trial [3,17]. Once participants submitted the prediction results for the public test set, the Kaggle competition platform reported their models' performance on the public leaderboard instantaneously. The final test set (n=506, ground truth labels were not disclosed to participants) was only available to participants after the model submission deadline, thus serving as an independent validation set that decided the final winners. The chest CT images in the training set, public test set, and final test set all came from multiple hospitals and had different qualities. In particular, the final test set contained more recent and higher quality data with thinner slice thickness than those in the two other sets [18].

To systematically compare the solutions developed by the award-winning teams, we acquired the source codes of the winning solutions and their documentation from the Kaggle news release after the conclusion of the competition. Per the rules of this Kaggle challenge, the source codes of these award-winning solutions were required to be released under open-source licenses approved by the Open Source Initiative [19] in order to facilitate free distribution and derivation of the solution codes [20]. The default license is the MIT license [20]. Under the open-source licenses approved by Open Source Initiative, the software can be freely used, modified, and shared [19].

Comparison of the Approaches and Their Performance

We compared the workflows of the top 10 solutions by examining and rerunning their source codes. For each solution, we inspected all steps taken from inputting the CT images to outputting the prediction. We documented the versions of the software package and platform dependencies of each solution.

The Kaggle Data Science Bowl used the log-loss function to evaluate the performance of the models [13]. The log-loss function

where n is the number of patients in the test set, y_i is 1 if patient i has lung cancer, 0 otherwise, and $\hat{y}_i$ is the predicted probability that patient i has lung cancer [13]. If the predicted outcome is set as 0.5 for all patients, the log-loss value would be ln (2) ≈ 0.69.

To investigate whether models with high performance in the public test set generalize to the images in the final test set, we computed the Spearman correlation coefficient of the log-loss

in the two test sets. All analyses were conducted using R version 3.6 (R Foundation for Statistical Computing).

Docker Image Generation for the Top Ten Solutions

We reproduced the results by recompiling the source codes and dependencies of each of the top ten solutions. Since the solutions used various platforms and different versions of custom software packages, many of which were not compatible with the most updated packages or mainstream release, we generated Docker images [16] to manage the software dependencies and patches required by each solution to enhance the reusability and reproducibility of the developed algorithms.

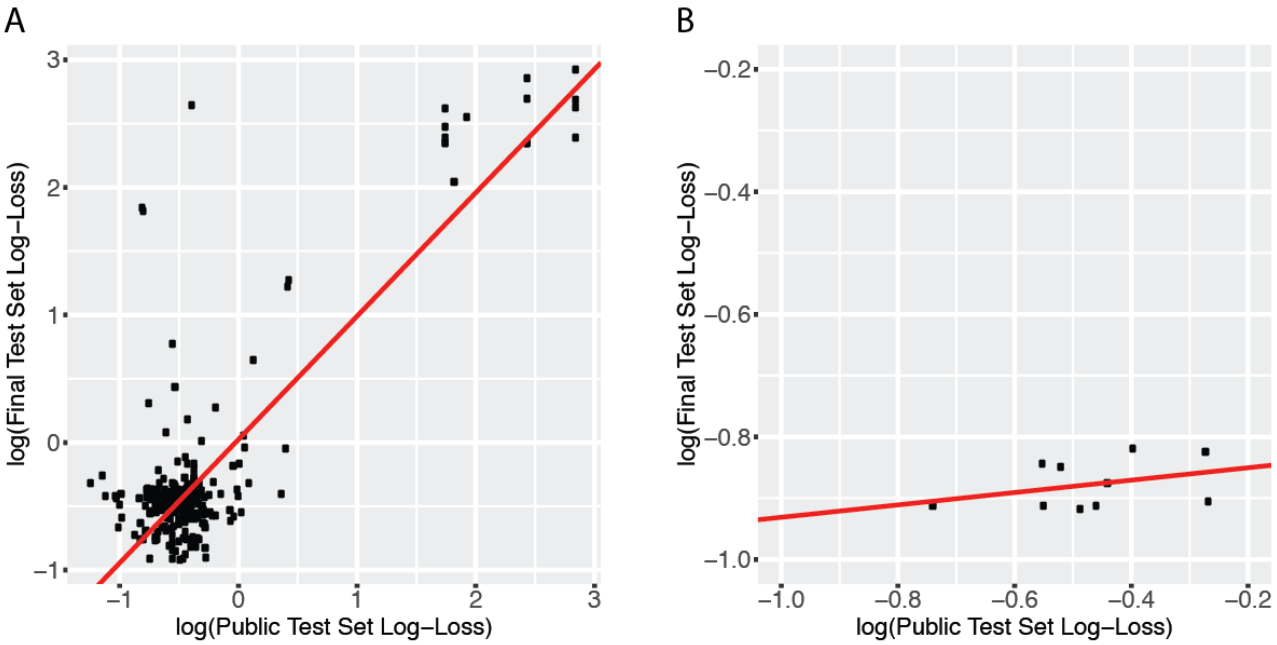
Results

Performance Comparison

Figure 1 summarizes the public and private test set scores of the top 250 teams that participated in the Kaggle Data Science Bowl. Results showed that the top 20 teams achieved a log-loss less than 0.5 in the final test set, and more than 80 submissions reached a log-loss less than 0.6 in the same set. However, these models had varying performances in the public test set. Surprisingly, 11 out of the top 50 teams had a public test set log-loss greater than 0.69, which was worse than blindly submitting "0.5" as the cancer probability for every patient. The correlation between the public test set scores and the final test set scores was weak among all teams that completed the contest (Spearman correlation coefficient = .23; Figure 2A). In the top 10 teams, the correlation is moderate (Spearman correlation coefficient = .39; Figure 2B).

Figure 1. The log-loss score distribution of the top 250 teams in the Kaggle Data Science Bowl Competition. The log-loss scores of the public test set and the final test set of each team were plotted. The red horizontal line indicates the log-loss of outputting the cancer probability as 0.5 for each patient. The blue horizontal line shows the log-loss of outputting cancer probability of each patient as the prevalence of cancer (0.26) in the training set.

Figure 2. A weak to moderate correlation between the log-loss scores of the public test set and the scores of the final test set. The red regression line shows the relation between the log-loss scores of the public test set and those of the final test set using a linear regression model. (A) The log-transformed scores of all participants who finished both stages of the Kaggle Data Science Bowl Competition were plotted. The Spearman correlation coefficient of the performance in the two test sets is .23. (B) The log-transformed scores of the top 10 teams defined by the final test set performance. The Spearman correlation coefficient among the top 10 teams is .39.



Data Workflow Comparison

Figure 3 summarizes the most frequently used strategy by the winning teams. Most solutions used additional publicly available datasets, generated lung segmentation, rescaled the voxels, and performed nodule segmentations before fitting the classification models. Table 1 compares the additional datasets, data preprocessing, segmentation, classification, implementation, and final test set scores of the top 10 solutions.

In addition to the training dataset provided by the Kaggle challenge, most teams used CT images and nodule annotations from other publicly available resources. Table 2 summarizes the sample size, availability of nodule locations, nodule segmentation, diagnoses, other characteristics of the Kaggle dataset, and additional datasets employed by the participants. Most of the top solutions used images and nodule segmentations from the Lung Nodule Analysis 2016 (LUNA16) challenge to develop their segmentation algorithms. LUNA16 is a closely related competition organized in 2016 with an aim to detect lung nodules in chest CT images [21,22]. Two teams also reported using the lung CT images, diagnostic annotations, and nodule location data from the International Society for Optics and Photonics (SPIE)–American Association of Physicists in Medicine (AAPM) Lung CT Challenge [23], but one of them did not incorporate this relatively small dataset (n=70) when building the final models. Only one of the top 10 teams did not use any additional datasets outside of the competition.

Frequently used image preprocessing steps include lung segmentation and voxel scaling. Voxel scaling ensures that the voxels of images from various CT scan protocols correspond to similar sizes of physical space. Variants of U-Net [24], VGGNet [25], and residual net (ResNet) [26] were commonly used as the nodule segmentation algorithms, and the nodule segmentation models trained on the LUNA16 dataset were often applied to the Data Science Bowl dataset.

After lung nodule segmentation, classification algorithms were employed to generate final cancer versus noncancer predictions. Most of the solutions leveraged existing ImageNet-based architecture and transfer learning [12,27]. All teams employed 2D or 3D convolutional neural networks (CNN). A few teams employed CNNs as feature extractors and used tree-based classifiers for this classification task.

Figure 3. A model of the informatics workflow used by most teams. In addition to the Kaggle training set, most teams obtained additional publicly available datasets with annotations. Lung segmentation, image rescaling, and nodule segmentation modules were commonly used before classification.

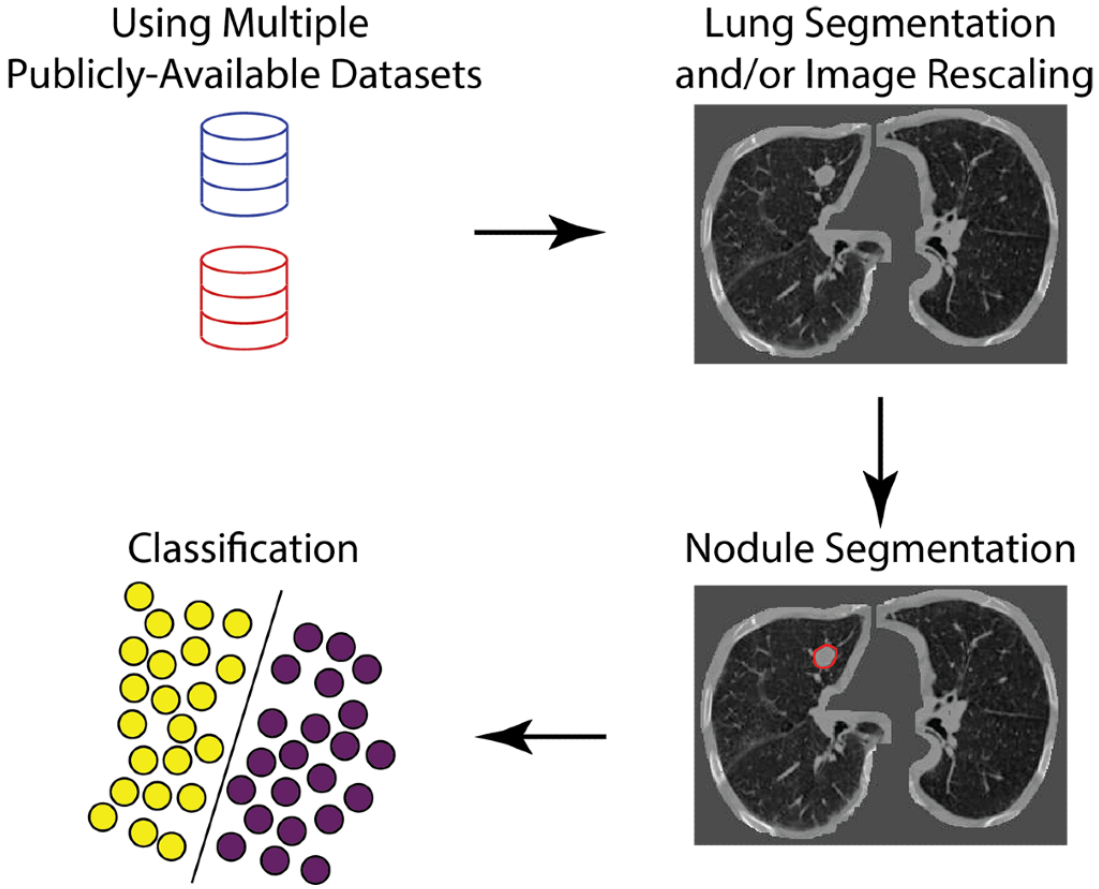


Table 1. Comparisons of the top-performing solutions of the Kaggle Data Science Bowl.

Rank	Team name	Additional datasets used	Data preprocessing	Nodule segmentation	Classification algorithms	Implementation	Final test set score
1	Grt123	LUNA16[a]	Lung segmentation, intensity normalization	Variant of U-Net	Neural network with a max-pooling layer and two fully connected layers	Pytorch	0.39975
2	Julian de Wit and Daniel Hammack	LUNA16, LIDC[b]	Rescale to 1×1×1	C3D[c], ResNet-like CNN[d]	C3D, ResNet-like CNN	Keras, Tensorflow, Theano	0.40117
3	Aidence	LUNA16	Rescale to 2.5×0.512×0.512 (for nodule detection) and 1.25×0.5×0.5 (for classification)	ResNet[e]	3D DenseNet[f] multitask model (different loss functions depending on the input source)	Tensorflow	0.40127
4	qfpxfd	LUNA16, SPIE-AAPM[g]	Lung segmentation	Faster R-CNN[h], with 3D CNN for false positive reduction	3D CNN inspired by VGGNet	Keras, Tensorflow, Caffe	0.40183
5	Pierre Fillard (Therapixel)	LUNA16	Rescale to 0.625×0.625×0.625, lung segmentation	3D CNN inspired by VGGNet	3D CNN inspired by VGGNet	Tensorflow	0.40409
6	MDai	None	Rescale to 1×1×1, normalize HU[i]	2D and 3D ResNet	3D ResNet + a Xgboost classifier incorporating CNN output, patient sex, # nodules, and other nodule features	Keras, Tensorflow, Xgboost	0.41629
7	DL Munich	LUNA16	Rescale to 1×1×1, lung segmentation	U-Net	2D and 3D residual neural network	Tensorflow	0.42751
8	Alex, Andre, Gilberto, and Shize	LUNA16	Rescale to 2×2×2	Variant of U-Net	CNN, tree-based classifiers (with better performance)	Keras, Theano, xgboost, extraTree	0.43019
9	Deep Breath	LUNA16, SPIE-AAPM[j]	Lung mask	Variant of SegNet	Inception-ResNet v2	Theano and Lasagne	0.43872
10	Owkin Team	LUNA16	Lung segmentation	U-Net, 3D VGGNet	Gradient boosting	Keras, Tensorflow, xgboost	0.44068

[a]LUNA16: Lung Nodule Analysis 2016.

[b]LIDC: Lung Image Database Consortium.

[c]C3D: convolutional 3D.

[d]ResNet-like CNN: residual net–like convolutional neural network.

[e]ResNet: residual net.

[f]DenseNet: dense convolutional network.

[g]SPIE-AAPM: International Society for Optics and Photonics–American Association of Physicists in Medicine Lung CT Challenge.

[h]R-CNN: region-based convolutional neural networks.

[i]HU: Hounsfield unit.

[j]Dataset has been evaluated but not used in building the final model.

Table 2. A summary of the chest computed tomography datasets employed by the participants.

Datasets	Number of CT[a] scan series	Data originated from multiple sites	Availability of nodule locations	Availability of nodule segmentations	Availability of patients' diagnoses (benign versus malignant)
Kaggle Data Science Bowl (this competition)	Training: 1397; public test set: 198; final test set: 506	Yes	No	No	Yes
Lung nodule analysis	888	Yes	Yes	Yes	Yes
SPIE-AAPM[b] Lung CT Challenge	70	No	Yes	No	Yes
Lung Image Database Consortium	1398	Yes	Yes	Yes	Yes

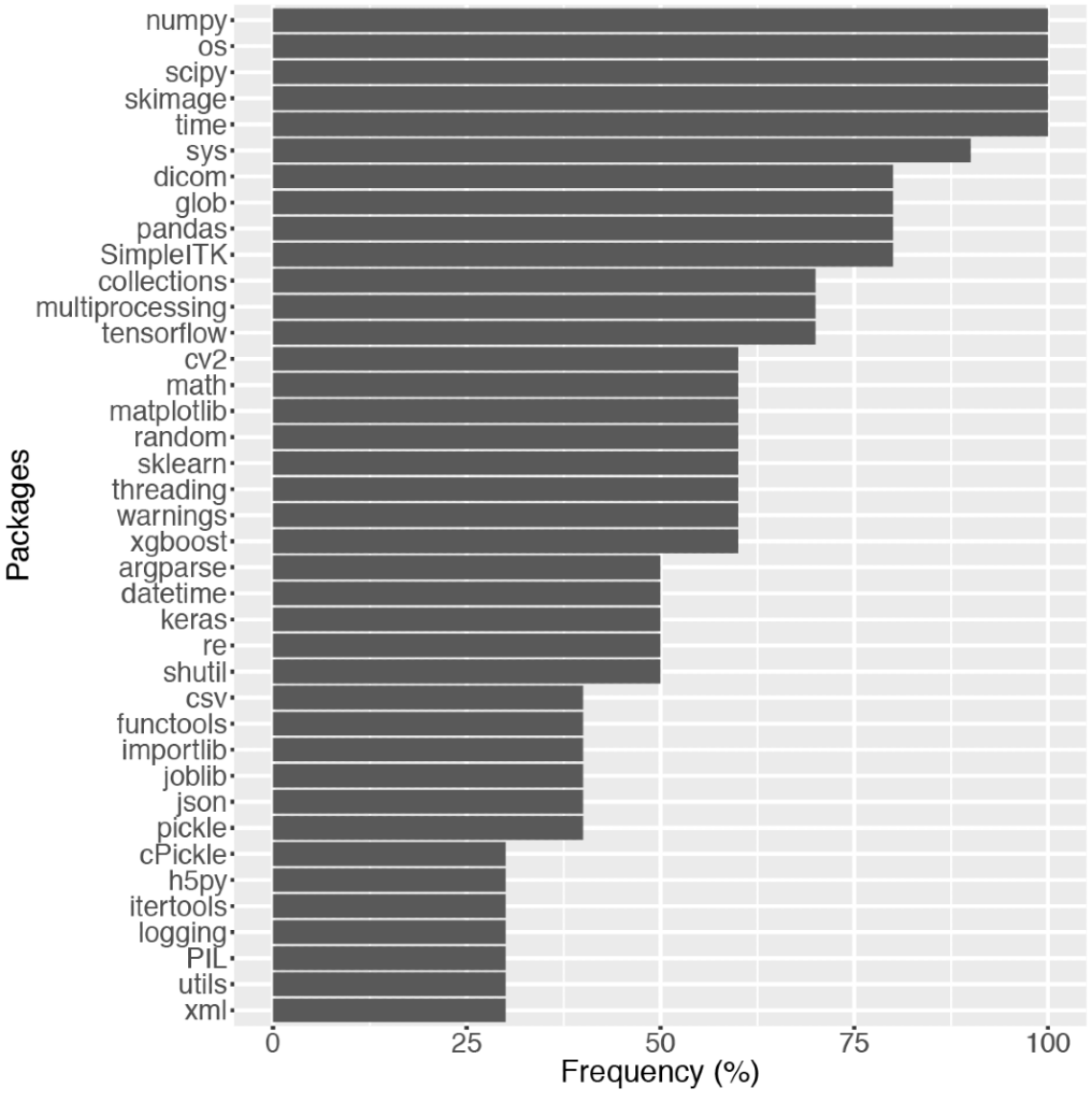
[a]CT: computed tomography.

[b]SPIE-AAPM: International Society for Optics and Photonics–American Association of Physicists in Medicine.

Comparison of the Implementation Platforms and Software Dependencies

Most of the winning teams developed their modules with Keras and Tensorflow. Only one team used Pytorch (the top-performing team), Caffe, or Lasagne. All of the top 10 teams employed a number of python packages for scientific computing and image processing, including NumPy, SciPy, and Scikit-image (skimage). A summary of package dependencies is shown in Figure 4. This reflected the popularity of the tools for processing chest CT images, building neural networks, and scientific computing among the top contestants of this contest.

Figure 4. The most widely used dependencies by the top 10 teams. The packages are ordered by their prevalence among the top teams. For simplicity, dependencies used by only one team are omitted from the figure.



Docker Images of the Top Solutions

To facilitate reusing the code developed by the top teams, we generated a Docker image for each of the available solutions. Our developed Docker images are redistributed under the open-source licenses chosen by the original developers [28]. Detailed instructions on accessing the Docker images can be found on GitHub [29].

Discussion

Principal Findings

This is the first study that systematically compared the algorithms and implementations of award-winning pulmonary nodule classifiers. Results showed that the majority of the best-performing solutions used additional datasets to train the pulmonary nodule segmentation models. The top solutions used different data preprocessing, segmentation, and classification algorithms. Nonetheless, they only differ slightly in their final test set scores.

The most commonly used data preprocessing steps were lung segmentation and voxel scaling [30]. For nodule classification, many solutions used CNNs. However, 2 of the top 10 teams employed tree-based methods for cancer versus noncancer classification. Tree-based approaches require a predefined set of image features, whereas CNNs allow data to refine the definition of features [31]. Given sufficient sample size, CNNs outperformed tree-based methods in many image-related tasks [12,32], whereas tree-based methods could reach satisfactory performance when the sample size was small, and they provided better model interpretability [33-35]. Since the conclusion of the contest, additional works on machine learning for CT evaluation have been published [36-40]. Nonetheless, these works reported similar strategies for data processing and classification overall.

To enhance the reproducibility of the developed modules, we generated a Docker image for each of the award-winning solutions. The Docker images contain all software dependencies and patches required by the source codes and are portable to various computing environments [16], which will expedite the application and improvement of the state-of-the-art CT analytical modules implemented by the contest winners.

Limitations

Since it was difficult to compile and release a large deidentified chest CT dataset to the public, the public test set only contains images from 198 patients. Leveraging the 5-digit precision of the log-loss value shown on the leaderboard, one participant implemented and shared a method for identifying all ground truth labels in the public test set during the competition [41]. Several participants successfully replicated this approach and got perfect scores on the public leaderboard. Thus, solutions with very low log-loss in the public test set may result from information leakage. Interestingly, among the top-10 models defined by the final test set, 2 performed worse than random guessing in the public test set, which raised concerns on their generalizability [42].

There are several approaches future contest organizers can take to ensure the generalizability of the developed models. First, a multistage competition can filter out the overfitted models using the first private test set and only allow reasonable models to advance to the final evaluation. In addition, organizers can discourage leaderboard probing by only showing the performance of a random subset of the public test data or limiting the number of submissions allowed per day. Finally, curating a larger test set can better evaluate the true model performance and reduce random variability [43]. If data deidentification is difficult, requiring contestants to submit their models to a secure computing environment rather than distributing the test data to the participants can minimize the risk of leaking identifiable medical information.

Conclusion

In summary, we compared, reproduced, and Dockerized state-of-the-art pulmonary nodule segmentation and classification modules. Results showed that many transfer learning approaches achieved reasonable accuracy in diagnosing chest CT images. Future works on additional data collections and validation will further enhance the generalizability of the current methods.

Acknowledgments

The authors express their appreciation to Dr Steven Seltzer for his feedback on the manuscript; Drs Shann-Ching Chen, Albert Tsung-Ying Ho, and Luke Kung for identifying the data resources; Dr Mu-Hung Tsai for pointing out the computing resources; and Ms Samantha Lemos and Nichole Parker for their administrative support. K-HY is a Harvard Data Science Fellow. This work was supported in part by the Blavatnik Center for Computational Biomedicine Award and grants from the Office of the Director, National Institutes of Health (grant number OT3OD025466), and the Ministry of Science and Technology Research Grant, Taiwan (grant numbers MOST 103-2221-E-006-254-MY2 and MOST 103-2221-E-168-019). The authors thank the Amazon Web Services Cloud Credits for Research, Microsoft Azure Research Award, and the NVIDIA Corporation for their support on the computational infrastructure. This work used the Extreme Science and Engineering Discovery Environment Bridges Pylon at the Pittsburgh Supercomputing Center (through allocation TG-BCS180016), which is supported by the National Science Foundation (grant number ACI-1548562).

Conflicts of Interest

None declared.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics 2012. CA Cancer J Clin 2015 Mar;65(2):87-108 [FREE Full text] [doi: 10.3322/caac.21262] [Medline: 25651787]
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Sep 12;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
3. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011 Aug 4;365(5):395-409 [FREE Full text] [doi: 10.1056/NEJMoa1102873] [Medline: 21714641]
4. Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013 May;143(5 Suppl):e93S-e120S [FREE Full text] [doi: 10.1378/chest.12-2351] [Medline: 23649456]
5. Lam VK, Miller M, Dowling L, Singhal S, Young RP, Cabebe EC. Community low-dose CT lung cancer screening: a prospective cohort study. Lung 2015 Feb;193(1):135-139. [doi: 10.1007/s00408-014-9671-9] [Medline: 25503535]
6. Moyer VA, U.S. Preventive Services Task Forsce. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2014 Mar 4;160(5):330-338. [doi: 10.7326/M13-2771] [Medline: 24378917]
7. Gurcan MN, Sahiner B, Petrick N, Chan H, Kazerooni EA, Cascade PN, et al. Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system. Med Phys 2002 Nov;29(11):2552-2558. [doi: 10.1118/1.1515762] [Medline: 12462722]
8. Singh SP, Gierada DS, Pinsky P, Sanders C, Fineberg N, Sun Y, et al. Reader variability in identifying pulmonary nodules on chest radiographs from the national lung screening trial. J Thorac Imaging 2012 Jul;27(4):249-254 [FREE Full text] [doi: 10.1097/RTI.0b013e318256951e] [Medline: 22627615]
9. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph 2007;31(4-5):198-211 [FREE Full text] [doi: 10.1016/j.compmedimag.2007.02.002] [Medline: 17349778]
10. Rubin GD, Lyo JK, Paik DS, Sherbondy AJ, Chow LC, Leung AN, et al. Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. Radiology 2005 Jan;234(1):274-283. [doi: 10.1148/radiol.2341040589] [Medline: 15537839]
11. Awai K, Murao K, Ozawa A, Komi M, Hayakawa H, Hori S, et al. Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance. Radiology 2004 Feb;230(2):347-352. [doi: 10.1148/radiol.2302030049] [Medline: 14752180]
12. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct;2(10):719-731 [FREE Full text] [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
13. Data Science Bowl 2017. Kaggle Inc. URL: https://www.kaggle.com/c/data-science-bowl-2017 [accessed 2020-07-12]
14. Healthcare 2020. Kaggle Inc. URL: https://www.kaggle.com/tags/healthcare [accessed 2020-07-12]
15. About Docker. URL: https://www.docker.com/company [accessed 2020-07-12]
16. Boettiger C. An introduction to Docker for reproducible research. SIGOPS Oper Syst Rev 2015 Jan 20;49(1):71-79. [doi: 10.1145/2723872.2723882]
17. National Lung Screening Trial Research Team, Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, et al. The National Lung Screening Trial: overview and study design. Radiology 2011 Jan;258(1):243-253 [FREE Full text] [doi: 10.1148/radiol.10091808] [Medline: 21045183]
18. Data Science Bowl 2017: data description. Kaggle Inc. URL: https://www.kaggle.com/c/data-science-bowl-2017/data [accessed 2020-07-12]
19. Open Source Initiative. URL: https://opensource.org/ [accessed 2020-07-12]
20. Data Science Bowl 2017: rules. Kaggle Inc. URL: https://www.kaggle.com/c/data-science-bowl-2017/rules [accessed 2020-07-12]
21. Setio AAA, Traverso A, de Bel T, Berens MS, Bogaard CVD, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Med Image Analysis 2017 Dec;42:1-13. [doi: 10.1016/j.media.2017.06.015]
22. Armato SG, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys 2011 Feb;38(2):915-931 [FREE Full text] [doi: 10.1118/1.3528204] [Medline: 21452728]
23. Armato SG, Drukker K, Li F, Hadjiiski L, Tourassi GD, Engelmann RM, et al. LUNGx Challenge for computerized lung nodule classification. J Med Imaging (Bellingham) 2016 Oct;3(4):044506 [FREE Full text] [doi: 10.1117/1.JMI.3.4.044506] [Medline: 28018939]
24. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Cham: Springer; 2015.

25. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv 2014:14091556.
26. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; 2016; Las Vegas. [doi: 10.1109/cvpr.2016.90]
27. Shin H, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016 May;35(5):1285-1298. [doi: 10.1109/tmi.2016.2528162]
28. DockerHub. URL: https://rebrand.ly/chestct [accessed 2020-07-12]
29. Docker images for chest CT analyses. URL: https://github.com/khyu/dsb_chestct [accessed 2020-07-12]
30. Lifton J, Malcolm A, McBride J, Cross K. The application of voxel size correction in X-ray computed tomography for dimensional metrology. 2013. URL: https://www.ndt.net/article/SINCE2013/content/papers/24_Lifton.pdf [accessed 2020-07-12]
31. Yu K, Wang F, Berry GJ, Ré C, Altman RB, Snyder M, et al. Classifying non-small cell lung cancer types and transcriptomic subtypes using convolutional neural networks. J Am Med Inform Assoc 2020 May 01;27(5):757-769. [doi: 10.1093/jamia/ocz230] [Medline: 32364237]
32. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015 Apr 11;115(3):211-252. [doi: 10.1007/s11263-015-0816-y]
33. Yu K, Berry GJ, Rubin DL, Ré C, Altman RB, Snyder M. Association of omics features with histopathology patterns in lung adenocarcinoma. Cell Syst 2017 Dec 27;5(6):620-627 [FREE Full text] [doi: 10.1016/j.cels.2017.10.014] [Medline: 29153840]
34. Yu K, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun 2016 Dec 16;7:12474 [FREE Full text] [doi: 10.1038/ncomms12474] [Medline: 27527408]
35. Yu K, Fitzpatrick MR, Pappas L, Chan W, Kung J, Snyder M. Omics AnalySIs System for PRecision Oncology (OASISPRO): a web-based omics analysis tool for clinical phenotype prediction. Bioinformatics 2018 Jan 15;34(2):319-320 [FREE Full text] [doi: 10.1093/bioinformatics/btx572] [Medline: 28968749]
36. Nishio M, Sugiyama O, Yakami M, Ueno S, Kubo T, Kuroda T, et al. Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. PLoS One 2018;13(7):e0200721 [FREE Full text] [doi: 10.1371/journal.pone.0200721] [Medline: 30052644]
37. da Silva GLF, Valente TLA, Silva AC, de Paiva AC, Gattass M. Convolutional neural network-based PSO for lung nodule false positive reduction on CT images. Comput Methods Programs Biomed 2018 Aug;162:109-118. [doi: 10.1016/j.cmpb.2018.05.006] [Medline: 29903476]
38. Ait Skourt B, El Hassani A, Majda A. Lung CT image segmentation using deep neural networks. Procedia Comput Sci 2018;127:109-113. [doi: 10.1016/j.procs.2018.01.104]
39. Lakshmanaprabu S, Mohanty SN, Shankar K, Arunkumar N, Ramirez G. Optimal deep learning model for classification of lung cancer on CT images. Future Gen Comput Syst 2019 Mar;92:374-382. [doi: 10.1016/j.future.2018.10.009]
40. Xie H, Yang D, Sun N, Chen Z, Zhang Y. Automated pulmonary nodule detection in CT images using deep convolutional neural networks. Pattern Recognition 2019 Jan;85:109-119. [doi: 10.1016/j.patcog.2018.07.031]
41. Trott O. The Perfect Score script. 2017. URL: https://www.kaggle.com/olegtrott/the-perfect-score-script [accessed 2020-05-24]
42. Yu K, Kohane IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf 2019 Mar;28(3):238-241. [doi: 10.1136/bmjqs-2018-008551] [Medline: 30291179]
43. Brain D, Webb GI. On the effect of data set size on bias and variance in classification learning. 1999. URL: https://pdfs.semanticscholar.org/6daf/14076a7ad2c4b4f1a8eb28778b9c641777e7.pdf [accessed 2020-07-12]

Abbreviations

AAPM: American Association of Physicists in Medicine
CNN: convolutional neural network
CT: computed tomography
LUNA16: Lung Nodule Analysis 2016
ResNet: residual net
skimage: Scikit-image
SPIE: International Society for Optics and Photonics

Edited by G Eysenbach; submitted 17.10.19; peer-reviewed by J Shull, M Morid; comments to author 18.05.20; revised version received 25.05.20; accepted 11.06.20; published 05.08.20.

Please cite as:
Yu KH, Lee TLM, Yen MH, Kou SC, Rosen B, Chiang JH, Kohane IS
Reproducible Machine Learning Methods for Lung Cancer Detection Using Computed Tomography Images: Algorithm Development and Validation
J Med Internet Res 2020;22(8):e16709
URL: https://www.jmir.org/2020/8/e16709
doi:10.2196/16709
PMID:32755895

Original Paper

Applying Machine Learning Models with An Ensemble Approach for Accurate Real-Time Influenza Forecasting in Taiwan: Development and Validation Study

Hao-Yuan Cheng[1*], MD, MSc; Yu-Chun Wu[2*], PhD; Min-Hau Lin[1], PhD; Yu-Lun Liu[1], MD, MSc; Yue-Yang Tsai[2], MSc; Jo-Hua Wu[2], MSc; Ke-Han Pan[2], MSc; Chih-Jung Ke[1], MSc; Chiu-Mei Chen[1], MSc; Ding-Ping Liu[1,3], PhD; I-Feng Lin[4], DrPH; Jen-Hsiang Chuang[1,4], MD, PhD

[1]Taiwan Centers for Disease Control, Taipei, Taiwan

[2]Value Lab, Acer Inc., Taipei, Taiwan

[3]National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

[4]Institute of Public Health, National Yang-Ming University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Jen-Hsiang Chuang, MD, PhD
Taiwan Centers for Disease Control
9F, No. 6, Linsen S. Road
Zhong-zheng District
Taipei, 100
Taiwan
Phone: 886 2 2395 9825
Email: jhchuang@cdc.gov.tw

Abstract

Background: Changeful seasonal influenza activity in subtropical areas such as Taiwan causes problems in epidemic preparedness. The Taiwan Centers for Disease Control has maintained real-time national influenza surveillance systems since 2004. Except for timely monitoring, epidemic forecasting using the national influenza surveillance data can provide pivotal information for public health response.

Objective: We aimed to develop predictive models using machine learning to provide real-time influenza-like illness forecasts.

Methods: Using surveillance data of influenza-like illness visits from emergency departments (from the Real-Time Outbreak and Disease Surveillance System), outpatient departments (from the National Health Insurance database), and the records of patients with severe influenza with complications (from the National Notifiable Disease Surveillance System), we developed 4 machine learning models (autoregressive integrated moving average, random forest, support vector regression, and extreme gradient boosting) to produce weekly influenza-like illness predictions for a given week and 3 subsequent weeks. We established a framework of the machine learning models and used an ensemble approach called stacking to integrate these predictions. We trained the models using historical data from 2008-2014. We evaluated their predictive ability during 2015-2017 for each of the 4-week time periods using Pearson correlation, mean absolute percentage error (MAPE), and hit rate of trend prediction. A dashboard website was built to visualize the forecasts, and the results of real-world implementation of this forecasting framework in 2018 were evaluated using the same metrics.

Results: All models could accurately predict the timing and magnitudes of the seasonal peaks in the then-current week (nowcast) (ρ=0.802-0.965; MAPE: 5.2%-9.2%; hit rate: 0.577-0.756), 1-week (ρ=0.803-0.918; MAPE: 8.3%-11.8%; hit rate: 0.643-0.747), 2-week (ρ=0.783-0.867; MAPE: 10.1%-15.3%; hit rate: 0.669-0.734), and 3-week forecasts (ρ=0.676-0.801; MAPE: 12.0%-18.9%; hit rate: 0.643-0.786), especially the ensemble model. In real-world implementation in 2018, the forecasting performance was still accurate in nowcasts (ρ=0.875-0.969; MAPE: 5.3%-8.0%; hit rate: 0.582-0.782) and remained satisfactory in 3-week forecasts (ρ=0.721-0.908; MAPE: 7.6%-13.5%; hit rate: 0.596-0.904).

Conclusions: This machine learning and ensemble approach can make accurate, real-time influenza-like illness forecasts for a 4-week period, and thus, facilitate decision making.

(J Med Internet Res 2020;22(8):e15394) doi:10.2196/15394

KEYWORDS

influenza; Influenza-like illness; forecasting; machine learning; artificial intelligence; epidemic forecasting; surveillance

Introduction

Seasonal influenza is one of the most prevalent infectious diseases in Taiwan, accounting for millions of cases, over tens of thousands of patient hospitalizations, and hundreds of deaths annually [1-4]. In Taiwan, the seasonal influenza epidemic typically begins in winter and continues to the end of the year until the spring of the next year [2]. However, the changeful influenza activity in subtropical areas like Taiwan sometimes causes problems in epidemic preparedness. For instance, in Taiwan, the 2015-2016 influenza epidemic, with H1N1 as the main circulating strain, was the biggest since the 2009 novel H1N1 outbreak. Nevertheless, H1N1 influenza activity was unexpectedly low in the following 2016-2017 influenza season, whereas H3N2 influenza activity peaked unpredictably in the summer season in 2017 and caused a severe epidemic [2,3].

Since 2004, to monitor changes in influenza activity, the Taiwan Centers for Disease Control (Taiwan CDC) has established real-time national influenza surveillance systems for influenza-like illness visits to hospitals and clinics [5,6]. The surveillance systems have minimal time lag in data collection; therefore, public-health professionals can immediately adjust their response almost in real time. The decision-making process, however, remains based only on past data (despite the short time lag). Influenza epidemic forecasting for upcoming weeks or months can provide more information for policymaking and is relevant for preparedness [7]. The ability to provide a short-term forecast in terms of epidemic magnitude is particularly vital for emergency departments during a long weekend or the Lunar New Year in Taiwan (eg, the Lunar New Year comprised 9 vacation days in 2019), during which time, influenza-like illness visits at emergency departments considerably increase (since outpatient services are closed), and sometimes patients crowd the emergency departments. In this situation, reliable forecasts are required to determine the surging capacity.

Many research teams have worked on influenza forecasting for a long time. Among the models used by researchers, the autoregressive integrated moving average (ARIMA) model is a methodology that is often chosen for seasonal influenza forecasts because of its advantage in dealing with time-series data [8,9], its satisfactory performance using data that are time dependent for short-term projection, and its widespread use in other health-related forecasting tasks [9-14]. Decision tree–based machine learning algorithms such as random forest and extreme gradient boosting also have their strengths in predictive analysis and forecasting, which has been shown in data science competitions such as Kaggle [15], influenza outbreaks [14], and foodborne disease trends [16]. A study in Canada [17] showed random forest models had better performance predicting influenza A virus frequency than that of ARIMA and generalized linear autoregressive integrated moving average models. Unlike ARIMA, random forest and extreme gradient boosting, as ensemble weak prediction models, have better performance dealing with high-dimension data [18], while support vector regression's strength is finding an optimal hyperplane with a nonlinear boundary [19,20]. Previous research has also demonstrated a successful combination of linear regression with nonlinear predictor, random forest, support vector regression, and extreme gradient boosting to predict dengue fever outbreak in the United States [21].

Instead of traditional surveillance data, researchers have also attempted using nontraditional data sources, such as Google Flu Trends and Flu Near You, to improve their forecasts since 2008 [22,23]. These data served as surrogate indicators or supplement data for influenza-like illness activity. Lasso regression, random forest, extreme gradient boosting, and support vector regression have been widely implemented to aggregate these data from Google search, Google trend, Wikipedia, and social media (such as Twitter and Baidu) in influenza forecasting [24-27]. The performance of elastic net and support vector regression was considered to be comparable in a study [26] which used the Baidu index as a predictor and predicted the number of influenza cases in China by support vector regression, and in a study [28] in France which used electronic health record data with historical epidemiology information for influenza-like illness incidence rate predictions.

On the other hand, researchers began to explore the possibility of simultaneously using multiple models or data sources to find an ensemble approach to produce more robust forecasts by combining the results of different forecasting models [11,29-32]. For seasonal influenza forecasting in the US, the empirical Bayes method has been used to integrate the forecasts from linear models using multiple data sources as predictors [29]. Kandula et al [32] also evaluated the performance of the susceptible-exposed-infectious-recovered-susceptible model, Bayesian weighted outbreaks, k-nearest neighbor, and a superensemble method when combining distinct forecast methods to predict influenza outbreaks in the United States. A meta-ensemble of statistical and mechanistic methods has shown better accuracy than individual methods [31,32].

Compared to internet data, which might easily be influenced by search engine marketing, the surveillance database in Taiwan can provide much more comprehensive data with a small time lag. These data sources are also easier to be maintained and reliable for a long-term decision-making system. Therefore, using the surveillance data, we aimed to develop a practical framework consisting of an ensemble model with machine learning models to combine the advantages of different forecasting models for real-time influenza-like illness predictions, and facilitate influenza preparedness.

Methods

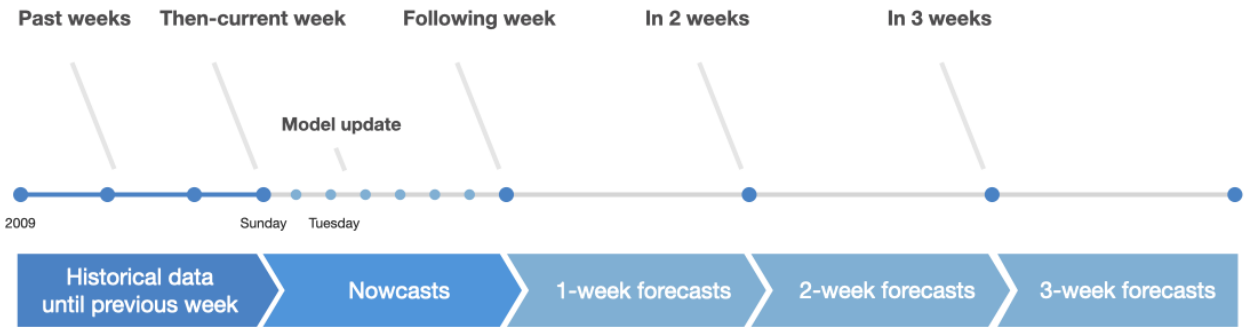
Data Source

The data we used to train and validate the machine learning algorithm included weekly data from the Real-Time Outbreak and Disease Surveillance System, the National Health Insurance Database, and the National Notifiable Disease Surveillance System [5,6]. The details and characteristics of the data sets are described in Multimedia Appendix 1. Other data used include the number of national holidays in each week, regular weekends, and long weekends. We used surveillance data from 2008-2017 to establish the framework of the forecasting models.

Forecasting Targets and the Renewal of the Surveillance Data and Models

The forecasting targets in our study were short-term forecasts—weekly number of influenza-like illness visits for the 4-week period after the most recent surveillance data. Real-Time Outbreak and Disease Surveillance System, National Health Insurance, and National Notifiable Disease Surveillance System databases were updated daily. Because of the potential delay in data entry by hospitals, all the models were automatically retrained and updated every Tuesday night using data up until the end of the previous week. The updated models would then produce the predictions of influenza-like illness visits for 4 weeks from that time point (the end of the previous week). Therefore, the initial forecast actually predicts the number of influenza-like illness visits in the then-current week (*nowcast*), whereas the *1-week*, *2-week*, and *3-week forecasts* represent the weekly predictions for each of the subsequent 3 weeks (Figure 1).

Figure 1. Timeline of historical data used for model training and forecasting periods.



Machine Learning Algorithms

Four machine learning algorithms—ARIMA, random forest, support vector regression, and extreme gradient boosting—were used to produce weekly influenza-like illness predictions for a 4-week period. We chose these algorithms, each with different characteristics and strengths, so that the forecasting task could benefit from the diversity of the machine learning algorithms.

To summarize the forecasts of the 4 different machine learning algorithms, we adopted the ensemble method called stacking [31,33] An ensemble model was trained using another support vector regression algorithm with a linear kernel that optimized the best regression between the observed number of influenza-like illness visits and the 4-week forecasts. A previous study [32] adopted a Bayesian model, which requires the prior distribution estimation to produce the ensemble forecast. We chose a support vector regression with a linear kernel model because it can produce a weighted-average forecast from 4 individual models without considering data distribution. By using the stacking method, the forecasts of different algorithms are automatically weighted and combined to produce the ensemble forecasts. The linear kernel was chosen because of the forecasting and efficient computing performance it showed in the training process. The hyperparameter tuning mechanism, described in the section that follows, was used to evaluate the performance of the ensemble model from the first week of 2015 to the 40th week of 2017.

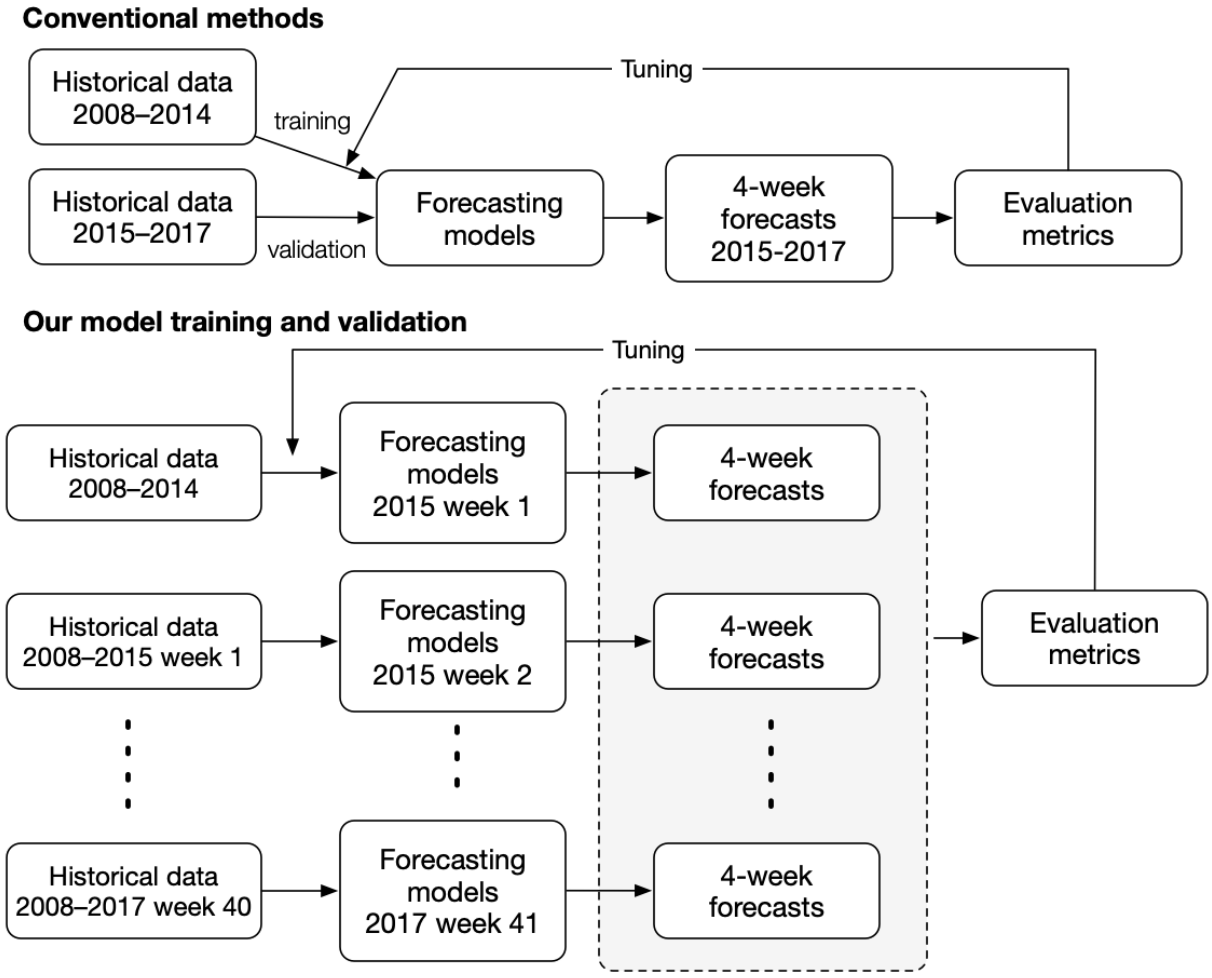
Feature Selection, Engineering, and Model Tuning

The initial features were selected after discussions with experts of Taiwan CDC. The number of past influenza-like illness visits in the 8 previous weeks (from the Real-Time Outbreak and Disease Surveillance System and National Health Insurance database) and the length of national holidays in a week were the basic features. We also included essential holidays, such as the Lunar New Year, in the feature set because it was believed to have a significant influence on influenza-like illness visits, especially in emergency departments. Our feature engineering work included moving average, moving difference with varying time lags, and the proportion of influenza-like illness visits to total medical visits (Multimedia Appendix 2).

We chose naïve (heuristic) mechanisms, instead of the conventional methods, for feature selection. We used surveillance data from 2008-2017 for feature selection and model tuning. We evaluated the overall forecasting performance of the algorithms during the first week of 2015 to the 40th week of 2017 by comparing the forecasts to observed historical data in the same period. Using this framework, we dynamically

retrained each model from zero every week to incorporate newly collected data (Figure 2):

Figure 2. The framework of feature selection and model tuning for our model training and validation compared to the conventional method.



1. For an individual week T from the first week of 2015, the training data set included the weekly influenza-like illness data from week 1 in 2008 to week T.

2. Given a set of features and fixed hyperparameter value, the model $f(\cdot|T, h)$, forecasting the h weeks ahead, was trained using the training data set at week T .

3. The number of influenza-like illness visits in week $T + h$ was forecasted by the trained $f(\cdot|T, h)$.

4. The forecasted number was compared to the observed number of influenza-like illness visits in week $T + h$ using the evaluation metrics.

5. For each week from the first week of 2015 to the 40th week of 2017, we repeated Step 1 to Step 4 and calculated the evaluation metrics for specific feature sets and hyperparameters. Then we selected the feature set and hyperparameters that performed best in evaluation metrics.

If we only used k-fold cross-validation during model training, look-ahead bias might have occurred when using time-series data with potential autocorrelation. The advantage of this framework avoided look-ahead bias and made use of all historical data before the week T to train the models in forecasting the weekly influenza-like illness visits of the week $T + h$ at the week T.

Evaluation Metrics

The metrics we used to evaluate the model performance included Pearson correlation (ρ), root mean squared error (RMSE), mean absolute percentage error (MAPE), and hit rate (Multimedia Appendix 3). Lower MAPE, lower RMSE, higher hit rate, and higher correlation indicated better forecasting performance.

Software and Visualization

We used data munging and feature engineering (dplyr), the time-series model, ARIMA (forecast) , random forest model (randomForest), support vector regression (e1071), and extreme gradient boosting model (xgboost) packages in R (version 3.4.4) on Ubuntu (version 14.0.4). The functions and hyperparameters that were used are listed in Multimedia Appendix 4. A visualization dashboard website was designed to display and compare the predictions of the 5 models (using D3.js and several JavaScript frameworks) [34].

Results

Real-Time Estimates (Nowcast)

The visualized comparison of the estimated epidemic curve in nowcasts and the observed number of influenza-like illness visits showed that all the models, especially the ensemble model, could predict the time and magnitude of the peaks of the influenza epidemic throughout the influenza season from 2015-2017, such as the peaks of the Lunar New Year vacation for each year and the peak of the summer flu in 2017 (Figure 3). All models could appropriately fit the epidemic curve of the outpatient (ρ=0.891-0.962) and emergency (ρ=0.802-0.967) departments (Table 1).

For 2015-2017, the nowcast prediction by the 4 machine learning models exhibited good accuracy (MAPE as low as 5.2%); however, the ensemble model (outpatient: ρ=0.956, MAPE 6.0%, hit rate 0.756; emergency: ρ= 0.967, MAPE 5.2%, hit rate 0.705) outperformed individual models.

Figure 3. Nowcasts (current week predictions) of the influenza-like illness visits in outpatient and emergency departments by the 5 machine learning models (colored lines) compared with the observed historical data (black line), 2015-2017. ARIMA: autoregressive integrated moving average; ILI: influenza-like illness; RF: random forest; SVR: support vector regression; XGB: extreme gradient boosting.

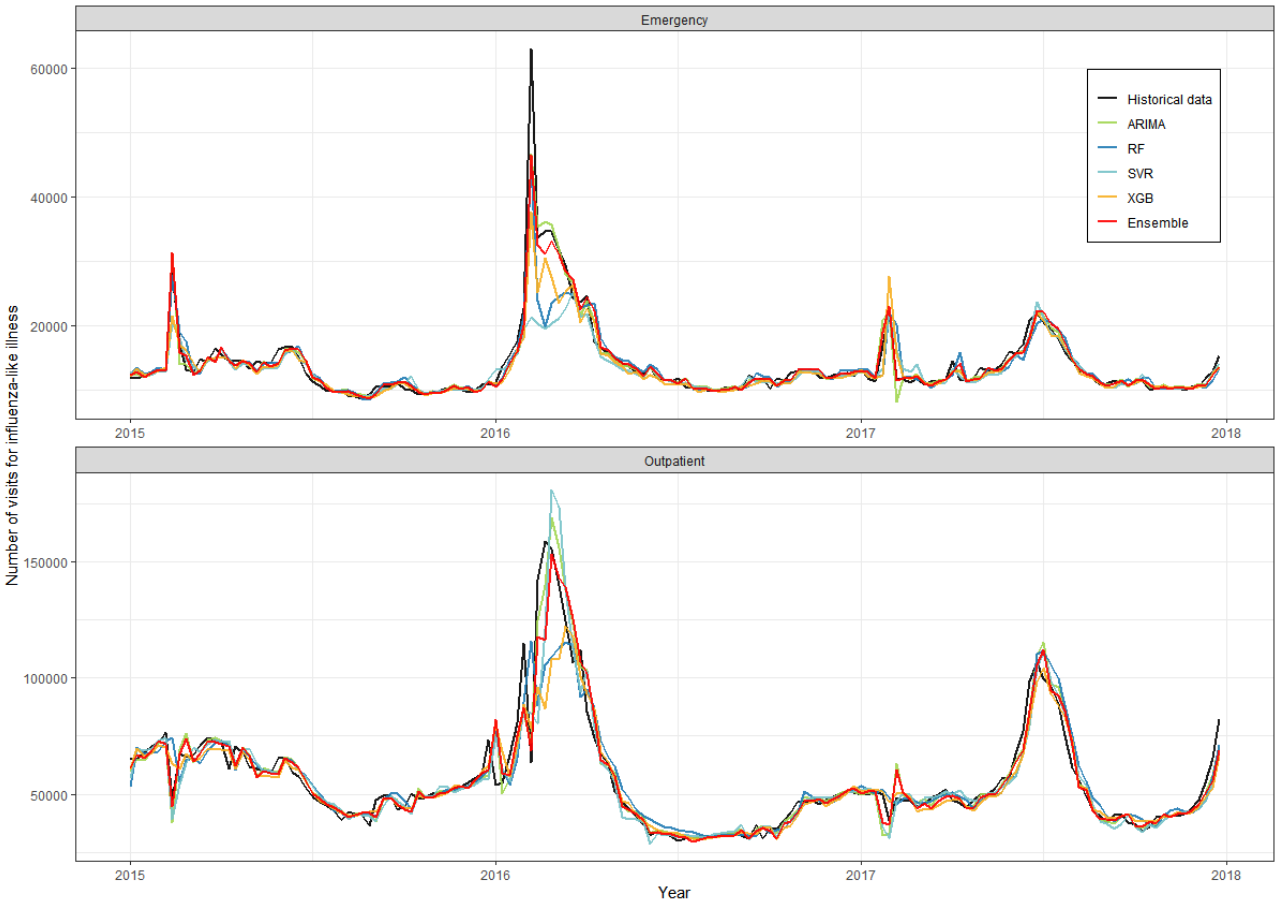


Table 1. The evaluation metrics of the 5 machine learning models for then-current week forecasts (nowcast), 1-week forecasts, 2-week forecasts, and 3-week forecasts for 2015 to 2017 data.

Time period	Outpatient influenza-like illness visits				Emergency influenza-like illness visits			
Model	RMSE[a]	MAPE[b], %	Hit rate	Pearson correlation coefficient	RMSE	MAPE, %	Hit rate	Pearson correlation coefficient
Nowcast (current week)								
ARIMA[c]	6621.9	6.5	0.744	0.962	1689.8	5.2	0.718	0.965
RF[d]	10773.1	9.2	0.577	0.891	2707.8	7.6	0.609	0.922
SVR[e]	9265.0	7.7	0.686	0.923	4189.8	7.1	0.686	0.802
XGB[f]	10063.6	7.3	0.635	0.915	2696.0	6.6	0.667	0.935
Ensemble	6903.5	6.0	0.756	0.956	1696.3	5.2	0.705	0.967
1-week								
ARIMA	11165.2	11.5	0.695	0.892	2562.1	8.3	0.747	0.918
RF	12256.6	11.7	0.701	0.855	3430.0	11.8	0.643	0.842
SVR	11573.6	10.7	0.740	0.874	4351.8	9.4	0.701	0.803
XGB	13604.7	11.0	0.695	0.836	3866.7	9.7	0.688	0.842
Ensemble	11752.8	10.1	0.721	0.874	2831.6	8.3	0.708	0.901
2-week								
ARIMA	15471.8	15.3	0.695	0.792	3206.2	10.1	0.727	0.867
RF	13464.5	13.8	0.721	0.823	3639.8	13.7	0.669	0.816
SVR	13972.9	13.7	0.708	0.808	4562.1	10.8	0.734	0.783
XGB	15317.7	13.6	0.727	0.785	4235.0	11.3	0.708	0.817
Ensemble	13758.1	12.0	0.727	0.823	3467.9	10.4	0.727	0.860
3-week								
ARIMA	19338.3	18.9	0.669	0.676	3836.8	12.0	0.688	0.801
RF	14310.9	14.9	0.753	0.796	3949.8	15.0	0.708	0.777
SVR	16004.3	15.9	0.786	0.743	4903.4	12.1	0.675	0.731
XGB	16888.8	15.7	0.708	0.723	4823.2	13.5	0.734	0.686
Ensemble	15193.6	13.3	0.773	0.780	3937.2	13.1	0.643	0.797

[a]RMSE: root mean squared error.

[b]MAPE: mean absolute percentage error.

[c]ARIMA: autoregressive integrated moving average.

[d]RF: random forest.

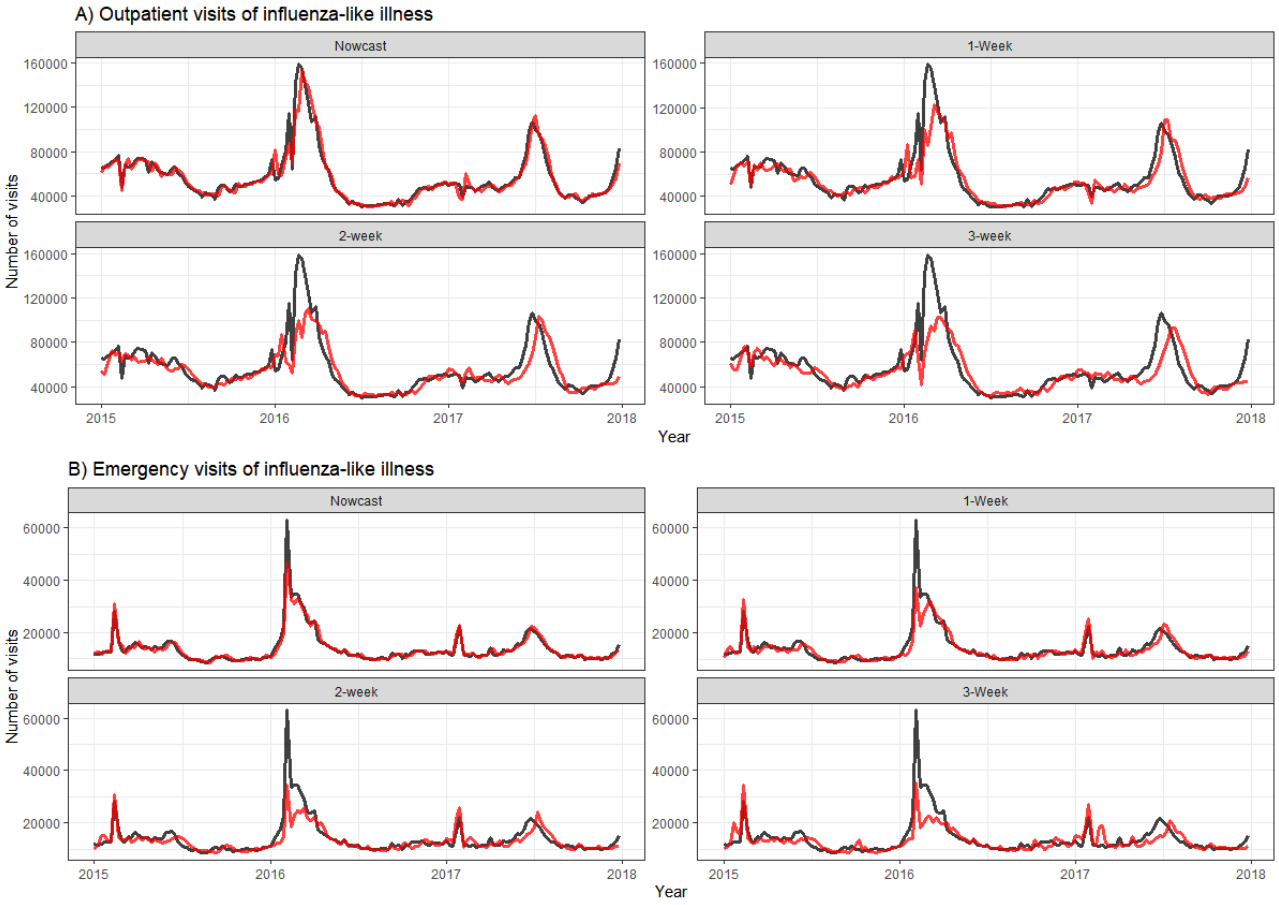
[e]SVR: support vector regression.

[f]XGB: extreme gradient boosting.

Forecasts for the Following 3 Weeks

The forecasts for the following 3 weeks using our ensemble model exhibited satisfactory performance for predicting the epidemic trend and successfully captured the epidemic peaks. Still, there were some time lags in peak prediction in the 1-, 2-, and 3-week forecasts (Figure 4). The accuracy slightly decreased with an increase in the forecast time horizons as well (MAPE: 8.3%-18.9%; hit rate: 0.643-0.786 in the 1-week, 2-week, and 3-week forecasts) (Table 1). Although the ARIMA model had the highest accuracy and hit rate in nowcasts, the random forest and support vector regression models performed better in the forecasts of the subsequent 2 and 3 weeks, particularly in terms of outpatient influenza-like illness visits.

Figure 4. Forecasts using the ensemble model (red) and the observed (black) number for influenza-like illness visits in (A) outpatient and (B) emergency departments, 2015-2017.



Real-World Application in 2018

We started using the framework with the 5 models in Taiwan CDC in early 2018. Since 2018, the nowcasts of our models has exhibited good accuracy (outpatient MAPE: 5.3%-5.8%; emergency MAPE: 5.7%-8.0%). Moreover, the 3-week forecasts maintained comparable accuracy to one another (outpatient MAPE: 8.8%-13.5%; emergency MAPE: 8.8%-13.5%; Table 2 and Multimedia Appendix 5). Hit rates of the nowcasts were 0.600-0.727 in outpatient and 0.582-0.782 in emergency department and remained at a high level in the 3-week forecasts (0.787-0.908 and 0.596-0.788 in outpatient and emergency department, respectively). All the models could approximately detect the declining trend when the magnitude of the epidemic had already reached a peak (Figure 5). The random forest and extreme gradient boosting model better identified the increasing trend during the early stage of the epidemic.

Table 2. Evaluation metrics of the 5 machine learning models for the current week forecasts (nowcast), 1-week forecasts, 2-week forecasts, and 3-week forecasts in 2018.

Model	Outpatient influenza-like illness visits				Emergency influenza-like illness visits			
	RMSE[a]	MAPE[b], %	Hit rate	Pearson correlation coefficient	RMSE	MAPE, %	Hit rate	Pearson correlation coefficient
Nowcast (current week)								
ARIMA[c]	6422.8	5.5	0.727	0.958	1125.3	5.7	0.782	0.965
RF[d]	6472.1	5.4	0.709	0.957	1305.8	6.2	0.709	0.952
SVR[e]	5343.2	5.8	0.655	0.969	2079.6	8.0	0.582	0.875
XGB[f]	7384.6	5.8	0.691	0.943	1643.6	6.2	0.673	0.923
Ensemble	6170.7	5.3	0.600	0.962	1751.0	6.5	0.727	0.912
1-week								
ARIMA	9874.7	9.0	0.741	0.897	1707.6	7.8	0.704	0.919
RF	8644.1	7.2	0.833	0.921	1861.2	7.9	0.759	0.899
SVR	7330.1	7.9	0.778	0.942	2643.7	10.6	0.759	0.798
XGB	9738.7	9.0	0.741	0.903	2353.3	8.3	0.685	0.836
Ensemble	9156.8	7.5	0.796	0.911	2363.0	8.4	0.722	0.832
2-week								
ARIMA	11630.6	11.8	0.811	0.851	1922.2	8.6	0.755	0.893
RF	10082.3	9.4	0.811	0.893	1975.5	7.7	0.736	0.888
SVR	9292.7	10.1	0.830	0.905	3031.0	12.8	0.566	0.730
XGB	11262.8	11.3	0.755	0.873	2371.6	7.9	0.698	0.843
Ensemble	10078.8	8.6	0.774	0.889	2389.4	8.6	0.755	0.835
3-week								
ARIMA	13656.7	13.5	0.865	0.787	1875.4	9.5	0.788	0.898
RF	10258.0	10.1	0.769	0.892	2041.9	8.8	0.692	0.877
SVR	9439.7	10.9	0.885	0.904	3106.0	13.5	0.596	0.721
XGB	12789.3	13.0	0.788	0.830	2259.6	7.6	0.692	0.890
Ensemble	9160.9	8.8	0.904	0.908	2478.4	9.6	0.769	0.814

[a]RMSE: root mean squared error.

[b]MAPE: mean absolute percentage error.

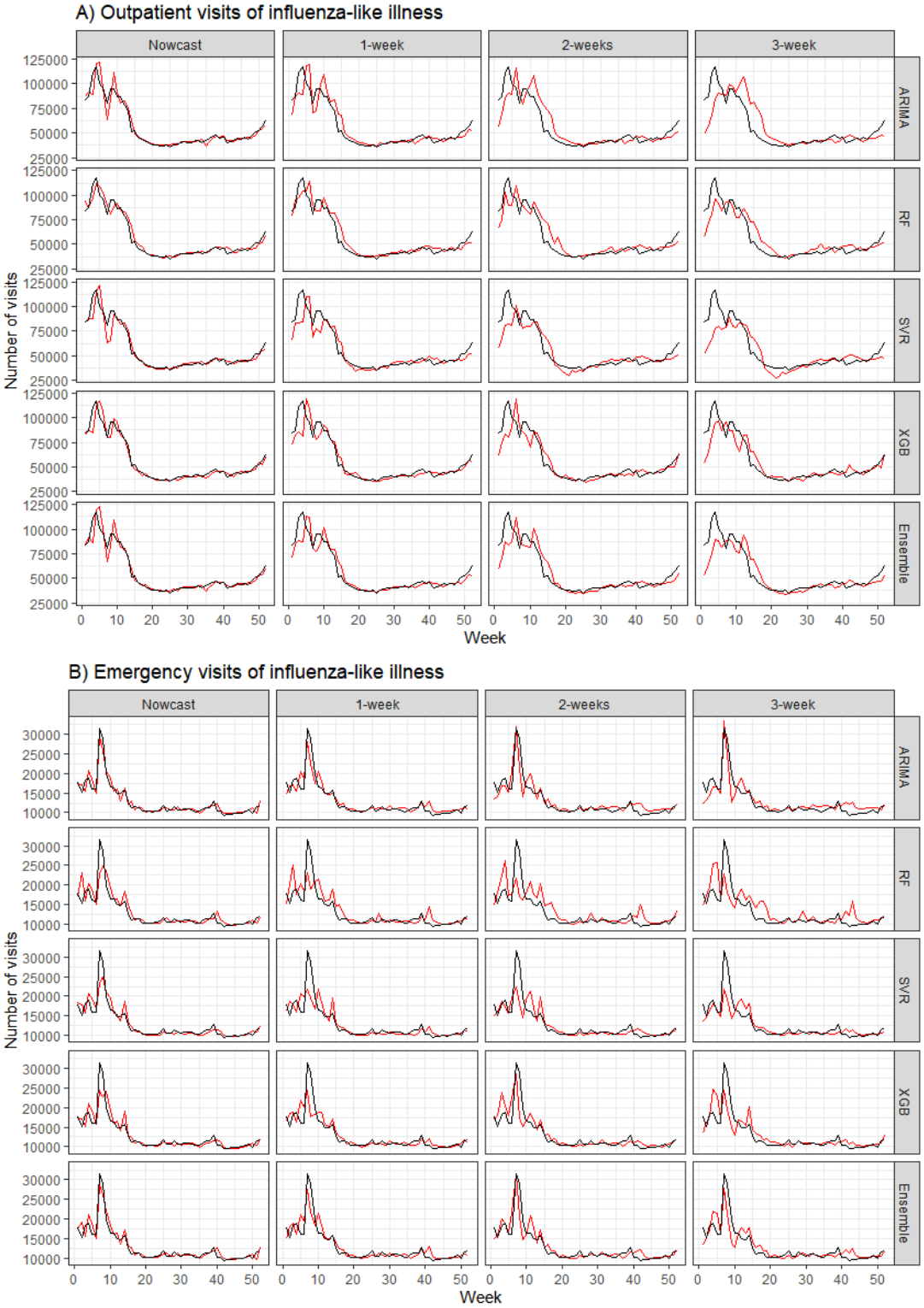
[c]ARIMA: autoregressive integrated moving average.

[d]RF: random forest.

[e]SVR: support vector regression.

[f]XGB: extreme gradient boosting.

Figure 5. Forecasts of the 5 machine learning models (red) and the observed number (black) of influenza-like illness visits in (A) outpatient and (B) emergency departments, 2018. ARIMA: autoregressive integrated moving average; MAPE: mean absolute percentage error; RF: random forest; RMSE: root mean squared error; SVR: support vector regression; XGB: extreme gradient boosting.



Visualization Dashboard of Forecasts

To easily compare the predictions of the 5 models, we created a visualization dashboard website to display the projections concurrently (Multimedia Appendix 6). We also provided the MAPEs and hit rates of all the models that were calculated using the recent 4-, 8-, and 52-week data. In this manner, policy makers could also consider accuracy when evaluating predictions.

Discussion

Principal Results

By using the influenza surveillance data from Taiwan CDC, we established a forecasting model framework that comprises 4 machine learning models and one ensemble model. Our ensemble approach and the framework of model training could provide highly precise forecasts of weekly influenza-like illness visits for a 4-week period. Then-current week forecasts (nowcasts) were the most accurate with MAPE as low as approximately 5% and hit rates of approximately 0.75. Because of the satisfactory hit rate, the change in the influenza-like illness visits in the forecasts for the 4-week period could be regarded as the estimated temporal trend forecasts of a future epidemic as well.

A comparison with models developed in other countries or areas revealed that our models could provide better accuracy with a very low MAPE, which was less than 10% in nowcasts and remained below 20% in the 4-week period forecasts. These results outperform previously-reported models with MAPE mostly greater than 15%-20% [11,25,35,36], suggesting that our models can provide promising predictivity of short-term forecasts on the epidemic magnitude; however, it is difficult to directly compare the performance among different algorithms developed in varying clinical settings and with varying data quality. As for short-term forecasting of the epidemic magnitude in Taiwan, an MAPE less than 10%, especially during the peak time, would be helpful when policymakers need to evaluate the required surging capacity. Because the weekly change of influenza-like illness visits is usually less than 10%-15% in Taiwan, an MAPE greater than 15%-20% might not reliably catch the shift in the epidemic.

The high accuracy of our models might be attributed to the comprehensive data set that we used [5]. The high coverage and good representativeness of the Real-Time Outbreak and Disease Surveillance System and National Health Insurance database allowed the forecasts to more accurately reflect the trends and magnitude of influenza-like illness without being affected by the bias that would have been caused by incompletely sampled data. Conversely, previous models in the US and other countries mostly relied on the sentinel surveillance system such as the influenza-like illness net in the US, which was mainly composed of volunteered sentinel clinics and had problems pertaining to completeness and representativeness [37-40]; the predictivity of the models might have been significantly impaired when they were trained using imprecise historical data. Researchers usually develop an algorithm using a specific period of historical data and then use the trained algorithm with newly-collected data to forecast; therefore, forecasting performance can become worse and worse over time and require periodic adjustment.

In contrast, with our method, models can be retrained every week using updated data. In this way, the algorithms learn from the updated data and maintain satisfactory performance even after being used by the Taiwan CDC for more than one year. In addition, our ensemble model was adapted from the stacking method and could summarize the forecasting outputs from the 4 basic machine learning algorithms with appropriate weighting [31-33]. The aim of our ensemble model was not to build the most accurate forecasting model for any given time. Since the 4 models select features independently from our data sources and had different forecasting performance in real-world applications, for example, ARIMA was usually a lagging indicator of the peak in the influenza season, while random forest and extreme gradient boosting predicted the peak better but tended to underestimate the magnitudes at the peak; therefore, by combining the forecasts from ARIMA and extreme gradient boosting model, the ensemble approach could overcome the disadvantages of each individual model and generate the most robust forecasts with stable performance.

In addition to completeness and representativeness, the Real-Time Outbreak and Disease Surveillance System and National Health Insurance database provided excellent timeliness for our forecasting models. Thanks to the nearly real-time data exchange of the Real-Time Outbreak and Disease Surveillance System and National Health Insurance database with a time lag of, at the most, 1-2 days [5], we could use influenza-like illness data from the previous week at the beginning the week and generate forecasts for a 4-week period that started every Tuesday, for any given week. Because of the delay in the collection of surveillance data, models developed in other countries usually acquire data with at least 1-2 weeks of delay. Thus, their 1-week forecast generated using historical data up until 2 weeks prior is actually the prediction for the previous week [11,25,40-42]. Compared to those models [43], the aforementioned short time delay made our forecast model, which can generate the forecast of a given week (nowcast), a real *real-time forecasting* model. This information can be of great help to the authorities for decision making concerning epidemic preparedness and interventions.

In order to resolve the timeliness problem of the influenza-like illness surveillance data, researchers have attempted to explore the use of social media data (such as Twitter and Facebook) or internet search data (such as Google search and Google Flu Trends) to develop forecasting models because these data can be collected in almost real-time [11,22,26,36,41,44]. However, the method of data collection, quality of social media data, and accuracy of the models still posed problems [22,26,36,41]. In our framework, we did not include social media data because of the following reasons. First, ideal sources of social media information have not yet been established in Taiwan. The largest social media website in Taiwan is Facebook. Still, it is rarely used in social media surveillance because of the hindrances in collecting personal posts from individual profiles (personal walls). A microblog such as Twitter is less prevalent in Taiwan netizens. Second, as for web search data, the Taiwan CDC conducts a regular weekly press release and usually causes a higher amount of search for the related terms on the day of the press release. For example, the searches for influenza significantly increase when an influenza-related news article is released. Therefore, it is difficult to determine whether the increase in the number of web searches, epidemic-related news, or social media discussions is due to an increase influenza-like illness visits or the effect on the media of the official press release. Conversely, access to medical service in Taiwan is easy, and our surveillance systems, such as the Real-Time Outbreak

and Disease Surveillance System and the National Health Insurance database, have already collected highly comprehensive data. Thus, we do not need to rely upon the use of social media as a supplementary data source for disease forecasting, especially for influenza-like illness.

For the models based on traditional frequency statistics such as ARIMA, it is relatively easy to produce a 95% confidence interval, but it is not similarly easy for random forest, support vector regression, and extreme gradient boosting models. Although some literature discussed how to generate prediction intervals for machine learning models like random forest, it is not practical to display 5 intervals on one chart simultaneously. Too much information only confuses the user and makes it difficult to interpret the trend from the 5 forecasts. As we introduced 4 forecasting models based on different algorithms, they already provide and demonstrate the variations in forecasts. When we combine these forecasts with the most robust forecasts from the ensemble model, decision makers can easily get an impression of the forecast without ignoring the potential outlier at one time. Thus, this framework can provide similar information to that provided by confidence intervals.

The models used were machine learning models, which were different from traditional mechanistic models and those such as the susceptible-infectious-recovered model [45,46]. The susceptible-infectious-recovered model considers the dynamics of infectious disease and other biological components. For example, a researcher could create a compartment to simulate the interaction dynamics between infected and immunized people to estimate the effects of vaccination. However, these models are usually built on the basis of historical data and are useful in evaluating the relationship between the different compartments. This characteristic makes such a model better for assessing the effectiveness of vaccination or other interventions on disease transmission, but poor in making future prediction since it is difficult to extrapolate the results because of unknown data at forecasting [43]. For example, when building a susceptible-infectious-recovered-V model, including the compartment V as vaccinated, we need to enter the possible number of vaccinated people in the near future if we want to use this model for forecasting.

Limitations

There are some limitations to our forecasting models. First, the predictivity of our models decreased with longer time horizons, and the best hit rate was only approximately 0.75, suggesting that our models are better at predicting the epidemic magnitude but not the trend. However, we could calibrate the forecasts by learning from the experience of the real-world application. For example, compared with traditional time-series models, such as the ARIMA model, we found that the random forest and support vector regression models may better predict the epidemic dynamics, when the models were applied in 2018. By combining the forecasts and human judgment, the decision-making process for future epidemics can be further ameliorated. Second, using other new deep learning algorithms, especially those with promising performance in time-series forecasting tasks, such as a recurrent neural network and long short-term memory networks, may help to improve the forecasting accuracy. Unlike sequential learning, we retrained the models from zero with mostly updated data every week to manage the time factor better. Our model is only designed for short-term forecasts not for the long-term epidemic change. A deep learning algorithm may be able to deal with this type of forecasting task. Further studies using other algorithms on different forecasting targets, such as the start of a seasonal influenza outbreak and its peak time, are still required to be able to provide more information.

Conclusions

Our project demonstrated real-time short-term forecasting models on weekly influenza-like illness visits using comprehensive influenza surveillance data. By using an ensemble approach to aggregate the forecasts of 4 machine learning models, we could provide accurate predictions for a 4-week period (nowcast and forecasts for the subsequent 3 weeks) to enhance epidemic preparedness.

Acknowledgments

We would like to thank the Acer Value Lab for providing technical support for this project (Fluforecast) and the data visualization website.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The description of data source for the machine learning algorithms.
[DOCX File , 18 KB - jmir_v22i8e15394_app1.docx]

Multimedia Appendix 2
The feature set and engineering used in machine learning algorithms.
[DOCX File , 18 KB - jmir_v22i8e15394_app2.docx]

Multimedia Appendix 3

The evaluation metrics for the algorithm performance.
[DOCX File , 20 KB - jmir_v22i8e15394_app3.docx]

Multimedia Appendix 4
Hyperparameters used in the machine learning models.
[DOCX File , 17 KB - jmir_v22i8e15394_app4.docx]

Multimedia Appendix 5
The trends of forecasting accuracy with increasing forecasting time horizons by difference evaluation metrics.
[PNG File , 105 KB - jmir_v22i8e15394_app5.png]

Multimedia Appendix 6
Visualization dashboard of the forecasts produced by the five machine learning models on the Fluforecast website.
[PNG File , 525 KB - jmir_v22i8e15394_app6.png]

References

1. Yang M, Tan EC, Su J. Cost-effectiveness analysis of quadrivalent versus trivalent influenza vaccine in Taiwan: A lifetime multi-cohort model. Hum Vaccin Immunother 2017 Jan 02;13(1):81-89 [FREE Full text] [doi: 10.1080/21645515.2016.1225636] [Medline: 27624648]
2. Yang J, Hsu S, Kuo C, Huang H, Huang T, Wang H, et al. An epidemic surge of influenza A(H3N2) virus at the end of the 2016-2017 season in Taiwan with an increased viral genetic heterogeneity. J Clin Virol 2018;99-100:15-21. [doi: 10.1016/j.jcv.2017.12.012] [Medline: 29278832]
3. Gong Y, Kuo R, Chen G, Shih S. Centennial review of influenza in Taiwan. Biomedical Journal 2018 Aug;41(4):234-241. [doi: 10.1016/j.bj.2018.08.002] [Medline: 30348266]
4. Su C, Tsou T, Chen C, Lin T, Chang S, Influenza Control Group, Infectious Disease Control Advisory Committee. Seasonal influenza prevention and control in Taiwan-Strategies revisited. J Formos Med Assoc 2019 Mar;118(3):657-663 [FREE Full text] [doi: 10.1016/j.jfma.2018.12.022] [Medline: 30648551]
5. Jian S, Chen C, Lee C, Liu D. Real-time surveillance of infectious diseases: Taiwan's experience. Health Secur 2017;15(2):144-153 [FREE Full text] [doi: 10.1089/hs.2016.0107] [Medline: 28418738]
6. Chuang J, Huang AS, Huang W, Liu M, Chou J, Chang F, et al. Nationwide surveillance of influenza during the pandemic (2009–10) and post-pandemic (2010–11) periods in Taiwan. PLoS ONE 2012 Apr 24;7(4):e36120. [doi: 10.1371/journal.pone.0036120] [Medline: 22545158]
7. Lipsitch M, Finelli L, Heffernan RT, Leung GM, Redd SC, 2009 H1n1 Surveillance Group. Improving the evidence base for decision making during a pandemic: the example of 2009 influenza A/H1N1. Biosecur Bioterror 2011 Jun;9(2):89-115 [FREE Full text] [doi: 10.1089/bsp.2011.0007] [Medline: 21612363]
8. Hillmer SC, Wei WWS. Time series analysis: univariate and multivariate methods. Journal of the American Statistical Association 1991 Mar;86(413):245. [doi: 10.2307/2289741]
9. Nsoesie EO, Brownstein JS, Ramakrishnan N, Marathe MV. A systematic review of studies on forecasting the dynamics of influenza outbreaks. Influenza Other Respir Viruses 2014 May;8(3):309-316 [FREE Full text] [doi: 10.1111/irv.12226] [Medline: 24373466]
10. Paul S, Mgbere O, Arafat R, Yang B, Santos E. Modeling and forecasting influenza-like illness (ILI) in Houston, Texas using three surveillance data capture mechanisms. Online J Public Health Inform 2017;9(2):e187 [FREE Full text] [doi: 10.5210/ojphi.v9i2.8004] [Medline: 29026453]
11. Xu Q, Gel YR, Ramirez LLR, Nezafati K, Zhang Q, Tsui K. Forecasting influenza in Hong Kong with Google search queries and statistical model fusion. PLoS One 2017;12(5):e0176690 [FREE Full text] [doi: 10.1371/journal.pone.0176690] [Medline: 28464015]
12. Feng H, Duan G, Zhang R, Zhang W. Time series analysis of hand-foot-mouth disease hospitalization in Zhengzhou: establishment of forecasting models using climate variables as predictors. PLoS One 2014;9(1):e87916 [FREE Full text] [doi: 10.1371/journal.pone.0087916] [Medline: 24498221]
13. Gharbi M, Quenel P, Gustave J, Cassadou S, La Ruche G, Girdary L, et al. Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. BMC Infect Dis 2011 Jun 09;11:166 [FREE Full text] [doi: 10.1186/1471-2334-11-166] [Medline: 21658238]
14. Kane MJ, Price N, Scotch M, Rabinowitz P. Comparison of ARIMA and random forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinformatics 2014 Aug 13;15:276 [FREE Full text] [doi: 10.1186/1471-2105-15-276] [Medline: 25123979]
15. Dmlc/Xgboost. GitHub. 2020. URL: https://github.com/dmlc/xgboost/blob/master/demo/README.md#usecases [accessed 2020-07-20]

16. Chen S, Xu J, Chen L, Zhang X, Zhang L, Li J. A regularization-based extreme gradient boosting approach in foodborne disease trend forecasting. Stud Health Technol Inform 2019 Aug 21;264:930-934. [doi: 10.3233/SHTI190360] [Medline: 31438060]
17. Petukhova T, Ojkic D, McEwen B, Deardon R, Poljak Z. Assessment of autoregressive integrated moving average (ARIMA), generalized linear autoregressive moving average (GLARMA), and random forest (RF) time series regression models for predicting influenza A virus frequency in swine in Ontario, Canada. PLoS One 2018;13(6):e0198313 [FREE Full text] [doi: 10.1371/journal.pone.0198313] [Medline: 29856881]
18. Bühlmann P, Yu B. Boosting with the L2 Loss. Journal of the American Statistical Association 2003 Jun;98(462):324-339. [doi: 10.1198/016214503000125]
19. Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and Machine Learning forecasting methods: concerns and ways forward. PLoS One 2018;13(3):e0194889 [FREE Full text] [doi: 10.1371/journal.pone.0194889] [Medline: 29584784]
20. Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing 2004 Aug;14(3):199-222. [doi: 10.1023/B:STCO.0000035301.49549.88]
21. Freeze J, Erraguntla M, Verma A. Data Integration and Predictive Analysis System for Disease Prophylaxis: Incorporating Dengue Fever Forecasts. 2018 Jan 03 Presented at: Hawaii International Conference on System Sciences; 2018/01/03; Hawaii. [doi: 10.24251/HICSS.2018.114]
22. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
23. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu Near You: crowdsourced symptom reporting spanning 2 influenza seasons. Am J Public Health 2015 Oct;105(10):2124-2130. [doi: 10.2105/AJPH.2015.302696] [Medline: 26270299]
24. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. PLoS One 2013;8(12):e83672 [FREE Full text] [doi: 10.1371/journal.pone.0083672] [Medline: 24349542]
25. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Comput Biol 2015 Oct;11(10):e1004513 [FREE Full text] [doi: 10.1371/journal.pcbi.1004513] [Medline: 26513245]
26. Liang F, Guan P, Wu W, Huang D. Forecasting influenza epidemics by integrating internet search queries and traditional surveillance data with the support vector machine regression model in Liaoning, from 2011 to 2015. PeerJ 2018;6:e5134 [FREE Full text] [doi: 10.7717/peerj.5134] [Medline: 29967755]
27. Alessa A, Faezipour M. A review of influenza detection and prediction through social networking sites. Theor Biol Med Model 2018 Dec 01;15(1):2 [FREE Full text] [doi: 10.1186/s12976-017-0074-5] [Medline: 29386017]
28. Poirier C, Lavenu A, Bertaud V, Campillo-Gimenez B, Chazard E, Cuggia M, et al. Real time influenza monitoring using hospital big data in combination with machine learning methods: comparison study. JMIR Public Health Surveill 2018 Dec 21;4(4):e11361 [FREE Full text] [doi: 10.2196/11361] [Medline: 30578212]
29. Ertem Z, Raymond D, Meyers LA. Optimal multi-source forecasting of seasonal influenza. PLoS Comput Biol 2018 Sep;14(9):e1006236 [FREE Full text] [doi: 10.1371/journal.pcbi.1006236] [Medline: 30180212]
30. Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, et al. Developing a dengue forecast model using machine learning: A case study in China. PLoS Negl Trop Dis 2017 Oct;11(10):e0005973 [FREE Full text] [doi: 10.1371/journal.pntd.0005973] [Medline: 29036169]
31. Reich NG, McGowan CJ, Yamana TK, Tushar A, Ray EL, Osthus D, et al. Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. PLoS Comput Biol 2019 Nov;15(11):e1007486 [FREE Full text] [doi: 10.1371/journal.pcbi.1007486] [Medline: 31756193]
32. Kandula S, Yamana T, Pei S, Yang W, Morita H, Shaman J. Evaluation of mechanistic and statistical methods in forecasting influenza-like illness. J R Soc Interface 2018 Jul;15(144):20180174 [FREE Full text] [doi: 10.1098/rsif.2018.0174] [Medline: 30045889]
33. Wolpert DH. Stacked generalization. Neural Networks 1992 Jan;5(2):241-259. [doi: 10.1016/s0893-6080(05)80023-1]
34. Bostock M, Ogievetsky V, Heer J. D^3 data-driven documents. IEEE Trans Visual Comput Graphics 2011 Dec;17(12):2301-2309. [doi: 10.1109/tvcg.2011.185]
35. Yang S, Santillana M, Brownstein JS, Gray J, Richardson S, Kou SC. Using electronic health records and internet search information for accurate influenza forecasting. BMC Infect Dis 2017 May 08;17(1):332 [FREE Full text] [doi: 10.1186/s12879-017-2424-7] [Medline: 28482810]
36. Lu FS, Hou S, Baltrusaitis K, Shah M, Leskovec J, Sosic R, et al. Accurate influenza monitoring and forecasting using novel internet data streams: a case study in the Boston metropolis. JMIR Public Health Surveill 2018 Jan 09;4(1):e4 [FREE Full text] [doi: 10.2196/publichealth.8950] [Medline: 29317382]
37. Clothier H, Turner J, Hampson A, Kelly H. Geographic representativeness for sentinel influenza surveillance: implications for routine surveillance and pandemic preparedness. Aust N Z J Public Health 2006 Aug;30(4):337-341. [doi: 10.1111/j.1467-842x.2006.tb00846.x] [Medline: 16956163]

38. Polgreen PM, Chen Z, Segre AM, Harris ML, Pentella MA, Rushton G. Optimizing influenza sentinel surveillance at the state level. Am J Epidemiol 2009 Nov 15;170(10):1300-1306 [FREE Full text] [doi: 10.1093/aje/kwp270] [Medline: 19822570]
39. Scarpino SV, Dimitrov NB, Meyers LA. Optimizing provider recruitment for influenza surveillance networks. PLoS Comput Biol 2012;8(4):e1002472 [FREE Full text] [doi: 10.1371/journal.pcbi.1002472] [Medline: 22511860]
40. Greenspan J. Preparing for ILINet 2.0. Online J Public Health Inform 2015 Feb 26;7(1):e25. [doi: 10.5210/ojphi.v7i1.5690]
41. Woo H, Cho Y, Shim E, Lee J, Lee C, Kim SH. Estimating influenza outbreaks using both search engine query data and social media data in South Korea. J Med Internet Res 2016 Jul 04;18(7):e177 [FREE Full text] [doi: 10.2196/jmir.4955] [Medline: 27377323]
42. Basile L, Oviedo de la Fuente M, Torner N, Martínez A, Jané M. Real-time predictive seasonal influenza model in Catalonia, Spain. PLoS One 2018;13(3):e0193651 [FREE Full text] [doi: 10.1371/journal.pone.0193651] [Medline: 29513710]
43. Reich NG, Brooks LC, Fox SJ, Kandula S, McGowan CJ, Moore E, et al. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. Proc Natl Acad Sci U S A 2019 Feb 19;116(8):3146-3154 [FREE Full text] [doi: 10.1073/pnas.1812594116] [Medline: 30647115]
44. Lu FS, Hattab MW, Clemente CL, Biggerstaff M, Santillana M. Improved state-level influenza nowcasting in the United States leveraging internet-based data and network approaches. Nat Commun 2019 Jan 11;10(1):147 [FREE Full text] [doi: 10.1038/s41467-018-08082-0] [Medline: 30635558]
45. Yang W, Karspeck A, Shaman J. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. PLoS Comput Biol 2014 Apr;10(4):e1003583 [FREE Full text] [doi: 10.1371/journal.pcbi.1003583] [Medline: 24762780]
46. Yang W, Lipsitch M, Shaman J. Inference of seasonal and pandemic influenza transmission dynamics. Proc Natl Acad Sci U S A 2015 Mar 03;112(9):2723-2728 [FREE Full text] [doi: 10.1073/pnas.1415012112] [Medline: 25730851]

Abbreviations

ARIMA: autoregressive integrated moving average model
MAPE: mean absolute percentage error
RMSE: root mean squared error

Edited by G Eysenbach; submitted 08.07.19; peer-reviewed by K Dhindsa, A Nguyen, E Frontoni; comments to author 21.10.19; revised version received 21.12.19; accepted 13.06.20; published 05.08.20.

Please cite as:
Cheng HY, Wu YC, Lin MH, Liu YL, Tsai YY, Wu JH, Pan KH, Ke CJ, Chen CM, Liu DP, Lin IF, Chuang JH
Applying Machine Learning Models with An Ensemble Approach for Accurate Real-Time Influenza Forecasting in Taiwan: Development and Validation Study
J Med Internet Res 2020;22(8):e15394
URL: https://www.jmir.org/2020/8/e15394
doi:10.2196/15394
PMID:32755888

Original Paper

A Conceptual Framework to Study the Implementation of Clinical Decision Support Systems (BEAR): Literature Review and Concept Mapping

Jhon Camacho[1,2], BEng, MSc; Manuela Zanoletti-Mannello[3], MD; Zach Landis-Lewis[4], MLIS, PhD; Sandra L Kane-Gill[5], MS, PharmD, FCCP, FCCM; Richard D Boyce[1], PhD

[1]Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[2]I&E Meaningful Research, Bogotá, Colombia

[3]School of Medicine, Pontificia Universidad Javeriana, Bogotá, Colombia

[4]Department of Learning Health Sciences, University of Michigan, Ann Arbor, MI, United States

[5]Department of Pharmacy and Therapeutics, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Jhon Camacho, BEng, MSc
I&E Meaningful Research
Calle 119a #70-31
Bogotá, 111121
Colombia
Phone: 57 315 695 5484
Email: jjcamachosanchez@gmail.com

Abstract

Background: The implementation of clinical decision support systems (CDSSs) as an intervention to foster clinical practice change is affected by many factors. Key factors include those associated with behavioral change and those associated with technology acceptance. However, the literature regarding these subjects is fragmented and originates from two traditionally separate disciplines: implementation science and technology acceptance.

Objective: Our objective is to propose an integrated framework that bridges the gap between the behavioral change and technology acceptance aspects of the implementation of CDSSs.

Methods: We employed an iterative process to map constructs from four contributing frameworks—the Theoretical Domains Framework (TDF); the Consolidated Framework for Implementation Research (CFIR); the Human, Organization, and Technology-fit framework (HOT-fit); and the Unified Theory of Acceptance and Use of Technology (UTAUT)—and the findings of 10 literature reviews, identified through a systematic review of reviews approach.

Results: The resulting framework comprises 22 domains: agreement with the decision algorithm; attitudes; behavioral regulation; beliefs about capabilities; beliefs about consequences; contingencies; demographic characteristics; effort expectancy; emotions; environmental context and resources; goals; intentions; intervention characteristics; knowledge; memory, attention, and decision processes; patient–health professional relationship; patient's preferences; performance expectancy; role and identity; skills, ability, and competence; social influences; and system quality. We demonstrate the use of the framework providing examples from two research projects.

Conclusions: We proposed BEAR (BEhavior and Acceptance fRamework), an integrated framework that bridges the gap between behavioral change and technology acceptance, thereby widening the view established by current models.

(J Med Internet Res 2020;22(8):e18388) doi:10.2196/18388

KEYWORDS

clinical decision support system; computerized decision support system; implementation science; technology acceptance; barriers; facilitators; determinants; decision support system

Introduction

Every year, significant amounts of resources are invested in medical research globally, an average of 0.19% of the gross domestic product in high-income countries [1]. All this effort has resulted in the exponential growth of scientific evidence. However, the translation of that knowledge into changes in clinical practice is advancing at a much lower rate, creating a growing knowledge-practice gap [2]. Reducing this gap requires not only the development and dissemination of evidence-based guidelines but also the integration of guideline recommendations into care processes and that health professionals change their practice. Clinical decision support systems (CDSSs) present a promising approach to address these challenges [3-6].

CDSSs encode clinical knowledge into computerized algorithms and combine them with patient-specific data to provide clinicians with information and decision guidance [7]. When successfully implemented, the ability of a CDSS to provide patient-specific decision support empowers health professionals to make timely decisions at the point of care while reducing medical errors [8,9]. Another benefit of this technology is that the transformation of clinical knowledge into algorithms allows for the correction of areas where documents (eg, clinical practice guidelines) are ambiguous or unclear [10-13]. CDSSs have been implemented to support care in several specialties [14-20], both in developed and developing countries [14,18,21].

Although several literature reviews have shown improvements in process measures after the implementation of CDSSs [8,22-24], the evidence of their effectiveness on clinical outcomes is still mixed [22,25,26]. This is partially explained because successful implementation of practice-change interventions is a multidimensional problem requiring attention to many factors [27-30]. Relevant factors include not only those internal and external to the health organization [29] but also those related to the clinicians' preferences and their mental model about their practice [28,30]. Additionally, CDSSs must be integrated into the clinical workflow and be accepted by the users.

We hypothesize that improving the implementation of CDSSs requires attention to factors related to practice change and also to those associated with technology acceptance, defined as the user's decision to use a technology system routinely [31]. Theories and frameworks about these topics can be found in the research fields of implementation science and technology acceptance. However, though drawing from similar sources (ie, psychology, sociology, and management science), these fields have developed into separate disciplines. Therefore, a researcher considering studying the implementation of a CDSS as a strategy to foster clinical change is confronted with a fragmented corpus of knowledge and the choice among conceptual frameworks, potentially missing or having to give up the contributions from one of these fields.

Another issue is that there are competing frameworks within the fields of implementation science and technology acceptance. Several authors have proposed theoretical models to explain the determinants of clinical practice change [32]. Similarly, several models attempt to explain the factors influencing the user's decision to use a technology system routinely [33,34]. Furthermore, several recent studies have identified determinants that are specific to the acceptance of CDSSs [35-40]. When seen together, these frameworks amount to too many concepts for a reasonable research project to use effectively.

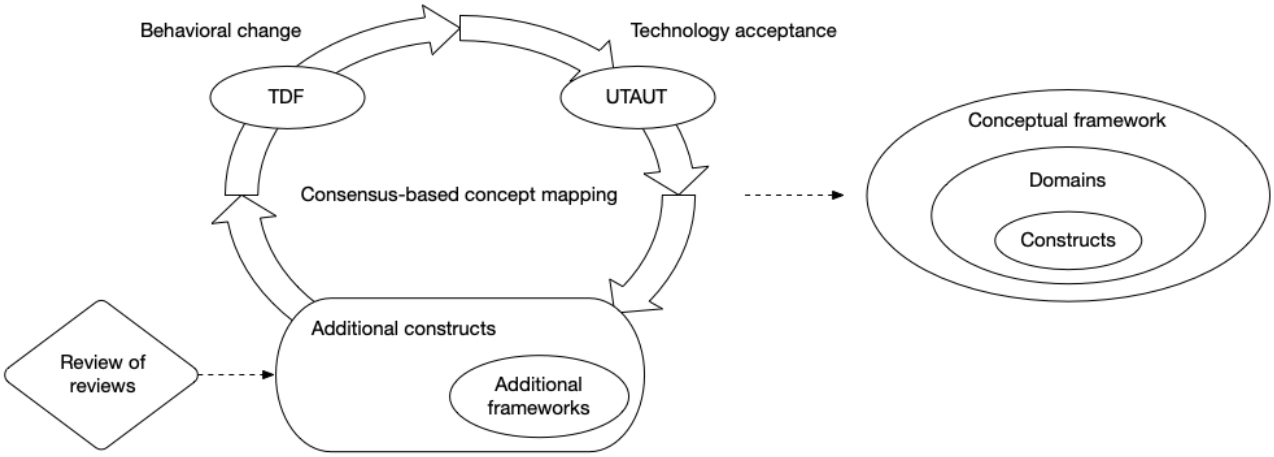
To address these issues, we propose BEAR (BEhavior and Acceptance fRamework), an integrated conceptual framework that bridges the gap between behavioral change and technology acceptance aspects of the implementation of CDSSs. BEAR synthesizes literature about factors influencing both practice change and acceptance of CDSSs from the health professional's perspective. Furthermore, BEAR seeks to capture the variability in the phenomenon of implementing CDSSs while providing an integrated tool that facilitates the design of research and evaluation projects.

Methods

Overview

We developed BEAR by employing an iterative process in which two investigators (JC and MZM) mapped constructs reported in the literature as determinants of behavioral change and acceptance of CDSSs (see Figure 1). In each iteration, both investigators mapped the constructs from a framework or a literature review into the emerging construct pool, starting with the Theoretical Domains Framework (TDF) [28,30] and the Unified Theory of Acceptance and Use of Technology (UTAUT) [34]. At the beginning of each iteration, the two investigators (JC and MZM) developed maps independently contrasting the information in the articles with the definitions in the construct pool. After that, the investigators discussed differences in their maps and agreed on modifications to the pool. These modifications encompassed the following: the inclusion of new constructs, in addition to agreeing to their definitions; changes in construct labels or definitions; and changes in the grouping of constructs into domains. The emerging framework was progressively documented in two files: one contained the definitions and the other contained the map to the original constructs (see Multimedia Appendices 1 and 2).

Figure 1. Framework development. TDF: Theoretical Domains Framework; UTAUT: Unified Theory of Acceptance and Use of Technology.



Base Frameworks

The first iteration comprised the mapping of two well-established frameworks: the TDF [28,30] and the UTAUT [34]. The TDF, proposed by Michie et al in 2005 [30] and revised by Cane et al in 2012 [28], comprises 84 theoretical constructs included in classic psychological theories about behavior change. The UTAUT, on the other hand, was proposed by Venkatesh et al in 2003 [34] to integrate the eight predominant models at the time about technology acceptance. The UTAUT comprises eight constructs that influence the regular use of a technology, directly or indirectly. The selection of these frameworks as a starting point was guided by the authors' previous experiences with the evaluation of medical informatics interventions.

Literature Review

Subsequent iterations comprised the mapping of constructs identified through a literature review of recent aggregative studies presenting determinants of the acceptance of CDSSs. To obtain the initial pool of references, we queried Scopus, MEDLINE (Medical Literature Analysis and Retrieval System Online), Embase, CINAHL (Cumulative Index of Nursing and Allied Health Literature), and PsycINFO. Figure 2 presents the search strategy used in Scopus. We constructed equivalent searches for the other databases.

Figure 2. Search strategy used in Scopus.

```
(TITLE-ABS-KEY("Decision support")) AND
(
   TITLE-ABS-KEY("Determinants") OR TITLE-ABS-KEY("Barriers") OR
   TITLE-ABS-KEY("Facilitators") OR TITLE-ABS-KEY("Factors") OR
   TITLE-ABS-KEY("Drivers") OR TITLE-ABS-KEY("Mediators") OR
   TITLE-ABS-KEY("Reasons")
) AND
(
   TITLE-ABS-KEY("Acceptance") OR TITLE-ABS-KEY("Adoption") OR
   TITLE-ABS-KEY("Uptake") OR TITLE-ABS-KEY("use")
) AND
(TITLE("Review") OR DOCTYPE(re)
) AND
(PUBYEAR > 2014))
```

Two investigators (JC and MZM) screened the initial reference pool and evaluated their titles and abstracts. For a reference to be selected, it needed to fulfill all of the following inclusion criteria: (1) address the topic of CDSSs, (2) employ literature review methods to obtain its results, and (3) include, among its results, determinants of the acceptance of CDSSs. Exclusion criteria were limited to the following: (1) the publication was not a research article (eg, abstract, viewpoint, commentary, editorial, or protocol) and (2) the full text was not in English or Spanish.

We limited the search strategy to articles published since 2014; we were working under the assumption that, although recent, the aggregate studies found would cover the relevant literature published before that year. That assumption was validated by documenting the period covered by each included review.

The investigators (JC and MZM) then extracted the constructs and their definitions from the full text of each article (see Multimedia Appendix 1). In cases where definitions or descriptions were not explicitly stated, the investigators reviewed the full text of the cited articles, including

supplemental materials. Furthermore, in cases where other frameworks were used to organize the review findings, those frameworks were included in the mapping exercise with their own iterations.

Domain Structure and Refinement

The grouping of constructs into domains was initially informed by the organization and definitions in the base frameworks (ie, TDF and UTAUT). Later on, new constructs identified from the literature reviews were contrasted with the domain definitions to choose their locations. In some cases, this process resulted in the creation of new domains or in changing previous definitions.

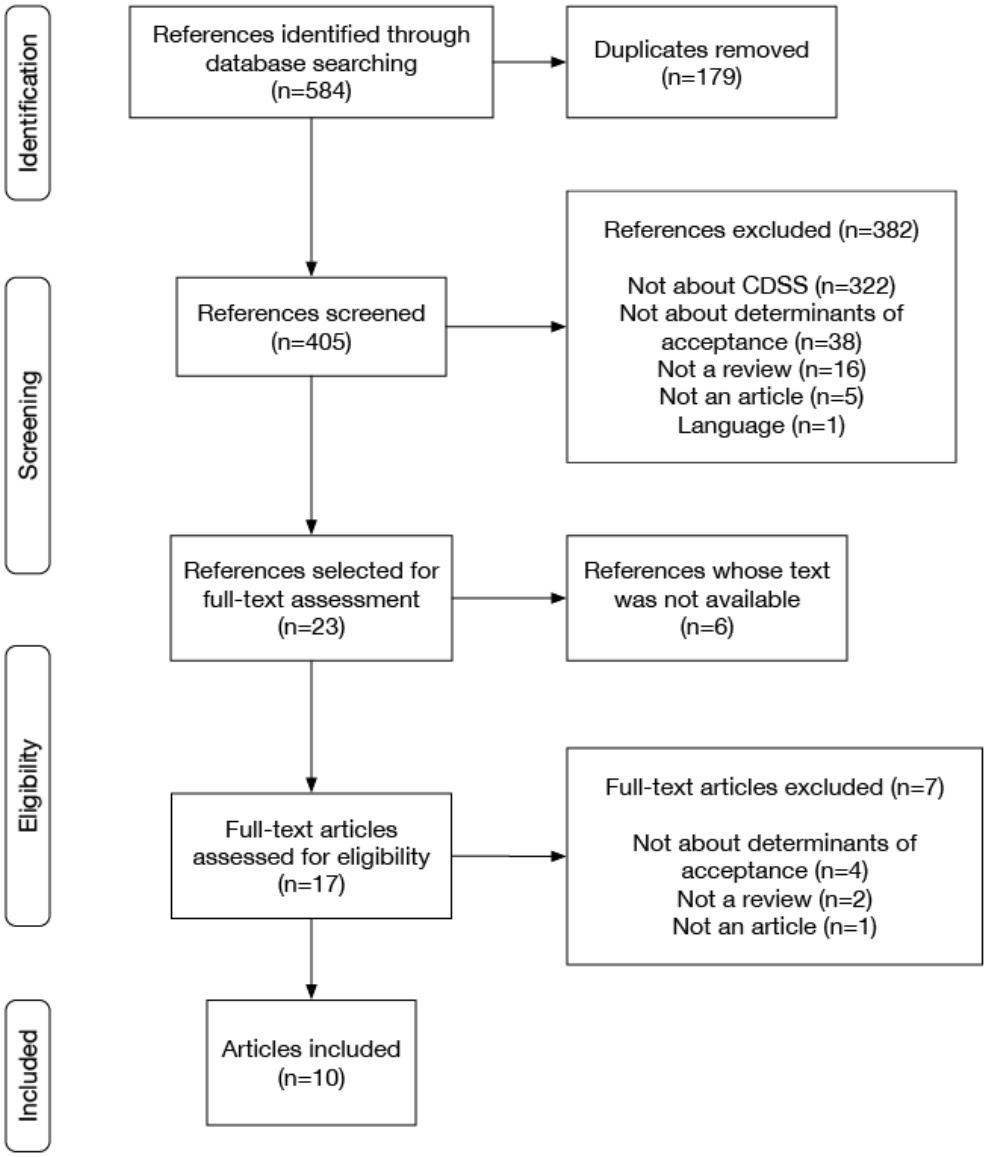
Finally, preliminary versions of the framework were discussed with the other authors (ZLL, SLKG, and RDB) and other colleagues, resulting in the refinement of construct labels, definitions, and grouping.

Results

Overview

The mapping process comprised 13 iterations, corresponding to 10 reviews and four frameworks. The initial literature search identified 584 references. After removing duplicates, 405 references passed through the screening process (see Figure 3). Of these, 23 were selected for full-text assessment; however, text from six of these articles was not available. After reviewing the 17 available articles, seven were excluded. Finally, 10 articles met all inclusion criteria [35-40,42,45]. Only one article was excluded based on language. Table 1 shows the characteristics of the reviews included in this study and any associated frameworks [29,35-45].

Figure 3. Flow diagram of the literature review. CDSS: clinical decision support system.



Included Reviews

The 10 reviews included in the mapping exercise (see Table 1) synthesize 219 studies from 1995 to 2018; these studies include participants from several populations of health professionals (ie, nurses, general practitioners, specialists, pharmacists, residents, laboratory technicians, physical therapists, medical

assistants, medical students, paramedics, psychologists, and social workers). Additionally, two frameworks were included: the CFIR (Consolidated Framework for Implementation Research) [29], which was used in Ross et al [37], and the HOT-fit (Human, Organization, and Technology-fit) framework [41], which was used in Kilsdonk et al [38] and Van Dort et al [44].

Table 1. Characteristics of included reviews.

Source	Number of references	Time span	Participants	Framework
Khong et al, 2015 [35]	16	2005-2014	Nurses, general practitioners, specialists, pharmacists, and medical assistants	N/A[a]
Khairat et al, 2018 [36]	14	1995-2015	Nurses, general practitioners, specialists, residents, and medical students	N/A
Ross et al, 2016 [37]	44	2002-2014	Nurses, general practitioners, specialists, laboratory technicians, physical therapists, paramedics, medical students, residents, pharmacists, and social workers	CFIR[b] [29]
Kilsdonk et al, 2017 [38]	35	2003-2015	Nurses, general practitioners, specialists, physical therapists, medical students, residents, pharmacists, and psychologists	HOT-fit[c] [41]
Miller et al, 2017 [39]	14	2003-2015	Nurses, general practitioners, specialists, pharmacists, medical assistants, and residents	N/A
Borum, 2018 [40]	9	2011-2016	Nurses, general practitioners, specialists, and pharmacists	N/A
Baig et al, 2019 [42]	22	2014-2016	Nurses, general practitioners, and specialists	N/A
Carter et al, 2019 [43]	13	2014-2017	Nurses, midwives, nurse students, specialists, and community health workers	N/A
Van Dort et al, 2019 [44]	13	2009-2018	General practitioners and specialists	HOT-fit [41]
Hussain et al, 2019 [45]	39	2008-2017	Nurses, general practitioners, and specialists	N/A

[a]N/A: not applicable.

[b]CFIR: Consolidated Framework for Implementation Research.

[c]HOT-fit: Human, Organization, and Technology-fit.

BEAR

Table 2 presents the constructs and domains included in the proposed framework [28,34,46-50]. The mapping of each construct into the sources is included in Multimedia Appendix 2. BEAR comprises 156 constructs arranged into 22 domains. Domain definitions are included in Table 2, whereas definitions for each construct are included in Multimedia Appendix 1.

Table 2. BEAR (BEhavior and Acceptance fRamework) constructs and domains.

Domain[a]	Domain definition	Constructs[b]
Knowledge	Awareness, understanding, or information about a subject that has been obtained by experience or study: based on [46]	Knowledge Knowledge of task environment Procedural knowledge Knowledge of the decision algorithm Knowledge of the patient's condition Previous experience with decision support technology
Skills, ability, and competence	An ability or proficiency acquired through training and practice [28]: based on [47]	Skills, ability, and competence Computer and mobile device skill Interpersonal skills Skills development
Role and identity	A coherent set of behaviors and displayed personal qualities of an individual in a social or work setting [28]: based on [47]	Individual identity Professional identity Organizational commitment Professional boundaries Professional role Professional autonomy
Beliefs about capabilities	Acceptance of the truth, reality, or validity about an ability, talent, or facility that a person can put to constructive use [28]: based on [47]	Beliefs about capabilities Empowerment Perceived behavioral control Professional confidence Self-confidence Self-efficacy Self-esteem
Beliefs about consequences	Acceptance of the truth, reality, or validity about outcomes of a behavior in a given situation [28]: based on [47]	Beliefs about consequences Anticipated regret Outcome expectancies Beliefs that technology would disrupt the delivery of care Characteristics of outcome expectancies Concerns about liability and responsibility Concerns over patient privacy
Attitudes	Relatively enduring and general evaluations of an object, person, group, issue, or concept on a dimension ranging from negative to positive. Attitudes provide summary evaluations of target objects and are often assumed to be derived from specific beliefs, emotions, and past behaviors associated with those objects [48].	Attitudes Interest in technology Perceived uselessness Optimism Pessimism Unrealistic optimism Attitude toward practice guidelines
Contingencies	A conditional probabilistic relationship between two events. Contingencies may be arranged via dependencies or they may emerge by accident [28]: citing [47].	Contingencies Consequences Reinforcement Incentives Punishment Rewards Sanctions

Domain[a]	Domain definition	Constructs[b]
Intentions	A conscious decision to perform a behavior; a resolve to act in a certain way or an impulse for purposeful action. In experiments, intention is often equated with goals defined by the task instruction [28]: citing [47].	Intentions Stability of intentions Stages of change—precontemplation Stages of change—contemplation Stages of change—preparation Stages of change—action Stages of change—maintenance
Goals	Mental representations of outcomes or end states that an individual wants to achieve [28]: based on [47]	Goals Goals—level of control (autonomous vs controlled) Goals—temporality (distal vs proximal) Target setting Goal priority Action planning Change plan
Memory, attention, and decision processes	The ability to retain information, focus selectively on aspects of the environment, and choose between two or more alternatives [28]: based on [47]	Memory Attention Attention control Decision process Cognitive overload and tiredness
Environmental context and resources	Any circumstance of a person's situation or environment that discourages or encourages the development of skills and abilities, independence, social competence, and adaptive behavior [28]: based on [47]	Environmental context Resources Environmental stressors Organizational structure Organizational culture and climate Assessment—skills Assessment—knowledge Assessment—performance Person × environment interaction Salient events and critical incidents Time availability—patient care Time availability—learning Technical support Technical infrastructure Facilities Implementation climate Tension for change Access to information and knowledge about the intervention
Social influences	Those interpersonal processes that can cause individuals to change their thoughts, feelings, or behaviors [28]: based on [47] The degree to which an individual perceives that others important to him or her believe he or she should use the new system [34]	Social influences Alienation Group conformity Group identity Group norms Leadership Intergroup conflict Modelling Power Social comparisons Social norms Social pressure Social support

Domain[a]	Domain definition	Constructs[b]
Emotions	A complex reaction pattern, involving experiential, behavioral, and physiological elements, by which the individual attempts to deal with a personally significant matter or event [28]: based on [47]	Emotions Affect Positive affect Negative affect Anxiety Burnout Depression Apprehension Fear Stress Frustration Uncertainty Dissatisfaction
Behavioral regulation	Anything aimed at managing or changing objectively observed or measured actions [28]: based on [47]	Behavioral regulation Breaking habit Self-monitoring
Intervention characteristics	Intervention attributes that facilitate or hinder its implementation. The intervention includes not only the system but also all processes and resources needed to deploy it.	Intervention characteristics Intervention source Adaptability Trialability Interoperability Implementation complexity Costs—initial Costs—recurrent Voluntariness of use
Performance expectancy	The degree to which an individual believes that using the system will help him or her to attain gains in job performance [34]	Performance expectancy Benefits for the patient Improved communication with other health professionals Improved access to knowledge Consistency of care Error prevention Time-saving Habituation
Effort expectancy	The effort an individual believes is required to implement or use the system	Effort expectancy Quality of the user interface Compatibility with the clinical workflow Access at the point of care Familiarization
Demographic characteristics	The characteristics of people who form a particular group, with reference to distribution, composition, or structure: based on [46,49]	Demographic characteristics Age Gender Professional experience Training level and educational level Nationality

Domain[a]	Domain definition	Constructs[b]
System quality	The degree to which the information and functions provided by the system meet the user's needs or expectations and give user satisfaction; the degree to which the system is free from deficiencies or defects: based on [50]	System quality System performance Output quality Output quality—accuracy Output quality—completeness Output quality—specificity Output quality—timeliness System reliability
Agreement with the decision algorithm	The degree to which the user agrees that the decision algorithm is a correct way to make the intended decision	Agreement with the decision algorithm Applicability to complex cases Evidence strength and quality
Patient–health professional relationship	The way the system affects the relationship between the health professional and the patient	Patient–health professional relationship Obtrusiveness Diminished eye contact Disruption of flow in conversation with the patient Knowledgeable image
Patient's preferences	The way the patient's preferences affect the health professional's decision about using the system	Patient's preferences Patient's decision not to follow the recommendation

[a]The way we include references in this column seeks to help the reader trace back the origin of each definition. In cases where we use the same text from the source (ie, a textual citation), we only include the reference number. In cases where the source text was adapted, we precede the reference number with the phrase "based on." In cases where the source is citing another source, we include a reference for the latter, preceded by the word "citing." Finally, definitions without a reference were developed by the authors.

[b]Construct definitions are included in Multimedia Appendix 1.

Discussion

Principal Findings

Our objective was to develop a framework, grounded in the literature about determinants of behavioral change and technology acceptance, that would be useful to researchers investigating the implementation of CDSSs as a strategy to foster the uptake of evidence-based recommendations.

Developing Strategy and Structure

The idea of BEAR originated in our search for a conceptual framework to guide our research in the use of CDSSs as a strategy to implement clinical practice guidelines. From the beginning, we realized that the effectiveness of such an approach would be mediated by aspects of behavioral change and technology acceptance. We found part of the guidance we were looking for in the TDF [28,30] and the UTAUT [34]. These frameworks provide constructs that address both aspects of the phenomenon, although at a higher level than the one we were seeking, particularly on the side of technology acceptance. For example, UTAUT includes the concept of *facilitating conditions*, defined as "the degree to which an individual believes that an organizational and technical infrastructure exists to support the use of the system" [34]. However, that definition is not enough to identify what specific facilitating conditions are missing, which we consider a necessary step for the development of interventions. This need for further detail is what led us to an iterative process by which we incorporated the findings of recent literature reviews about determinants of the acceptance of CDSSs.

To capture the variability in the phenomenon of implementing CDSSs, we sought to include constructs that are specific enough to facilitate the identification of what is different between one implementation experience and another. For example, we added constructs to represent four quality aspects for the system's output: accuracy, completeness, specificity, and timeliness. However, recognizing that it is unlikely that we have identified every relevant concept, we also included general constructs, in this case *output quality*. Along the same lines, most domains include a construct that shares the domain's label. In cases where a domain label corresponds to a group of concepts (eg, *memory, attention, and decision processes*), we do not include a general construct but only those representing each constituting concept.

The decision to include both specific and general constructs led us to make two more decisions about the framework's structure: (1) to have only one grouping level (ie, domain and construct) and (2) to include each construct only once, inside the domain where, in our opinion, the construct fits better. With these decisions, we sought to control complexity while maintaining detail. The resulting domains sort the constructs thematically; that is, constructs included in a domain represent determinants that could influence the concept represented by the domain, instead of particularizations of that concept.

Use Cases

BEAR is not a parsimonious framework. We believe this is both a strength and a limitation. On the one hand, we expect that the level of detail facilitates the identification of actionable determinants; on the other hand, using the whole framework could be difficult, particularly in quantitative-oriented projects.

However, for most projects, using every construct in the framework is not necessary or even advisable.

Our recommendation is to use the framework at the domain level during the initial stages of research design, particularly when discussing scope. Later, in qualitative-oriented studies, BEAR could be utilized to develop data collection guides for interviews, focus groups, or observations. This could be done at the domain level, in the case of exploratory studies, or based on selected constructs. During analysis, BEAR could serve as an initial coding schema, either at the domain or construct levels, raising the researcher's awareness of determinants and supporting the identification of categories.

In quantitative-oriented studies, besides informing decisions about scope and research questions, BEAR could support the search for theories and measurement tools. In both cases, we recommend reviewing the definitions and references provided in Multimedia Appendix 1.

In the next section, we present two examples of how we have used BEAR in our research.

Use Example: A CDSS to Support Chronic Obstructive Pulmonary Disease Active Case-Finding

We are currently using BEAR in a qualitative-oriented project whose objective is to identify barriers and facilitators to the use of a CDSS to support the implementation of a case-finding recommendation included in the Colombian chronic obstructive pulmonary disease (COPD) clinical practice guideline.

According to the recommendation, suspicion of airflow obstruction could be established by the identification of specific risk factors and symptoms. Once the suspicion is established, spirometry should be ordered to confirm the limitation in the airflow [51].

In the study, we explained the recommendation to primary-level physicians and asked them to use a CDSS, implemented as a mobile app, during their patient encounters. The CDSS asked a series of questions about the patient's clinical history and current symptoms. Using this information, the system applied a decision algorithm to establish the suspicion of COPD, in which case it recommended that the participant order a spirometry test to discard or confirm the diagnosis.

After 2 months using the system, we interviewed the participants to explore their experiences applying the recommendation and using the system. The guide used in these interviews was developed using BEAR at the domain level. Table 3 presents selected questions included in the guide. Designing the interview around BEAR's domains allowed us to explore both the behavioral change and the technology acceptance aspects of the intervention.

Table 3. Example 1 questions.

Domain	Questions
Knowledge	Before this project, did you know about this recommendation?
	Given the information you had before, and what we have given you in the project, do you consider that you have all the information you need to carry out the screening?
Role and identity	Do you consider that screening for COPD[a] cases is part of the primary care physician's responsibility or should it be assigned to someone else?
Performance expectancy	Was the app useful in the process of implementing the recommendation?
	How did you use it?
Agreement with the decision algorithm	Can you think of anything that the ministry could change in the content of the recommendation to make it easier to meet the goal of detecting COPD cases early?

[a]COPD: chronic obstructive pulmonary disease.

Data collected in the interviews were analyzed thematically using BEAR's constructs as the initial coding schema. During the initial analysis, the transcript below—adapted from the original data in Spanish—was coded under the following constructs: *patient–health professional relationship, diminished eye-contact*, and *patient's preferences*.

> *...nowadays, we hardly see the patient, we are always [gestures representing the use of a keyboard]. We are all the time writing in the health record...In fact, some patients get upset. They complain that I do not look them in the eye. I try to look at them while writing in the computer, but I don't have the ability yet.* [Participant]
>
> *Does this mean that they got upset when you used the application on your phone?* [Interviewer]
>
> *No, because I tell them, "Look, I am going to use this app to help in the diagnosis," and I show them my phone.* [Participant]

Later in the analysis, the review of the content in these codes led to development of two categories: *perceived loss of attention* and *negotiating the use of the device with the patient*. The first category refers to the way the patient seems to interpret and resent that he has lost the physician's attention when the latter is using the computer. The second category refers to the way physicians prevent complaints about themselves when using the phone during the encounter by telling the patients what they are using the phone for and sometimes including them in the process of using the system. The relationship between these categories allows us to recognize differential effects over the *patient–health professional relationship* of CDSSs implemented as mobile apps and as desktop applications.

Use Example: Clinician Responsiveness to the CDSS in Clinical Practice

Alert fatigue is a common problem for clinicians who use technology designed to improve patient safety. Evidence-based strategies to overcome alert fatigue are lacking, especially in the intensive care unit (ICU). There is an evidence gap, as discussed in a recent review and guidance document [52].

The goal of the project in this use example is to provide effective strategies for the management of alert fatigue in the ICU. The behavioral change of interest is increasing clinician responsiveness to CDSS alerts provided during patient care (eg, ordering medications). Formative research needs to be completed to understand the barriers and facilitators to clinicians' responsiveness to alerts. To meet these goals, a mixed methods approach was applied using a survey and in-depth interviews conducted with critical care clinicians. Questions were developed based on BEAR at the domain level. A sample of selected questions is provided in Table 4.

Table 4. Example 2 questions.

Domain	Questions
Skills, ability, and competence	Do you feel competent to respond to the alerts you are receiving?
Beliefs about consequences	What do you think will happen if you do not respond to alerts?
Social influences	How responsive are your peers to alerts?
Emotions	How frequently does receiving an alert lead to an evoked emotional response?
Behavioral regulation	What would encourage you to be more responsive to alerts?
Performance expectancy	To what extent are the alerts useful?
Effort expectancy	How easy is it to respond to the alerts?

BEAR in Relation to Other Frameworks

Out of 122 constructs identified in the literature review, 52 (42.6%) mapped to the TDF (see Multimedia Appendix 2). This supports our initial assumption that, in the context of CDSSs, behavioral change and technology acceptance are interrelated. Since the TDF was selected as a source from the beginning, it is not a surprise that both its constructs and structure had a substantial influence on the resulting framework. However, the TDF's constructs emerged from the mapping process with some modifications. In some cases, these changes corresponded to the integration of constructs (eg, the TDF's *skills*, *ability*, and *competence*, whose definitions in the American Psychological Association Dictionary of Psychology [48] are similar). In other cases, we changed the construct definition to facilitate its interpretation in the context of clinical practice change. For instance, the TDF's definition for modelling—"In developmental psychology the process in which one or more individuals or other entities serve as examples (models) that a child will copy" [28]—was changed to "The process in which one or more individuals or other entities serve as examples (models) for a person to copy." We believe these alterations do not substantially change the meaning of the affected TDF constructs, but rather improve their applicability.

The CFIR [29] and HOT-fit framework [41] were also part of the mapping process. However, several of their constructs were not included in the resulting framework due to differences in the scope. Whereas BEAR deals with behavioral change and technology acceptance from the individual's perspective, the CFIR considers the implementation as a whole, integrating other perspectives (ie, those related to the organization, the government and health system, and the implementation project [29]). In some cases, those perspectives intersected. For example, the CFIR's *inner setting—implementation climate* construct represents an organizational characteristic that influences the individual's behavior. However, in other cases (eg, the CFIR's *process—reflecting and evaluating* construct), we did not identify a direct influence over the individual. The same happened with the HOT-fit framework's *organization* domain [41]. We recognize that the level of influence of a particular construct over the individual's behavior could be a matter of debate—indeed, we had several discussions about it during the mapping process—thus, we used our better judgment. For information about the mapping of specific constructs, we refer the reader to Multimedia Appendix 2.

This is certainly not the first attempt to apply technology acceptance models in health care [31,53]. The majority of these attempts have tried to adapt the Technology Acceptance Model (TAM) [33], one of UTAUT's eight contributing frameworks [34]. BEAR has similarities and differences with these works. On the one hand, BEAR attempts to cover a wide range of possible determinants, but it does not make statements about the magnitude of their influence on each other or the individual's behavior. In other words, BEAR does not attempt to state a theory about technology acceptance. Instead, BEAR is meant as an exploratory tool that allows for the identification of determinants that could be articulated into hypotheses and potentially form the basis of interventions. We hope that the study of those hypotheses in different health care contexts results in a future theory that is able to explain and predict practice change in the context of CDSSs.

On the other hand, literature reviews have found that the TAM has low explanatory power in health care environments [31,53]. The authors of these studies attribute this lack of fit to professional differences between health professionals and other workers [53] and to the fact that the TAM does not completely incorporate the emotional and cultural aspects of health care decision making [31]. Our interpretation of these findings is that behavioral change determinants operate differently in health care in comparison to other work environments. If that is correct,

it could be expected that bridging the gap between behavioral change and technology acceptance brings forward the missing pieces in the puzzle.

Finally, our objective of this study was to develop a conceptual framework, not a theory. A theory serves as an explanation of a phenomenon that in many cases allows for the prediction of outcomes. However, our objective was not to predict an outcome or the relative weight of each determinant in the explanation of an outcome, but to synthesize and organize all potential determinants reported in the reviewed literature. That is why we do not state any conclusion about the relationship between specific constructs, besides grouping them into domains to facilitate organization and presentation, nor their relative contribution to the success in the implementation of CDSSs. In this sense, BEAR is akin to other determinant frameworks [54], such as the TDF [28,30] or the CFIR [29], rather than a theory, such as the UTAUT [34].

Acknowledgments

This research was funded in part by the US National Institute on Aging (K01AG044433). We thank Socorro Moreno, Ps, MSc, and María Fernanda Parra, Ps, for their comments and insights regarding preliminary versions of the framework.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions of constructs in BEAR (BEhavior and Acceptance fRamework).
[DOC File , 224 KB - jmir_v22i8e18388_app1.doc]

Multimedia Appendix 2
Mapping of constructs in BEAR (BEhavior and Acceptance fRamework).
[XLS File (Microsoft Excel File), 379 KB - jmir_v22i8e18388_app2.xls]

References

1. Gross domestic R&D expenditure on health (health GERD) as a % of gross domestic product (GDP). World Health Organization. 2020 Jan. URL: https://www.who.int/research-observatory/indicators/gerd_gdp/en/ [accessed 2019-11-05]
2. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: Understanding time lags in translational research. J R Soc Med 2011 Dec;104(12):510-520 [FREE Full text] [doi: 10.1258/jrsm.2011.110180] [Medline: 22179294]
3. Pearson S, Moxey A, Robertson J, Hains I, Williamson M, Reeve J, et al. Do computerised clinical decision support systems for prescribing change practice? A systematic review of the literature (1990-2007). BMC Health Serv Res 2009 Aug 28;9:154 [FREE Full text] [doi: 10.1186/1472-6963-9-154] [Medline: 19715591]
4. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. JAMA 2005 Mar 09;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
5. Murphy EV. Clinical decision support: Effectiveness in improving quality processes and clinical outcomes and factors that may influence success. Yale J Biol Med 2014 Jun;87(2):187-197 [FREE Full text] [Medline: 24910564]
6. Holroyd BR, Bullard MJ, Graham TA, Rowe BH. Decision support technology in knowledge translation. Acad Emerg Med 2007 Nov;14(11):942-948 [FREE Full text] [doi: 10.1197/j.aem.2007.06.023] [Medline: 17766733]
7. Eberhardt J, Bilchik A, Stojadinovic A. Clinical decision support systems: Potential with pitfalls. J Surg Oncol 2012 Apr 01;105(5):502-510. [doi: 10.1002/jso.23053] [Medline: 22441903]
8. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. BMJ 2005 Apr 02;330(7494):765 [FREE Full text] [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]
9. Reis WC, Bonetti AF, Bottacin WE, Reis AS, Souza TT, Pontarolo R, et al. Impact on process results of clinical decision support systems (CDSSs) applied to medication use: Overview of systematic reviews. Pharm Pract (Granada) 2017;15(4):1036 [FREE Full text] [doi: 10.18549/PharmPract.2017.04.1036] [Medline: 29317919]
10. Codish S, Shiffman R. A model of ambiguity and vagueness in clinical practice guideline recommendations. In: Proceedings of the American Medical Informatics Association (AMIA) Annual Symposium. 2005 Presented at: American Medical Informatics Association (AMIA) Annual Symposium; October 22-26, 2005; Washington, DC p. 146-150.
11. Lee PY, Liew SM, Abdullah A, Abdullah N, Ng CJ, Hanafi NS, et al. Healthcare professionals' and policy makers' views on implementing a clinical practice guideline of hypertension management: A qualitative study. PLoS One 2015;10(5):e0126191 [FREE Full text] [doi: 10.1371/journal.pone.0126191] [Medline: 25942686]

12. Solà I, Carrasco JM, Díaz Del Campo P, Gracia J, Orrego C, Martínez F, et al. Attitudes and perceptions about clinical guidelines: A qualitative study with Spanish physicians. PLoS One 2014;9(2):e86065 [FREE Full text] [doi: 10.1371/journal.pone.0086065] [Medline: 24505253]
13. Peleg M, González-Ferrer A. Guidelines and workflow models. In: Greenes RA, editor. Clinical Decision Support: The Road to Broad Adoption. 2nd edition. London, UK: Academic Press; 2014:435-464.
14. Blank A, Prytherch H, Kaltschmidt J, Krings A, Sukums F, Mensah N, et al. "Quality of prenatal and maternal care: Bridging the know-do gap" (QUALMAT study): An electronic clinical decision support system for rural Sub-Saharan Africa. BMC Med Inform Decis Mak 2013 Apr 10;13:44 [FREE Full text] [doi: 10.1186/1472-6947-13-44] [Medline: 23574764]
15. Moja L, Kwag KH, Lytras T, Bertizzolo L, Brandt L, Pecoraro V, et al. Effectiveness of computerized decision support systems linked to electronic health records: A systematic review and meta-analysis. Am J Public Health 2014 Dec;104(12):e12-e22. [doi: 10.2105/AJPH.2014.302164] [Medline: 25322302]
16. Cox JL, Parkash R, Abidi SS, Thabane L, Xie F, MacKillop J, IMPACT-AF Investigators. Optimizing primary care management of atrial fibrillation: The rationale and methods of the Integrated Management Program Advancing Community Treatment of Atrial Fibrillation (IMPACT-AF) study. Am Heart J 2018 Jul;201:149-157 [FREE Full text] [doi: 10.1016/j.ahj.2018.04.008] [Medline: 29807323]
17. Vecchio N, Belardi D, Benzadón M, Seoane L, Daquarti G, Scazzuso F. Impact of a hybrid decision support system to improve the adherence to implantable cardioverter defibrillator therapy guidelines for primary prevention of sudden cardiac death. Int J Med Inform 2018 Jun;114:76-80. [doi: 10.1016/j.ijmedinf.2018.03.016] [Medline: 29673607]
18. Shao AF, Rambaud-Althaus C, Samaka J, Faustine AF, Perri-Moore S, Swai N, et al. New algorithm for managing childhood illness using mobile technology (ALMANACH): A controlled non-inferiority study on clinical outcome and antibiotic use in Tanzania. PLoS One 2015;10(7):e0132316 [FREE Full text] [doi: 10.1371/journal.pone.0132316] [Medline: 26161535]
19. Dehghani Soufi M, Samad-Soltani T, Shams Vahdati S, Rezaei-Hachesu P. Decision support system for triage management: A hybrid approach using rule-based reasoning and fuzzy logic. Int J Med Inform 2018 Jun;114:35-44. [doi: 10.1016/j.ijmedinf.2018.03.008] [Medline: 29673601]
20. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M, Sainz-de-Abajo B, Robles M, García-Gómez JM. Mobile clinical decision support systems and applications: A literature and commercial review. J Med Syst 2014 Jan;38(1):4. [doi: 10.1007/s10916-013-0004-y] [Medline: 24399281]
21. O'Reilly-Shah VN, Kitzman J, Jabaley CS, Lynde GC. Evidence for increased use of the Society of Pediatric Anesthesia Critical Events Checklist in resource-limited environments: A retrospective observational study of app data. Paediatr Anaesth 2018 Feb;28(2):167-173. [doi: 10.1111/pan.13305] [Medline: 29285834]
22. Souza NM, Sebaldt RJ, Mackay JA, Prorok JC, Weise-Kelly L, Navarro T, CCDSS Systematic Review Team. Computerized clinical decision support systems for primary preventive care: A decision-maker-researcher partnership systematic review of effects on process of care and patient outcomes. Implement Sci 2011 Aug 03;6:87 [FREE Full text] [doi: 10.1186/1748-5908-6-87] [Medline: 21824381]
23. Mickan S, Atherton H, Roberts NW, Heneghan C, Tilson JK. Use of handheld computers in clinical practice: A systematic review. BMC Med Inform Decis Mak 2014 Jul 06;14:56 [FREE Full text] [doi: 10.1186/1472-6947-14-56] [Medline: 24998515]
24. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: A systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
25. Chan WV, Pearson TA, Bennett GC, Cushman WC, Gaziano TA, Gorman PN, et al. ACC/AHA special report: Clinical practice guideline implementation strategies: A summary of systematic reviews by the NHLBI Implementation Science Work Group: A report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. Circulation 2017 Feb 28;135(9):e122-e137. [doi: 10.1161/CIR.0000000000000481] [Medline: 28126839]
26. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. mHealth adoption by health care professionals: A systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
27. Lau R, Stevenson F, Ong BN, Dziedzic K, Treweek S, Eldridge S, et al. Achieving change in primary care--causes of the evidence to practice gap: Systematic reviews of reviews. Implement Sci 2016 Mar 22;11:40 [FREE Full text] [doi: 10.1186/s13012-016-0396-4] [Medline: 27001107]
28. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
29. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
30. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: A consensus approach. Qual Saf Health Care 2005 Feb;14(1):26-33 [FREE Full text] [doi: 10.1136/qshc.2004.011155] [Medline: 15692000]

31. Ward R. The application of technology acceptance and diffusion of innovation models in healthcare informatics. Health Policy Technol 2013 Dec;2(4):222-228. [doi: 10.1016/j.hlpt.2013.07.002]
32. Tabak RG, Khoong EC, Chambers DA, Brownson RC. Bridging research and practice: Models for dissemination and implementation research. Am J Prev Med 2012 Sep;43(3):337-350 [FREE Full text] [doi: 10.1016/j.amepre.2012.05.024] [Medline: 22898128]
33. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: A comparison of two theoretical models. Manage Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
34. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
35. Khong PCB, Holroyd E, Wang W. A critical review of the theoretical frameworks and the conceptual factors in the adoption of clinical decision support systems. Comput Inform Nurs 2015 Dec;33(12):555-570. [doi: 10.1097/CIN.0000000000000196] [Medline: 26535769]
36. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: Critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
37. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of eHealth: A systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
38. Kilsdonk E, Peute L, Jaspers M. Factors influencing implementation success of guideline-based clinical decision support systems: A systematic review and gaps analysis. Int J Med Inform 2017 Feb;98:56-64. [doi: 10.1016/j.ijmedinf.2016.12.001] [Medline: 28034413]
39. Miller K, Mosby D, Capan M, Kowalski R, Ratwani R, Noaiseh Y, et al. Interface, information, interaction: A narrative review of design and functional requirements for clinical decision support. J Am Med Inform Assoc 2018 May 01;25(5):585-592 [FREE Full text] [doi: 10.1093/jamia/ocx118] [Medline: 29126196]
40. Borum C. Barriers for hospital-based nurse practitioners utilizing clinical decision support systems: A systematic review. Comput Inform Nurs 2018 Apr;36(4):177-182. [doi: 10.1097/CIN.0000000000000413] [Medline: 29360699]
41. Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for health information systems: Human, organization and technology-fit factors (HOT-fit). Int J Med Inform 2008 Jun;77(6):386-398. [doi: 10.1016/j.ijmedinf.2007.08.011] [Medline: 17964851]
42. Baig MM, GholamHosseini H, Moqeem AA, Mirza F, Lindén M. Clinical decision support systems in hospital care using ubiquitous devices: Current issues and challenges. Health Informatics J 2019 Sep;25(3):1091-1104. [doi: 10.1177/1460458217740722] [Medline: 29148314]
43. Carter J, Sandall J, Shennan AH, Tribe RM. Mobile phone apps for clinical decision support in pregnancy: A scoping review. BMC Med Inform Decis Mak 2019 Nov 12;19(1):219 [FREE Full text] [doi: 10.1186/s12911-019-0954-1] [Medline: 31718627]
44. Van Dort BA, Zheng WY, Baysari MT. Prescriber perceptions of medication-related computerized decision support systems in hospitals: A synthesis of qualitative research. Int J Med Inform 2019 Sep;129:285-295. [doi: 10.1016/j.ijmedinf.2019.06.024] [Medline: 31445268]
45. Hussain M, Reynolds T, Zheng K. Medication safety alert fatigue may be reduced via interaction design and clinical role tailoring: A systematic review. J Am Med Inform Assoc 2019 Oct 01;26(10):1141-1149 [FREE Full text] [doi: 10.1093/jamia/ocz095] [Medline: 31206159]
46. Cambridge Dictionary. URL: https://dictionary.cambridge.org/ [accessed 2019-04-26]
47. VandenBos GR, editor. APA Dictionary of Psychology. Washington, DC: American Psychological Association; 2007.
48. APA Dictionary of Psychology. Washington, DC: American Psychological Association URL: https://dictionary.apa.org/ [accessed 2019-04-23]
49. Medical Subject Headings. National Library of Medicine. URL: https://www.nlm.nih.gov/mesh/meshhome.html [accessed 2019-04-27]
50. Akrani G. What is product quality? Definition. Meaning. Importance. Kalyan City Life Blog. 2013 May 21. URL: https://kalyan-city.blogspot.com/2013/05/what-is-product-quality-definition.html [accessed 2019-04-29]
51. Londoño D, García O, Celis C, Giraldo M, Casas A, Torres C, et al. Guía de práctica clínica basada en la evidencia para la prevención, diagnóstico, tratamiento y seguimiento de la enfermedad pulmonar obstructiva crónica (EPOC) en población adulta. Acta Médica Colombiana. Volume 39 No 2 (Supl 3). 2014. URL: http://www.actamedicacolombiana.com/anexo/articulos/2014/02S3-2014-00.pdf [accessed 2020-05-22]
52. Kane-Gill SL, O'Connor MF, Rothschild JM, Selby NM, McLean B, Bonafide CP, et al. Technologic distractions (Part 1): Summary of approaches to manage alert quantity with intent to reduce alert fatigue and suggestions for alert fatigue metrics. Crit Care Med 2017 Sep;45(9):1481-1488. [doi: 10.1097/CCM.0000000000002580] [Medline: 28682835]
53. Holden RJ, Karsh B. The technology acceptance model: Its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
54. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]

Abbreviations

BEAR: BEhavior and Acceptance fRamework
CDSS: clinical decision support system
CFIR: Consolidated Framework for Implementation Research
CINAHL: Cumulative Index of Nursing and Allied Health Literature
COPD: chronic obstructive pulmonary disease
HOT-fit: Human, Organization, and Technology-fit
ICU: intensive care unit
MEDLINE: Medical Literature Analysis and Retrieval System Online
TAM: Technology Acceptance Model
TDF: Theoretical Domains Framework
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by G Eysenbach; submitted 02.03.20; peer-reviewed by S Sarbadhikari, W Pian, S Sabarguna; comments to author 18.03.20; revised version received 11.05.20; accepted 03.06.20; published 06.08.20.

Please cite as:
Camacho J, Zanoletti-Mannello M, Landis-Lewis Z, Kane-Gill SL, Boyce RD
A Conceptual Framework to Study the Implementation of Clinical Decision Support Systems (BEAR): Literature Review and Concept Mapping
J Med Internet Res 2020;22(8):e18388
URL: https://www.jmir.org/2020/8/e18388
doi:10.2196/18388
PMID:32759098

Original Paper

Text Processing for Detection of Fungal Ocular Involvement in Critical Care Patients: Cross-Sectional Study

Sally L Baxter[1,2], MD, MSc; Adam R Klie[3], BS; Bharanidharan Radha Saseendrakumar[4], BE; Gordon Y Ye[2]; Michael Hogarth[2], MD

[1]Viterbi Family Department of Ophthalmology and Shiley Eye Institute, University of California San Diego, La Jolla, CA, United States

[2]Division of Biomedical Informatics, Department of Medicine, University of California San Diego, La Jolla, CA, United States

[3]Bioinformatics and Systems Biology, University of California San Diego, La Jolla, CA, United States

[4]Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, United States

Corresponding Author:
Sally L Baxter, MD, MSc
Viterbi Family Department of Ophthalmology and Shiley Eye Institute
University of California San Diego
9415 Campus Point Dr. MC 0946
La Jolla, CA, 92093
United States
Phone: 1 858 534 8858
Email: s1baxter@health.ucsd.edu

Abstract

Background: Fungal ocular involvement can develop in patients with fungal bloodstream infections and can be vision-threatening. Ocular involvement has become less common in the current era of improved antifungal therapies. Retrospectively determining the prevalence of fungal ocular involvement is important for informing clinical guidelines, such as the need for routine ophthalmologic consultations. However, manual retrospective record review to detect cases is time-consuming.

Objective: This study aimed to determine the prevalence of fungal ocular involvement in a critical care database using both structured and unstructured electronic health record (EHR) data.

Methods: We queried microbiology data from 46,467 critical care patients over 12 years (2000-2012) from the Medical Information Mart for Intensive Care III (MIMIC-III) to identify 265 patients with culture-proven fungemia. For each fungemic patient, demographic data, fungal species present in blood culture, and risk factors for fungemia (eg, presence of indwelling catheters, recent major surgery, diabetes, immunosuppressed status) were ascertained. All structured diagnosis codes and free-text narrative notes associated with each patient's hospitalization were also extracted. Screening for fungal endophthalmitis was performed using two approaches: (1) by querying a wide array of eye- and vision-related diagnosis codes, and (2) by utilizing a custom regular expression pipeline to identify and collate relevant text matches pertaining to fungal ocular involvement. Both approaches were validated using manual record review. The main outcome measure was the documentation of any fungal ocular involvement.

Results: In total, 265 patients had culture-proven fungemia, with Candida albicans (n=114, 43%) and Candida glabrata (n=74, 28%) being the most common fungal species in blood culture. The in-hospital mortality rate was 121 (46%). In total, 7 patients were identified as having eye- or vision-related diagnosis codes, none of whom had fungal endophthalmitis based on record review. There were 26,830 free-text narrative notes associated with these 265 patients. A regular expression pipeline based on relevant terms yielded possible matches in 683 notes from 108 patients. Subsequent manual record review again demonstrated that no patients had fungal ocular involvement. Therefore, the prevalence of fungal ocular involvement in this cohort was 0%.

Conclusions: MIMIC-III contained no cases of ocular involvement among fungemic patients, consistent with prior studies reporting low rates of ocular involvement in fungemia. This study demonstrates an application of natural language processing to expedite the review of narrative notes. This approach is highly relevant for ophthalmology, where diagnoses are often based on physical examination findings that are documented within clinical notes.

(J Med Internet Res 2020;22(8):e18855) doi:10.2196/18855

KEYWORDS

fungemia; fungal endophthalmitis; fungal ocular involvement; electronic health records; diagnosis codes; regular expressions; natural language processing; unstructured data

Introduction

Fungal ocular infection can be highly morbid and vision-threatening. In contrast to bacterial infections, which tend to result from exogenous causes, fungal eye infections more frequently arise endogenously from fungal bloodstream infections [1]. Consequently, they are often diagnosed in inpatient settings, particularly among critical care patients. Although rates of ocular involvement among fungemic patients were reported in early studies to range from 10% to 45% [2-5], most studies in the last two decades have reported rates less than 5% in both adults and children [6-10], a trend attributed to improved systemic antifungal therapies. Although Ghodrasa et al found a slightly higher rate of ocular involvement of 9.2% (22/238) over 5 years, ophthalmic consultation changed management in only 3.8% (9/238) of cases [11]. In short, these studies have shown that a very low percentage of patients with fungemia who are referred for ophthalmological consultation demonstrate ocular involvement.

Prior studies have ascertained positive cases based on manual review of ophthalmic consultation notes, a labor-intensive process, as multiple years of records need to be reviewed given that fungal endophthalmitis is now a relatively rare entity. However, the widespread adoption of electronic health records (EHRs) offers opportunities to facilitate efficient case detection. The most common approach is to utilize structured EHR data in the form of billing or diagnosis codes or other discrete data fields (eg, physiologic measurements, laboratory values, microbiology data). However, structured data comprise a small proportion of the overall data found in EHRs—prior analyses have found that more than 80% of EHR data are unstructured [12]. Natural language processing (NLP) methods have the potential to leverage these unstructured data and are increasingly employed for biomedical applications [13-16]. One impactful application of NLP is information extraction from free-text clinical notes [17-19], which is especially relevant for ophthalmology, where many diagnoses are based on physical examination findings described in free-text notes rather than structured data such as laboratory values. NLP has been applied to extract data on visual acuity [20] and intracameral antibiotic injections and posterior capsular rupture [21]. It has also been used to extract surgical laterality and intraocular lens implant power and model information [22], glaucoma-related characteristics [23], measurements of epithelial defects and stromal infiltrates in microbial keratitis [24], as well as identification of herpes zoster ophthalmicus [25] and pseudoexfoliation [26].

In this study, we evaluated the prevalence of fungal ocular involvement in the Medical Information Mart for Intensive Care III (MIMIC-III), a cohort of over 46,000 critical care patients over 12 years. Our objective was to detect cases of any fungal ocular involvement (including fungal endophthalmitis, vitritis, and chorioretinitis). In addition to the traditional approach of using manual record review of all patients, we also used queries based on discrete diagnosis codes and developed a novel NLP-based approach for rapid detection of relevant free-text clinical notes related to the diagnosis.

Methods

Study Population

The study population consisted of all patients with fungemia in MIMIC-III, a database of patients admitted to critical care units at the Beth Israel Deaconess Medical Center, a tertiary care hospital in Boston, Massachusetts [27]. It includes deidentified data from over 46,000 critical care patients (adults and neonates) from 2001-2012. In addition to structured data elements such as admission information, demographics, laboratory values, diagnosis codes, intervention/procedure codes, microbiology data, medications, and physiologic measurements, MIMIC-III also includes deidentified free-text clinical notes such as provider admission and progress notes, discharge summaries, consultation notes, and free text reports of electrocardiogram and imaging studies [27]. All authors underwent appropriate Health Insurance Portability and Accountability Act (HIPAA) and Collaborative Institutional Training Initiative (CITI) research ethics training and completed the required data use agreements before accessing any data from MIMIC-III. The University of California San Diego Institutional Review Board (IRB) ruled that approval was not required for this study, as it qualified as non-human subjects research. Consent was not obtained, given that participants were not identifiable. The research adhered to the tenets of the Declaration of Helsinki and conformed with all country, federal, and state laws.

The population of interest consisted of all patients with fungal bloodstream infections, as these individuals would typically be referred for ophthalmological evaluation according to current guidelines from the Infectious Disease Society of America and standard practice patterns [28]. We identified these patients via a structured query language (SQL) query of the “MICROBIOLOGY EVENTS” table in MIMIC-III, selecting for all patients with documented positive blood cultures containing fungal organisms. All fungal species were included. This query yielded 265 patients.

Case Detection via Processing of Structured Data

We defined the outcome of interest as the development of fungal ocular involvement during the documented hospitalization involving culture-positive fungemia. First, we screened for fungal ocular involvement using structured data consisting of International Classification of Disease version 9 (ICD-9) diagnosis codes. As the data in MIMIC-III spanned 2001-2012, there were no ICD-10 codes, which were not available until 2015. We obtained ICD-9 codes by querying the “DIAGNOSES_ICD” table in MIMIC-III, which entails hospital-assigned diagnoses for each patient during each hospitalization and joining the codes with the “D_ICD_DIAGNOSES” table, which is a dictionary of ICD-9

codes and descriptions. To account for known limitations of billing and diagnosis codes such as incomplete coding by clinicians and insufficient granularity of codes [29], we included a broad range of diagnosis codes related to endophthalmitis, retinal disorders, chorioretinal inflammations, vitreous disorders, visual disturbances, and blindness and low vision (see Multimedia Appendix 1 for specific ICD-9 codes). Records for patients with relevant diagnosis codes were then reviewed to determine whether there was any documentation of fungal ocular involvement.

Case Detection via Natural Language Processing of Unstructured Data

We also screened for cases of fungal ocular involvement by analyzing unstructured data consisting of free-text notes available in the MIMIC-III database. Given the large volume of clinical notes for the study population (26,830 total notes for 265 patients), we used NLP methods to facilitate note review and identification of cases. We used a regular expression pipeline based on string matching to extract relevant text strings from notes surrounding the term(s) specified in the regular expression. Regular expressions are strings written in a standardized computational grammar that can be used to identify patterns in free text; they offer an effective strategy for many NLP problems involving pattern matching [30]. Regular expressions have been used in several biomedical contexts to extract information such as conditions and medications from free text [31-34]. Here, we used a variety of terms embedded in regular expressions, ranging from specific single phrases (eg, "fungal endophthalmitis") to multiterm expressions (eg, "ophth|opth|eye|ocul|retina|fundus|dilat|endophthalmitis|chorioretinitis|vitritis"). We used a string-matching approach for its simplicity and straightforward implementation. Common misspellings, such as "opth-" instead of "ophth-," were accounted for in the search. In a strategy similar to the one used for the diagnosis code queries, a broad array of terms were included in the regular expressions in order to err on the side of sensitivity, improving the likelihood of case detection, particularly because fungal ocular involvement is a rare condition. We did not include "Candida" in the regular expressions because this term had been removed from the MIMIC-III database as part of de-identification protocols since "Candida" can be a female first name. Therefore, any mentions of "Candida endophthalmitis" would be obscured as "[**Female First Name (un)**] endophthalmitis." However, because terms such as "endophthalmitis," "vitritis," "chorioretinitis," "ocular involvement," and other terms related to the eye and fungal infections were included, our strategy was designed to maximize the likelihood of identifying those cases without explicitly including "Candida" in the regular expressions.

We recorded the number of text strings resulting from each pattern-matching script as well as the run time for each script to produce an output. Text strings and subsequently, full clinical notes were manually reviewed to determine whether fungal ocular involvement was documented. The Python package *re* was used as the regular expression interpreter [35]. Open-source coding details are available on GitHub [36].

Manual Record Review

A full manual record review was performed for all 265 fungemic patients in this cohort by a practicing ophthalmologist (SB). This involved review of all structured (eg, diagnosis codes, microbiology values) and unstructured (eg, notes) data for each patient to identify any mention of fungal ocular involvement. The rationale for this was to establish a "gold standard" using a traditional approach and to ascertain whether there were any additional cases of fungal ocular involvement that may not have been identified by querying diagnosis codes or regular expression-based searches.

Other Patient Characteristics

In addition to the outcome of fungal ocular involvement for each patient, we also recorded demographic data, fungal species present in blood culture, and known risk factors for fungemia such as the presence of indwelling catheters, recent major surgery, diabetes, immunosuppression (either from underlying conditions or induced medically by immune-suppressing medications such as chemotherapy), history of intravenous drug use, and hyperalimentation. We obtained data regarding these various factors via SQL queries of the relevant tables in MIMIC-III based on current procedural terminology (CPT) events and ICD-9 codes for diagnoses and procedures. Definitions for risk factors that we used in the queries are detailed in Multimedia Appendix 2. Data were processed and analyzed using R [37].

Results

Figure 1 depicts an overview of the study methodology and results. In total, 265 patients in the MIMIC-III database had positive blood cultures containing fungal species. The mean (SD) age was 62.3 (16.8) years. A majority (199/265, 75%) self-identified as white, and there were slightly more males (154/265, 58%) than females (Table 1). The mean (SD) length of stay was 24.6 (27) days. The in-hospital mortality rate for these fungemic patients was high, with 121 (41%) who died during hospitalization, reflecting the severity of illness. The most common fungal species found on blood cultures were *Candida albicans* (114, 43%), *Candida glabrata* (74, 28%), *Candida parapsilosis* (32, 12%), and *Candida tropicalis* (21, 8%).

Over one-half of fungemic patients had recent major surgery (n=148, 55.8%), over one-third had indwelling central catheters (n=96, 36.2%), and over one-quarter carried a diagnosis of diabetes (n=69, 26.0%; Table 1). Other known risk factors for fungemia, such as cancer, immunosuppressed status from chemotherapy or steroids, intravenous drug use, and hyperalimentation, were relatively uncommon (all <10%) in this cohort.

Figure 1. Screening for fungal ocular involvement using approaches leveraging structured and unstructured data within the Medical Information Mart for Intensive Care III (MIMIC-III). SQL: Structured Query Language; ICD: International Classification of Disease.

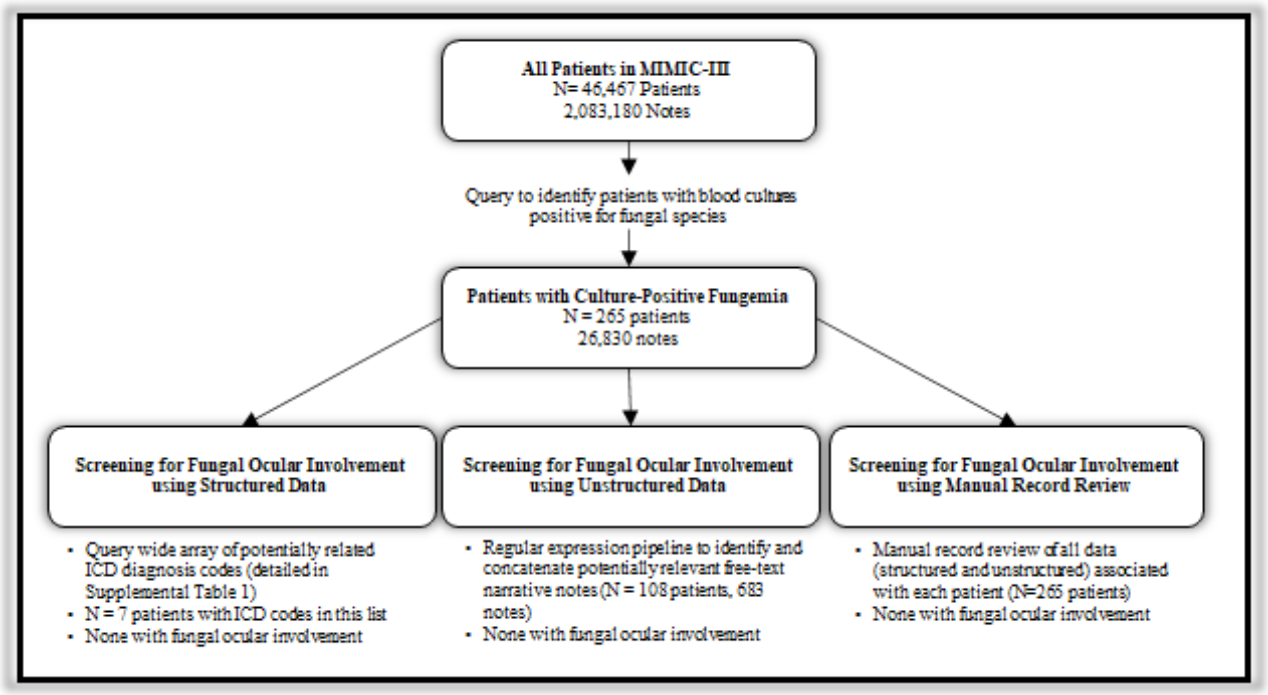


Table 1. Characteristics of patients with fungemia in MIMIC-III (N=265).

Characteristics	Value, n (%)
Gender	
Female	111 (41.9)
Male	154 (58.1)
Ethnicity	
White	199 (75.1)
Asian	6 (2.3)
Black/African American	22 (8.3)
Hispanic or Latino	5 (1.9)
Other	3 (1.1)
Patient declined or unknown	30 (11.3)
Risk factors for fungemia	
Indwelling central catheter	96 (36.2)
Diabetes	69 (26.0)
Cancer	19 (7.2)
Chemotherapy	4 (1.5)
Steroids	7 (2.6)
Intravenous drug use	11 (4.2)
Bone marrow transplant	0 (0)
Hyperalimentation	2 (0.8)
Surgery	148 (55.8)

Screening for Fungal Ocular Involvement Using Structured Data

A query of the MIMIC-III diagnosis tables among this cohort of fungemic patients using a broad range of eye- and vision-related ICD-9 codes (Multimedia Appendix 1) yielded 7 unique patients. For these patients, there were 1079 total associated clinical notes in the database. Examination of these records revealed that none had fungal ocular involvement. For example, one patient had been coded as having unilateral vision loss, but based on a review of her records, this was attributed to ischemic optic neuritis, not a fungal infection. Another patient endorsed blurry vision and poor visual acuity following a transfer out of the intensive care unit, prompting dilated fundus examination—there were diabetic proliferative changes noted but no evidence of fungal ocular involvement. Other reasons for ICD-9 codes related to vision changes among the fungemic patients in the cohort included occipital lobe stroke and posterior subcapsular cataracts. Out of all fungemic patients with ICD-9 codes related to vision changes, none were found to have fungal ocular involvement as the cause of their visual symptoms.

Screening for Fungal Ocular Involvement Using Unstructured Data

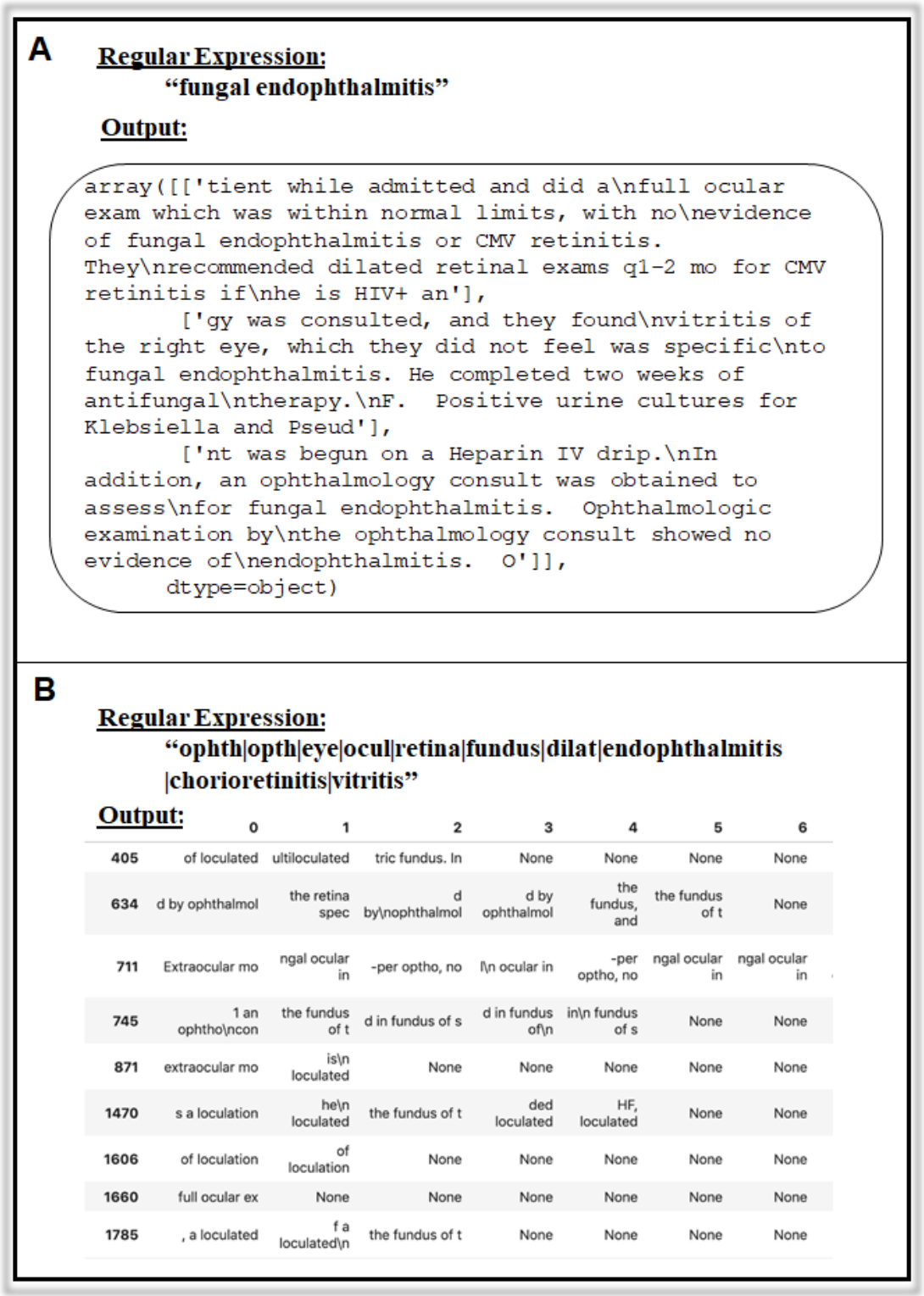
To leverage the free-text narrative notes in this database, we used a range of regular expressions to identify potential cases of fungal ocular involvement. These regular expressions formed the main component of pattern-matching scripts that returned the matches and surrounding text. First, we used algorithms centered on single phrases, such as “fungal endophthalmitis,” “fungal endopthalmitis” (accounting for common misspelling), and “fungal ocular involvement.” “Fungal endophthalmitis” yielded 3 notes, “fungal endophthalmitis” yielded one note, and “fungal ocular involvement” yielded no notes.

Next, we used a broader range of terms related not only to potential findings from fungal ocular involvement (eg, endophthalmitis, chorioretinitis, vitritis) but also to general eye examinations (eg, dilation, fundus, eye). This approach yielded 4000 notes from 255 patients. However, many mentions of “eye” and “dilat” had no relationship to fungal ocular involvement as these were very general terms. To improve specificity, we removed “eye” and “dilat” from the regular expression, resulting in 683 notes from 180 patients. Examples of the output from the regular expressions are depicted in Figure 2. Our strategy included not only the identification of relevant strings but also the concatenation of all relevant text strings together for each patient to facilitate a subsequent review. Based on a manual review of the relevant text strings identified by the algorithm in these notes, followed by a manual review of the full-length notes, we still did not identify a single case of fungal endophthalmitis or any variation of fungal ocular involvement.

Manual Record Review

A full manual record review was conducted for all 265 patients, encompassing a full-text review of all 26,830 notes as well as all structured data. There was no evidence of fungal ocular involvement for any patient in the cohort, even after manual record review.

Figure 2. Examples of output from regular expressions screening for fungal endophthalmitis. (A) Output with 100 characters on each side from 3 patients with "fungal endophthalmitis" mentioned. (B) Output with 5 characters on each side from a multi-term regular expression. Rows indicate patients, and columns indicate notes. In this example illustration, only the first 6 notes are displayed for a sample of patients. "None" indicates no matches from the regular expression were found in the note.



	0	1	2	3	4	5	6
405	of loculated	ultiloculated	tric fundus. In	None	None	None	None
634	d by ophthalmol	the retina spec	d by\nophthalmol	d by ophthalmol	the fundus, and	the fundus of t	None
711	Extraocular mo	ngal ocular in	-per optho, no	l\n ocular in	-per optho, no	ngal ocular in	ngal ocular in
745	1 an ophtho\ncon	the fundus of t	d in fundus of s	d in fundus of\n	in\n fundus of s	None	None
871	extraocular mo	is\n loculated	None	None	None	None	None
1470	s a loculation	he\n loculated	the fundus of t	ded loculated	HF, loculated	None	None
1606	of loculation	of loculation	None	None	None	None	None
1660	full ocular ex	None	None	None	None	None	None
1785	, a loculated	f a loculated\n	the fundus of t	None	None	None	None

Discussion

A major finding from this study was there was not a single case of fungal ocular involvement identified among culture-proven fungemic patients in the MIMIC-III database, which encompasses >50,000 hospital admissions for >46,000 patients between 2000 and 2012. This result is consistent with several other studies where rates of fungal ocular involvement were very low, generally <5% in studies conducted over the last two decades [6-10]. Most of those studies examined periods of 5 years or less, whereas here we did not detect a single case over 12 years. These patients demonstrated several risk factors associated with fungemia, as one-half had recent major surgery

(n=148, 55.8%), over one-third had indwelling central catheters (n=96, 36.2%), and over one-quarter carried a diagnosis of diabetes (n=69, 26.0%; Table 1). Indicative of the severity of infection and illness, this cohort had a high mortality rate, with over 40% of patients dying during the documented hospitalization. Therefore, there is no evidence to suggest that the lack of fungal ocular involvement was because this cohort was healthier or lower risk than those in prior studies.

Some have argued that targeted ophthalmic screening of patients with fungemia, rather than the universal screening that is currently recommended [28], may be a better use of personnel and financial resources, given the low rates of fungal ocular involvement and the rarity of changes in clinical management based on ophthalmic consultation [7,8,11,38,39]. However, prior studies have not rigorously described criteria for guiding targeted screening. Although there appears to be a consensus that nonverbal patients who are unable to communicate visual symptoms should be screened, there have not been rigorous reports of further risk stratification. Presumably, one reason for this is that the incidence of fungal ocular involvement is so low that achieving sufficient numbers for appropriately powered statistical models is difficult. With increasing efforts toward building multicenter clinical data warehouses, where data from multiple centers can be aggregated and analyzed together, there may be increased opportunities in the future to improve risk stratification.

However, the increasing volume of data in the current era of EHRs presents challenges. For example, in this cohort, there were almost 27,000 notes associated with fungemic patients. To address this, we developed a regular expression pipeline as an NLP-based approach to facilitate note review. NLP has been used to abstract findings from radiology and pathology reports [40-44], for identifying phenotypes from narrative notes [45-48], and specifically for ophthalmic data extraction such as visual acuity and surgical complications [20-22]. NLP has been shown to enhance case detection compared to structured diagnosis codes alone, for example, for identifying cases of pseudoexfoliation syndrome [26] and herpes zoster ophthalmicus [25]. Structured diagnosis codes have known limitations such as incomplete or inaccurate coding, insufficient granularity, and the fact that clinicians may not code for every condition at every encounter [29]. In this specific context, diagnosis codes likely have low sensitivity for fungal ocular involvement because ophthalmologists who examine patients serve in consultant roles rather than as primary providers. Therefore coding is typically performed by non-ophthalmologists who may not be familiar with eye-related diagnoses.

Our regular expression pipeline efficiently identified relevant text strings, extracted the associated notes, and subsequently collated them together for further review. Run times were short, at less than 3.5 seconds, much faster than manually combing through 27,000 notes, which took several weeks. We used simple regular expressions based on string-matching that were straightforward to implement. These were run on a standard local machine without requiring a high-performance computing infrastructure. We are unaware of previous reports of using NLP-based information extraction to identify cases of fungal ocular involvement and could not find any reference to this type of application in a PubMed search. Therefore, this approach represents a novel efficient "search" of a large volume and variety of critical care notes to identify the most relevant notes related to this diagnosis. Improving search functionalities will be crucial as EHRs generate an increasing volume of unstructured data requiring analysis.

Several important points arise when considering the issues of sensitivity and specificity of the structured versus unstructured approaches. Because we knew fungal ocular involvement to be a relatively rare condition based on prior studies, we erred on the side of sensitivity and cast a wide net in terms of included diagnosis codes as well as regular expressions. For this reason, we also did not include negation terms in the regular expressions in order to avoid inadvertently excluding any records with pertinent eye findings. At first glance, the structured diagnosis code query may seem to have been "better" than the unstructured NLP approach since it identified fewer patients and therefore appeared more specific and more efficient. However, in several prior studies, using diagnosis codes alone yielded fewer positive cases, resulting in under-detection and decreased sensitivity [25,26]. In this study, we could not assess the relative sensitivity and specificity of each approach, given that there were zero cases of fungal ocular involvement. However, because we manually reviewed the remaining notes that were not returned by the search, we were confident that these search terms did not "miss" any positive cases. Future studies of cohorts that include positive cases would benefit from comparing these metrics across different approaches that leverage structured data, unstructured data, or both. In addition, refining search strategies to include more nuanced approaches such as negation terms would also be relevant in future work.

We considered the possibility that there may have been true cases of fungal ocular involvement that were missed due to one or more of the following: (1) we did not include the appropriate diagnosis codes in our query, (2) we did not include the appropriate terms in our regular expressions, or (3) rare conditions may have been removed from the database as part of de-identification protocols. To address the first two possibilities, we performed a full manual record review of all data (structured and unstructured) for these 265 fungemic patients. There were still no cases of fungal ocular involvement. To address the possibility of removal for de-identification purposes, we contacted the principal investigators of the MIMIC-III database, who confirmed that rare diseases were not excluded from the database as part of any de-identification procedures. Therefore, we are confident that our findings reflect the true prevalence of fungal ocular involvement among patients with culture-proven fungemia in this database.

This study had some limitations. First, some of these patients may have developed fungal ocular involvement in the regular wards or the outpatient setting after being downgraded or discharged from critical care units, and this would not have been reflected in the inpatient critical care records analyzed. In addition, this database was based on a single academic medical center, so it is not clear whether these findings would generalize to other settings. However, fungemic patients tend to have significant co-morbidities and are, therefore, often treated at the tertiary care level. Despite being limited to a single center,

the 12-year longitudinal period of study encompassed in the database allowed for a large number of patients to be analyzed. There may have been patients with highly suspected fungemia and eye findings presumed to be fungal endophthalmitis. However, we did not include them in this analysis due to the difficulty of precisely defining "highly suspected fungemia" using either diagnosis codes or text. By restricting the analysis to patients with culture-proven fungemia, we increased certainty around the denominator in the prevalence calculation. Finally, only five patients had ICU providers who had copied and pasted ophthalmic exam findings verbatim into their notes. For the remaining patients, ICU notes contained summaries stating the ophthalmology consultation did not show any evidence of ocular involvement, but not a direct recapitulation of ophthalmic exam findings. Some patients may have had positive eye exam findings that were not captured in ICU provider notes, which is a limitation of using this particular data source. However, in clinical practice, typically, the detection of fungal ocular involvement is a rare and notable event such that the likelihood of an ICU provider not documenting a positive case would be low.

In summary, we demonstrated an application of NLP-based methods to a large-scale clinical database to gain insights about the prevalence of fungal ocular involvement. As clinical research increasingly includes unstructured narrative notes in addition to structured data, NLP will play a growing role in phenotyping. This approach will be critical for ophthalmological entities where details and descriptions are often embedded within notes rather than within structured data fields.

Acknowledgments

The authors would like to thank Shamim Nemati (UCSD Health Sciences Department of Biomedical Informatics), Fatemeh Amrollahi (UCSD Health Sciences Department of Biomedical Informatics), and Jit Battacharya (Amazon Web Services) for technical support during the project.

This study was supported by the National Institutes of Health (Bethesda, MD, Grants T15LM011271 and T32GM008806). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funders had no role in the design or conduct of this research.

Authors' Contributions

SLB conceived the study. SLB, ARK, GYY, and BRS performed the data analysis. All authors participated in data interpretation. SLB drafted the manuscript, and all authors revised it for important intellectual content and provided final approval of the version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Structured diagnosis codes used to screen for fungal ocular involvement. ICD-9 = International Classification of Diseases, version 9.
[DOCX File , 21 KB - jmir_v22i8e18855_app1.docx]

Multimedia Appendix 2
Definitions of features extracted from MIMIC-III regarding patient characteristics and risk factors for fungemia.
[DOCX File , 22 KB - jmir_v22i8e18855_app2.docx]

References

1. Durand ML. Bacterial and Fungal Endophthalmitis. Clin Microbiol Rev 2017 Jul;30(3):597-613 [FREE Full text] [doi: 10.1128/CMR.00113-16] [Medline: 28356323]
2. Henderson DK, Edwards JE, Montgomerie JZ. Hematogenous candida endophthalmitis in patients receiving parenteral hyperalimentation fluids. J Infect Dis 1981 May;143(5):655-661. [doi: 10.1093/infdis/143.5.655] [Medline: 6787141]
3. Brooks RG. Prospective study of Candida endophthalmitis in hospitalized patients with candidemia. Arch Intern Med 1989 Oct;149(10):2226-2228. [Medline: 2802888]
4. McDonnell PJ, McDonnell JM, Brown RH, Green WR. Ocular involvement in patients with fungal infections. Ophthalmology 1985 May;92(5):706-709. [doi: 10.1016/s0161-6420(85)33989-1] [Medline: 3874382]
5. Bross J, Talbot GH, Maislin G, Hurwitz S, Strom BL. Risk factors for nosocomial candidemia: a case-control study in adults without leukemia. Am J Med 1989 Dec;87(6):614-620. [doi: 10.1016/s0002-9343(89)80392-4] [Medline: 2589396]
6. Feman SS, Nichols JC, Chung SM, Theobald TA. Endophthalmitis in patients with disseminated fungal disease. Trans Am Ophthalmol Soc 2002;100:67-70; discussion 70 [FREE Full text] [Medline: 12545679]

7. Adam MK, Vahedi S, Nichols MM, Fintelmann RE, Keenan JD, Garg SJ, et al. Inpatient Ophthalmology Consultation for Fungemia: Prevalence of Ocular Involvement and Necessity of Funduscopic Screening. Am J Ophthalmol 2015 Nov;160(5):1078-1083.e2. [doi: 10.1016/j.ajo.2015.07.033] [Medline: 26235339]
8. Dozier CC, Tarantola RM, Jiramongkolchai K, Donahue SP. Fungal eye disease at a tertiary care center: the utility of routine inpatient consultation. Ophthalmology 2011 Aug;118(8):1671-1676. [doi: 10.1016/j.ophtha.2011.01.038] [Medline: 21550121]
9. Huynh N, Chang HP, Borboli-Gerogiannis S. Ocular involvement in hospitalized patients with candidemia: analysis at a Boston tertiary care center. Ocul Immunol Inflamm 2012 Apr;20(2):100-103. [doi: 10.3109/09273948.2011.646383] [Medline: 22409562]
10. Donahue SP, Hein E, Sinatra RB. Ocular involvement in children with candidemia. Am J Ophthalmol 2003 Jun;135(6):886-887. [doi: 10.1016/s0002-9394(02)01705-1] [Medline: 12788131]
11. Ghodasra DH, Eftekhari K, Shah AR, VanderBeek BL. Outcomes, impact on management, and costs of fungal eye disease consults in a tertiary care setting. Ophthalmology 2014 Dec;121(12):2334-2339 [FREE Full text] [doi: 10.1016/j.ophtha.2014.06.042] [Medline: 25113869]
12. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA 2013 Apr 03;309(13):1351-1352. [doi: 10.1001/jama.2013.393] [Medline: 23549579]
13. Cai T, Giannopoulos AA, Yu S, Kelil T, Ripley B, Kumamaru KK, et al. Natural Language Processing Technologies in Radiology Research and Clinical Applications. Radiographics 2016;36(1):176-191 [FREE Full text] [doi: 10.1148/rg.2016150080] [Medline: 26761536]
14. Savova GK, Tseytlin E, Finan S, Castine M, Miller T, Medvedeva O, et al. DeepPhe: A Natural Language Processing System for Extracting Cancer Phenotypes from Clinical Records. Cancer Res 2017 Nov 01;77(21):e115-e118 [FREE Full text] [doi: 10.1158/0008-5472.CAN-17-0615] [Medline: 29092954]
15. Jackson R, Patel R, Velupillai S, Gkotsis G, Hoyle D, Stewart R. Knowledge discovery for Deep Phenotyping serious mental illness from Electronic Mental Health records. F1000Res 2018;7:210 [FREE Full text] [doi: 10.12688/f1000research.13830.2] [Medline: 29899974]
16. Zeng Z, Deng Y, Li X, Naumann T, Luo Y. Natural Language Processing for EHR-Based Computational Phenotyping. IEEE/ACM Trans Comput Biol Bioinform 2019;16(1):139-153 [FREE Full text] [doi: 10.1109/TCBB.2018.2849968] [Medline: 29994486]
17. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. J Biomed Inform 2018 Jan;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]
18. Ohno-Machado L. Realizing the full potential of electronic health records: the role of natural language processing. J Am Med Inform Assoc 2011 Jan;18(5):539-549 [FREE Full text] [doi: 10.1136/amiajnl-2011-000501] [Medline: 21846784]
19. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. JMIR Med Inform 2019 Apr 27;7(2):e12239 [FREE Full text] [doi: 10.2196/12239] [Medline: 31066697]
20. Baughman DM, Su GL, Tsui I, Lee CS, Lee AY. Validation of the Total Visual Acuity Extraction Algorithm (TOVA) for Automated Extraction of Visual Acuity Data From Free Text, Unstructured Clinical Records. Transl Vis Sci Technol 2017 Mar;6(2):2 [FREE Full text] [doi: 10.1167/tvst.6.2.2] [Medline: 28299240]
21. Liu L, Shorstein NH, Amsden LB, Herrinton LJ. Natural language processing to ascertain two key variables from operative reports in ophthalmology. Pharmacoepidemiol Drug Saf 2017 Apr;26(4):378-385 [FREE Full text] [doi: 10.1002/pds.4149] [Medline: 28052483]
22. Wang SY, Pershing S, Tran E, Hernandez-Boussard T. Automated extraction of ophthalmic surgery outcomes from the electronic health record. Int J Med Inform 2020 Jan;133:104007. [doi: 10.1016/j.ijmedinf.2019.104007] [Medline: 31706228]
23. Barrows Jr RC, Busuioc M, Friedman C. Limited parsing of notational text visit notes: ad-hoc vs. NLP approaches. Proc AMIA Symp 2000:51-55 [FREE Full text] [Medline: 11079843]
24. Maganti N, Tan H, Niziol LM, Amin S, Hou A, Singh K, et al. Natural Language Processing to Quantify Microbial Keratitis Measurements. Ophthalmology 2019 Dec;126(12):1722-1724. [doi: 10.1016/j.ophtha.2019.06.003] [Medline: 31307829]
25. Zheng C, Luo Y, Mercado C, Sy L, Jacobsen SJ, Ackerson B, et al. Using natural language processing for identification of herpes zoster ophthalmicus cases to support population-based study. Clin Exp Ophthalmol 2019 Jan;47(1):7-14. [doi: 10.1111/ceo.13340] [Medline: 29920898]
26. Stein JD, Rahman M, Andrews C, Ehrlich JR, Kamat S, Shah M, et al. Evaluation of an Algorithm for Identifying Ocular Conditions in Electronic Health Record Data. JAMA Ophthalmol 2019 May 01;137(5):491-497 [FREE Full text] [doi: 10.1001/jamaophthalmol.2018.7051] [Medline: 30789656]
27. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
28. Pappas PG, Kauffman CA, Andes D, Benjamin DK, Calandra TF, Edwards JE, Infectious Diseases Society of America. Clinical practice guidelines for the management of candidiasis: 2009 update by the Infectious Diseases Society of America. Clin Infect Dis 2009 Mar 01;48(5):503-535 [FREE Full text] [doi: 10.1086/596757] [Medline: 19191635]

29. Stein JD, Lum F, Lee PP, Rich WL, Coleman AL. Use of health care claims data to study patients with ophthalmologic conditions. Ophthalmology 2014 May;121(5):1134-1141 [FREE Full text] [doi: 10.1016/j.ophtha.2013.11.038] [Medline: 24433971]
30. Lane H, Howard C, Hapke HM. Information extraction (named entity extraction and question answering) · Natural Language Processing in Action: Understanding, analyzing, and gener. In: Natural Language Processing in Action. Shelter Island, New York: Manning Publications; 2019.
31. Hinchcliff M, Just E, Podlusky S, Varga J, Chang RW, Kibbe WA. Text data extraction for a prospective, research-focused data mart: implementation and validation. BMC Med Inform Decis Mak 2012 Sep 13;12:106 [FREE Full text] [doi: 10.1186/1472-6947-12-106] [Medline: 22970696]
32. Weeks HL, Beck C, McNeer E, Williams ML, Bejan CA, Denny JC, et al. medExtractR: A targeted, customizable approach to medication extraction from electronic health records. J Am Med Inform Assoc 2020 Mar 01;27(3):407-418. [doi: 10.1093/jamia/ocz207] [Medline: 31943012]
33. Bui DDA, Zeng-Treitler Q. Learning regular expressions for clinical text classification. J Am Med Inform Assoc 2014;21(5):850-857 [FREE Full text] [doi: 10.1136/amiajnl-2013-002411] [Medline: 24578357]
34. Kim YS, Yoon D, Byun J, Park H, Lee A, Kim IH, et al. Extracting information from free-text electronic patient records to identify practice-based evidence of the performance of coronary stents. PLoS One 2017;12(8):e0182889 [FREE Full text] [doi: 10.1371/journal.pone.0182889] [Medline: 28800599]
35. 7.2. re — Regular expression operations. The Python Standard Library. URL: https://docs.python.org/2/library/re.html [accessed 2020-02-03]
36. Klie AR. fungal_endophthalmitis. GitHub. 2020. URL: https://github.com/adamklie/fungal_endophthalmitis [accessed 2020-03-23]
37. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2013. URL: http://www.R-project.org/ [accessed 2020-01-01]
38. Vinikoor MJ, Zoghby J, Cohen KL, Tucker JD. Do all candidemic patients need an ophthalmic examination? Int J Infect Dis 2013 Mar;17(3):e146-e148 [FREE Full text] [doi: 10.1016/j.ijid.2012.12.014] [Medline: 23369606]
39. El-Abiary M, Jones B, Williams G, Lockington D. Fundoscopy screening for intraocular candida in patients with positive blood cultures-is it justified? Eye (Lond) 2018 Nov;32(11):1697-1702 [FREE Full text] [doi: 10.1038/s41433-018-0160-x] [Medline: 29980768]
40. Hripcsak G, Austin JHM, Alderson PO, Friedman C. Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. Radiology 2002 Jul;224(1):157-163. [doi: 10.1148/radiol.2241011118] [Medline: 12091676]
41. Chapman WW, Fizman M, Chapman BE, Haug PJ. A comparison of classification algorithms to automatically identify chest X-ray reports that support pneumonia. J Biomed Inform 2001 Feb;34(1):4-14 [FREE Full text] [doi: 10.1006/jbin.2001.1000] [Medline: 11376542]
42. Crowley RS, Castine M, Mitchell K, Chavan G, McSherry T, Feldman M. caTIES: a grid based system for coding and retrieval of surgical pathology reports and tissue specimens in support of translational research. J Am Med Inform Assoc 2010;17(3):253-264 [FREE Full text] [doi: 10.1136/jamia.2009.002295] [Medline: 20442142]
43. Coden A, Savova G, Sominsky I, Tanenblatt M, Masanz J, Schuler K, et al. Automatically extracting cancer disease characteristics from pathology reports into a Disease Knowledge Representation Model. J Biomed Inform 2009 Oct;42(5):937-949 [FREE Full text] [doi: 10.1016/j.jbi.2008.12.005] [Medline: 19135551]
44. Strauss JA, Chao CR, Kwan ML, Ahmed SA, Schottinger JE, Quinn VP. Identifying primary and recurrent cancers using a SAS-based natural language processing algorithm. J Am Med Inform Assoc 2013;20(2):349-355 [FREE Full text] [doi: 10.1136/amiajnl-2012-000928] [Medline: 22822041]
45. Denny JC, Peterson JF, Choma NN, Xu H, Miller RA, Bastarache L, et al. Extracting timing and status descriptors for colonoscopy testing from electronic medical records. J Am Med Inform Assoc 2010;17(4):383-388 [FREE Full text] [doi: 10.1136/jamia.2010.004804] [Medline: 20595304]
46. Seyfried L, Hanauer DA, Nease D, Albeiruti R, Kavanagh J, Kales HC. Enhanced identification of eligibility for depression research using an electronic medical record search engine. Int J Med Inform 2009 Dec;78(12):e13-e18 [FREE Full text] [doi: 10.1016/j.ijmedinf.2009.05.002] [Medline: 19560962]
47. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA 2011 Aug 24;306(8):848-855. [doi: 10.1001/jama.2011.1204] [Medline: 21862746]
48. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. Ann Intern Med 1995 May 01;122(9):681-688. [doi: 10.7326/0003-4819-122-9-199505010-00007] [Medline: 7702231]

Abbreviations

CITI: Collaborative Institutional Training Initiative

CPT: Current Procedural Terminology
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
ICD: International Classification of Disease
IRB: Institutional Review Board
MIMIC-III: Medical Information Mart for Intensive Care III
NLP: natural language processing
SQL: Structured Query Language

Edited by G Eysenbach; submitted 23.03.20; peer-reviewed by S Wang, GE Iyawa; comments to author 11.04.20; revised version received 21.04.20; accepted 13.06.20; published 14.08.20.

Please cite as:
Baxter SL, Klie AR, Radha Saseendrakumar B, Ye GY, Hogarth M
Text Processing for Detection of Fungal Ocular Involvement in Critical Care Patients: Cross-Sectional Study
J Med Internet Res 2020;22(8):e18855
URL: https://www.jmir.org/2020/8/e18855
doi:10.2196/18855
PMID:32795984

Original Paper

Self-Selection of Bathroom-Assistive Technology: Development of an Electronic Decision Support System (Hygiene 2.0)

Manon Guay[1,2], PhD; Karine Latulippe[3], PhD; Claudine Auger[4,5], PhD; Dominique Giroux[6,7,8], PhD; Noémie Séguin-Tremblay[2], MSc; Josée Gauthier[9], B Erg; Catherine Genest[10], MSc; Ernesto Morales[7,11], PhD; Claude Vincent[7,11], PhD

[1]School of Rehabilitation, Université de Sherbrooke, Sherbrooke, QC, Canada
[2]Center for Research on Aging, Sherbrooke, QC, Canada
[3]Department of Teaching and Learning Studies, Université Laval, Quebec, QC, Canada
[4]School of Rehabilitation, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada
[5]Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal, Montreal, QC, Canada
[6]Center of Excellence on Aging Quebec, Quebec, QC, Canada
[7]Department of Rehabilitation, Faculty of Medicine, Université Laval, Quebec, QC, Canada
[8]VITAM - Centre de recherche en santé durable, Quebec, QC, Canada
[9]CIUSSS de la Mauricie-et-du-Centre-du-Québec, Trois-Rivières, QC, Canada
[10]CIUSSS de la Capitale-Nationale, Quebec, QC, Canada
[11]Center for Interdisciplinary Research in Rehabilitation and Social Integration, Quebec, QC, Canada

Corresponding Author:
Manon Guay, PhD
School of Rehabilitation
Université de Sherbrooke
3001, 12e avenue Nord
Sherbrooke, QC
Canada
Phone: 1 819 780 2220 ext 45484
Email: Manon.Guay@USherbrooke.ca

Abstract

Background: A clinical algorithm (Algo) in paper form is used in Quebec, Canada, to allow health care workers other than occupational therapists (OTs) to make bathroom adaptation recommendations for older adults. An integrated knowledge transfer process around Algo suggested an electronic version of this decision support system (electronic decision support system [e-DSS]) to be used by older adults and their caregivers in search of information and solutions for their autonomy and safety in the bathroom.

Objective: This study aims to (1) create an e-DSS for the self-selection of bathroom-assistive technology by community-dwelling older adults and their caregivers and (2) assess usability with lay users and experts to improve the design accordingly.

Methods: On the basis of a user-centered design approach, the process started with content identification for the prototype through 7 semistructured interviews with key informants of various backgrounds (health care providers, assistive technology providers, and community services) and 4 focus groups (2 with older adults and 2 with caregivers). A thematic content transcript analysis was carried out and used during the creation of the prototype. The prototype was refined iteratively using think-aloud and observation methods with a clinical expert (n=1), researchers (n=3), OTs (n=3), older adults (n=3), and caregivers (n=3), who provided information on the usability of the e-DSS.

Results: Overall, 4 themes served as the criteria for the prototype of the electronic Algo (Hygiene 2.0 [H_2.0]): focus (safety, confidentiality, well-being, and autonomy), engage, facilitate (simplify, clarify, and illustrate), and access. For example, users first pay attention to the images (engage and illustrate) that can be used to depict safe postures (safety), illustrate questions embedded in the decision support tool (clarify and illustrate), and demonstrate the context of the use of assistive technology (safety and clarify).

Conclusions: The user-centered design of H_2.0 allowed the cocreation of an e-DSS in the form of a website, in line with the needs of community-dwelling older adults and their caregivers seeking bathroom-assistive technology that enables personal

hygiene. Each iteration improved usability and brought more insight into the users' realities, tailoring the e-DSS to the implementation context.

(J Med Internet Res 2020;22(8):e16175) doi:10.2196/16175

KEYWORDS

hygiene; activities of daily living; decision aids; occupational therapy; aging; self-help devices; universal design; accidental falls; mobile phone

Introduction

Background

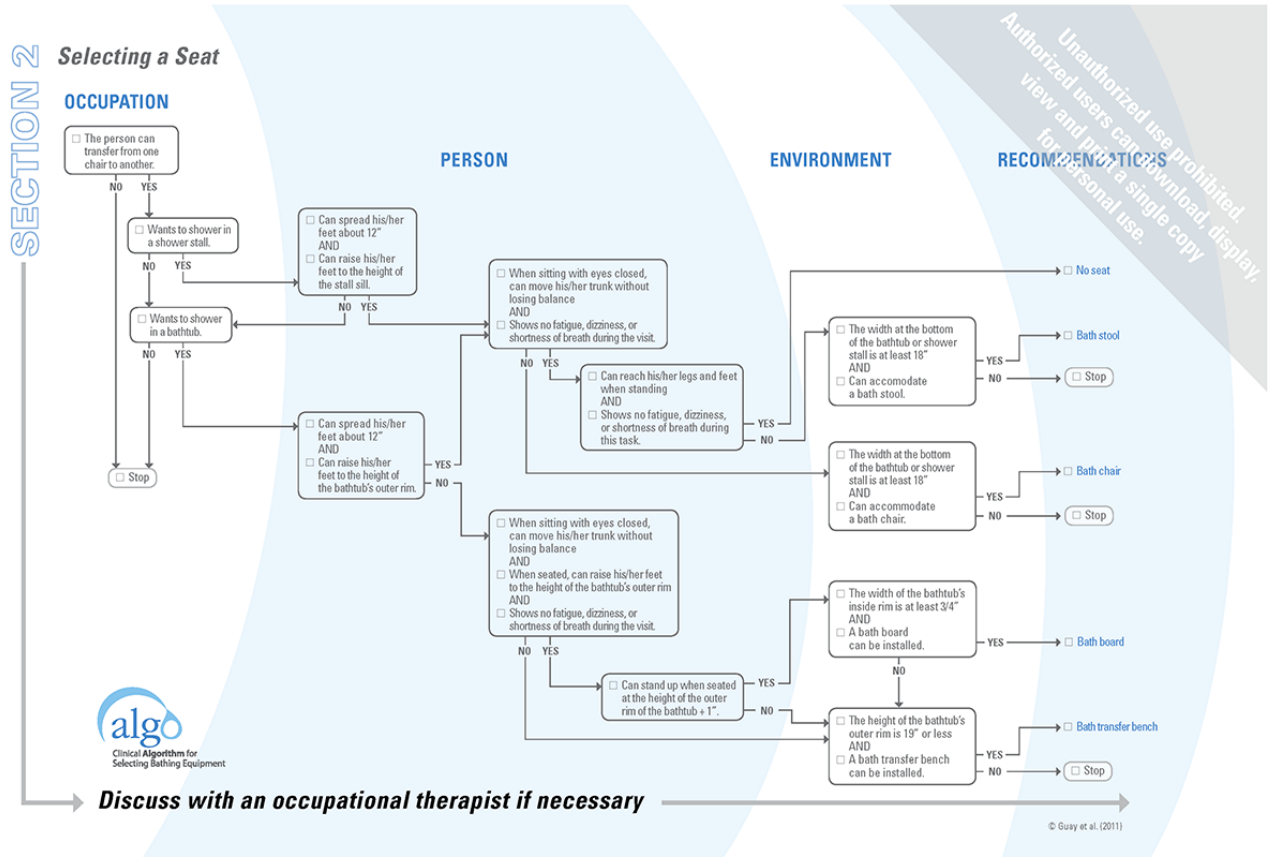
Performing personal hygiene is an essential daily activity for health and dignity that commonly becomes difficult with aging. A survey of 28,406 noninstitutionalized Canadians (50-104 years old) revealed that the prevalence of disability increased with age, exponentially increasing when considering the oldest old [1]. At 90 years of age, 21% of Canadians reported requiring assistance to wash themselves [1]. In such cases, adapting the bathroom environment with assistive technologies, such as bath seats, grab bars, or nonslip mats, is a common recommendation to promote autonomy and safety [2,3].

To address this issue in Quebec, a mainly French-speaking province of Canada, the clinical *algo*rithm Algo [4] has been proposed allowing occupational therapists (OTs) to collaborate with non-OTs: health care workers other than OTs, such as home health aides, social workers, and nurse assistants. In a situation deemed straightforward (ie, clients of standard morphology with predictable occupational performance in bath transfer in their standard shower stall or bathtub) [5], Algo supports decision making for home care clients who need recommendations on full-body hygiene, considering their preferences as well as their abilities and the actual physical environment in which they live [6].

Indeed, Algo is a clinical algorithm in paper form paired with a user guide that outlines the logical steps for non-OTs to select bathing equipment (Figure 1). It also identifies complex cases to be referred to an OT [6]. Older adults receive 1 of 9 possible recommendations [6], with or without assistive technology, specific to their bathing situation (ie, standing without a seat in the bathtub, sitting on a bath stool in the shower stall, or stop and refer to an OT). The content of the Algo was developed through an integrated knowledge approach [6,7], and psychometric studies [8-12] revealed, for example, that it guided non-OTs toward a bath seat that meets the needs of community-dwelling older adults in the majority of cases (mean 84%, SD 9%) [9]. The appropriateness rate of seats recommended by non-OTs did not statistically differ from that of 2 OTs [9].

In 2015, approximately half (48%) of the targeted end users knew about Algo, with half of these (24%) reporting that they had begun the implantation process in their clinical settings [13]. Since then, community OTs have ideated about converting the knowledge embedded in Algo into an electronic version to be used by their clients [14]. It appears to these OTs that Algo could be the foundation for an electronic decision support system (e-DSS) to help people make informed decisions about assistive technologies [15]. Usually, e-DSSs include first, a knowledge base; second, a program to combine that knowledge with user-specific information; and third, an interface used to collect data about the user and to provide the user with relevant information [16].

Figure 1. Section 2 of the paper format Algo.



Objectives

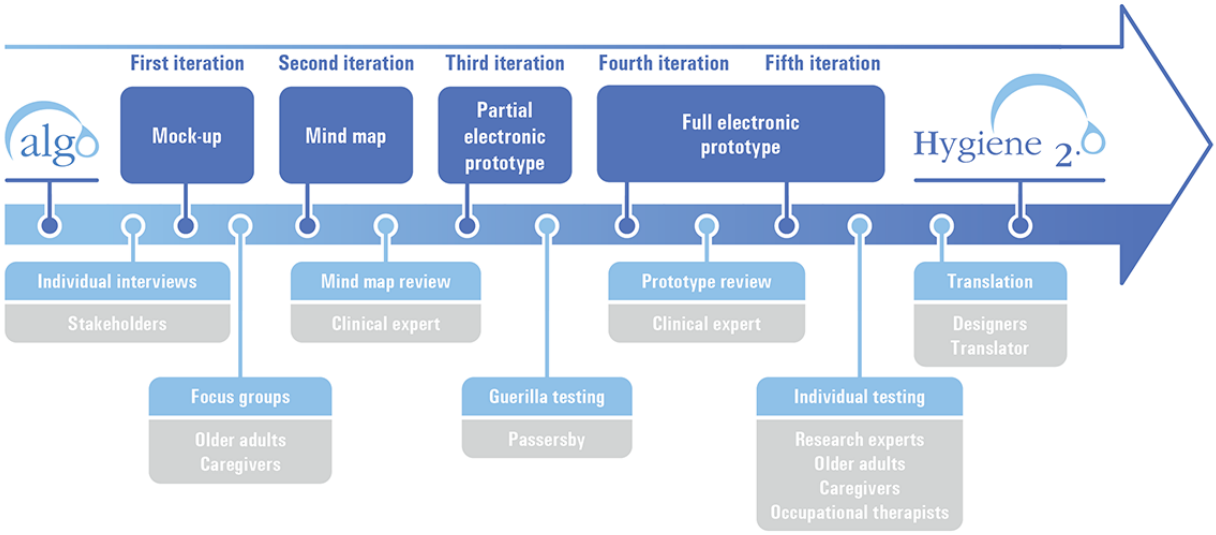
This study aims to adapt Algo's paper version to an e-DSS for older adults and caregivers experiencing difficulty when performing personal hygiene. Specifically, the objectives were to (1) create an e-DSS for the self-selection of bathroom-assistive technology by community-dwelling older adults and their caregivers, and (2) assess usability with lay users and experts to improve the design accordingly.

Methods

Design Process

A user-centered design method [17] was conducted, in which an ongoing iterative process facilitated dialogue between potential users and designers (Figure 2). The designers are the principal investigator (an OT) and a research professional (background in mechanical engineering and cognitive ergonomics). The term older adults refers to community-dwelling adults aged at least 65 years who have difficulty completing their personal hygiene routine and their caregivers refers to people concerned about or assisting a community-dwelling older adult with their personal hygiene. Although Algo is the name of the algorithm in paper form, Hygiene 2.0 (H_2.0) is the name given to the electronic version of the decision support system.

Figure 2. Iterative design process.



First Iteration

In addition to the knowledge gathered during our previous studies, 7 semistructured interviews were conducted with stakeholders, recruited by word-of-mouth, to participate in the research project for their professional experience by providing information on bathing assistive technology for the elderly or selecting and providing such technology. The interviews aimed at gaining deeper knowledge of the needs an e-DSS should address in their work context. In a private room in their work settings, participants were encouraged to express their point of view following open-ended questions from an interview guide developed by the research team and iteratively modified (Multimedia Appendix 1). Interviews were recorded and transcribed. Qualitative data analysis principles of thematic data condensation were first conducted by the interviewer using Microsoft Word and Excel software. A total of 2 members of the research team reviewed and clarified the theme definitions by analyzing the transcription extracts iteratively (individual coding and dyad work sessions). Memos were used to facilitate reflexivity and research team discussions.

The designers considered these themes while creating the first version of an electronic Algo as they represent user needs related to this e-DSS. A mock-up was designed on PowerPoint as a rough draft using 6 questions extracted from Algo and pictures taken by the designers to illustrate them (Figure 3). The mock-up allowed the collection of participants' spontaneous comments and questions regarding specific features, facilitating the iterative design process.

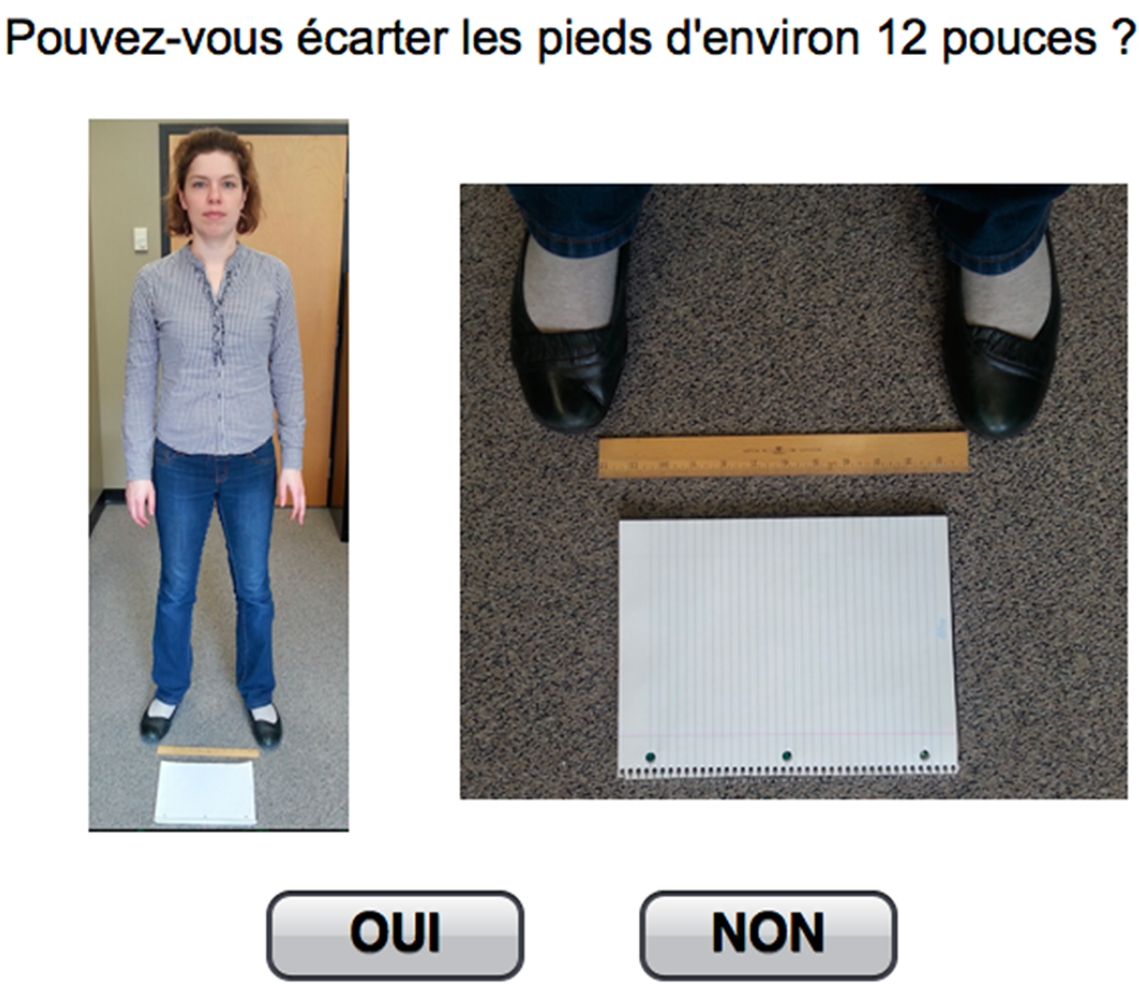
A total of 4 focus groups were organized to give users the opportunity to express their experience regarding the choice of assistive technologies, their needs regarding health information, their use of computer technology, and their opinion on the PowerPoint mock-up (Multimedia Appendix 2). A moderator (the research professional), a content expert (the principal investigator), and a graduate student responsible for logistics were present in the room ensuring privacy.

Overall, 2 of those focus groups included older adults and were constituted with a systematic sampling procedure. Every fifth volunteer aged 65 years on a list from the Research Center on Aging was contacted. A research assistant reviewed with the volunteer, over the phone, the following inclusion criteria: having experienced difficulties with bathing per the definition of the *Functional Autonomy Measurement System* [18]. The exclusion criteria were cognitive impairment limiting expression or comprehension and inability to speak French.

Additionally, 2 focus groups were conducted with caregivers. Recruitment was performed in collaboration with 2 community resources (a domestic help service and a volunteer bureau), which helped identify caregivers for people having difficulty with bathing. The same exclusion criteria applied to older adults in the previous focus groups were applied to the caregivers.

All 4 focus groups were transcribed verbatim and analyzed with the methodology described above for the interviews, but using NVivo software (QSR International) to analyze the data. Modifications were made to the PowerPoint mock-up after each focus group. A grid relating the themes to specific solutions for the prototype was implemented to verify that each general theme emerging through the coding would be translated into practical solutions and that each modification to the prototype corresponded to the previously established themes (Multimedia Appendix 3). In addition to considering the user needs and context, tracking and correcting usability challenges were also considered within the iterative evolution of the prototype. Emerging knowledge was integrated in an ongoing process into the different versions of the prototypes.

Figure 3. Excerpt from the PowerPoint mock-up (translation: Can you spread your feet about 12 inches? Options: Yes, No).



Second Iteration

The analysis of the focus groups integrated with the results from the interviews revealed that an offline mobile app and a responsive website would be the preferred formats for users of the adapted Algo. To do so, the Algo algorithm was broken down and reorganized into a diagram, referred to as a mind map [19], considering previously gathered data regarding the users and the context of use. For example, the item order in the paper form of Algo was modified to make the e-DSS easy to understand and minimize the steps before obtaining an assistive technology recommendation.

An OT with 25 years of clinical experience in home care for older adults was hired to conduct a thorough review of the mind map. She was familiar with Algo in paper form and had trained OTs and non-OTs on using the same. The expert conducted a careful reading of the results of the qualitative analysis and the mind map prototype. She was asked to provide her professional opinions on equivalency, noting missing or superfluous information, as well as on literacy (target audience: 10-year-old reader). A total of five 2-hour unstructured interviews were conducted with her (for a total of 10 hours), encouraging the think-aloud process and modifying the mind map live with her throughout the meetings.

Third Iteration

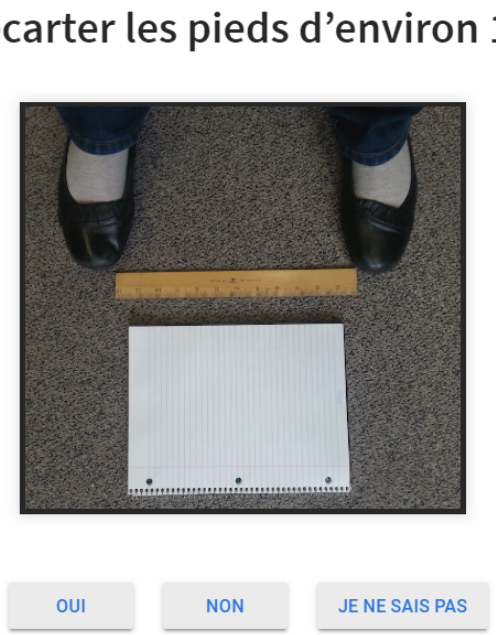
The conception of a partial electronic prototype was initiated with the involvement of a programmer. The questions of the mind map relating to the recommendation “sitting on a bath stool in the shower stall” were programmed.

In parallel, to identify the best practices in web design for older adults, a literature review was conducted with the assistance of 2 librarians. The databases Education Resources Information Center (ERIC), AgeLine, Medical Literature Analysis and Retrieval System Online (MEDLINE), and Cumulated Index to Nursing and Allied Health Literature (CINAHL) were searched for publications before January 1, 2014, in French or English with such keywords as *Internet**, *health**, *health education**, *instructional design**, *product design**, and *older computer users**. Of the 47 references selected by the principal investigator for their relevance to the subject, the guidelines from Chaffin and Maddux [20], Czaja [21], Fisk et al [22], Nielsen [23], the National Institute on Aging (NIA) [24], and norms International Organization for Standardization (ISO) 9241-210 [25] and ISO 9241-11 [26] were applicable to this

study. Relevant publications regarding literacy and usability published since then were also considered during the design: US Federal Plain Language Guidelines [27], a guide for communicating in the field of health from the Montreal Health and Social Services Agency [28], and a guide from the US Department of Health on the design of easy-to-use health websites [29].

The partial electronic prototype (Figure 4) was used for guerrilla testing, which consists of validating a design by conducting quick usability tests with passersby in a public space [30]. In our case, this was done during the poster sessions of 2 geriatric and rehabilitation scientific conferences in Sherbrooke and Montreal, Canada. Volunteers were recruited among attendees (interdisciplinary Masters or PhD students, rehabilitation researchers, research professionals, health care professionals, and decision makers). Although these settings did not include older adults, they did include other stakeholders and offered a good opportunity to rapidly get feedback on the preliminary design. Specifically, these tests aimed at verifying the font and size of the text, the page layout, the understanding of questions and associated images, and the swift functioning of the navigation. After giving them minimal contextual setting, the programmer and a designer, respectively, observed and took note of the volunteers' comments and behavior while they tested the prototype presented on a tablet. As the partial prototype was pared down and the testing method was quick (approximately 5 min), the comments were brief. The programmer and designer discussed these comments and notes were taken. Aspects with the greatest impact on usability (eg, loading of the web pages was slow, hyperlinks were not understood as well as buttons) were considered to make corresponding changes. When the comments were contradictory between volunteers and no trend could be identified, they were kept to be verified during the following tests.

Figure 4. Excerpt from the partial prototype (translation: Can your family member or loved one stand with his or her feet about 12 inches [30 cm] apart? Options: Yes, No, I Don't Know).



Fourth Iteration

A full prototype was programmed, with the process being performed backward, meaning each of Algo's 9 potential recommendations was programmed and added to the prototype one at a time. After every addition, the designers reviewed the prototype independently, navigating systematically and randomly, to identify potential errors and suggest improvements. They compared the equivalence between questions within the original paper form of Algo and the prototype. Every discrepancy was noted and discussed among the research team to verify the rationale for changes within the results.

This procedure was repeated with a clinical OT; the expert hired for the second iteration was recruited. She had to independently compare the algorithm structure of the paper form and the electronic form and question discrepancies while thinking aloud. The interview lasted 134 min. Designers gathered observations in person, verifying that they had a rationale within the study results, iteratively modifying the grid relating the themes to specific solutions for the prototype. Incoherencies were discussed among designers and the prototype was modified accordingly to have it ready for usability testing with potential users. As these users advised during previous iterations, a professional photographer was hired, and 3 older adults were recruited by word-of-mouth to appear in pictures depicting the questions in the prototype.

Fifth Iteration

The prototype was first tested with 3 researchers, recruited by word-of-mouth, having previously worked as clinical OTs. Although they were not the target users, they had clinical experience, and as researchers tend to be very thorough, these tests acted as a comprehensive review of the prototype. Moreover, 2 categories of users and 1 category of stakeholders were then tested: older adults, caregivers, and OTs. A sample of 3 participants from each category was formed as it is the most resource-efficient way to conduct testing [31]. Recruitment of

the older adults and caregivers followed the same procedure described for the focus groups. Tests were conducted at each participant's home. Overall, 3 OTs were recruited by word-of-mouth and tested in their work settings.

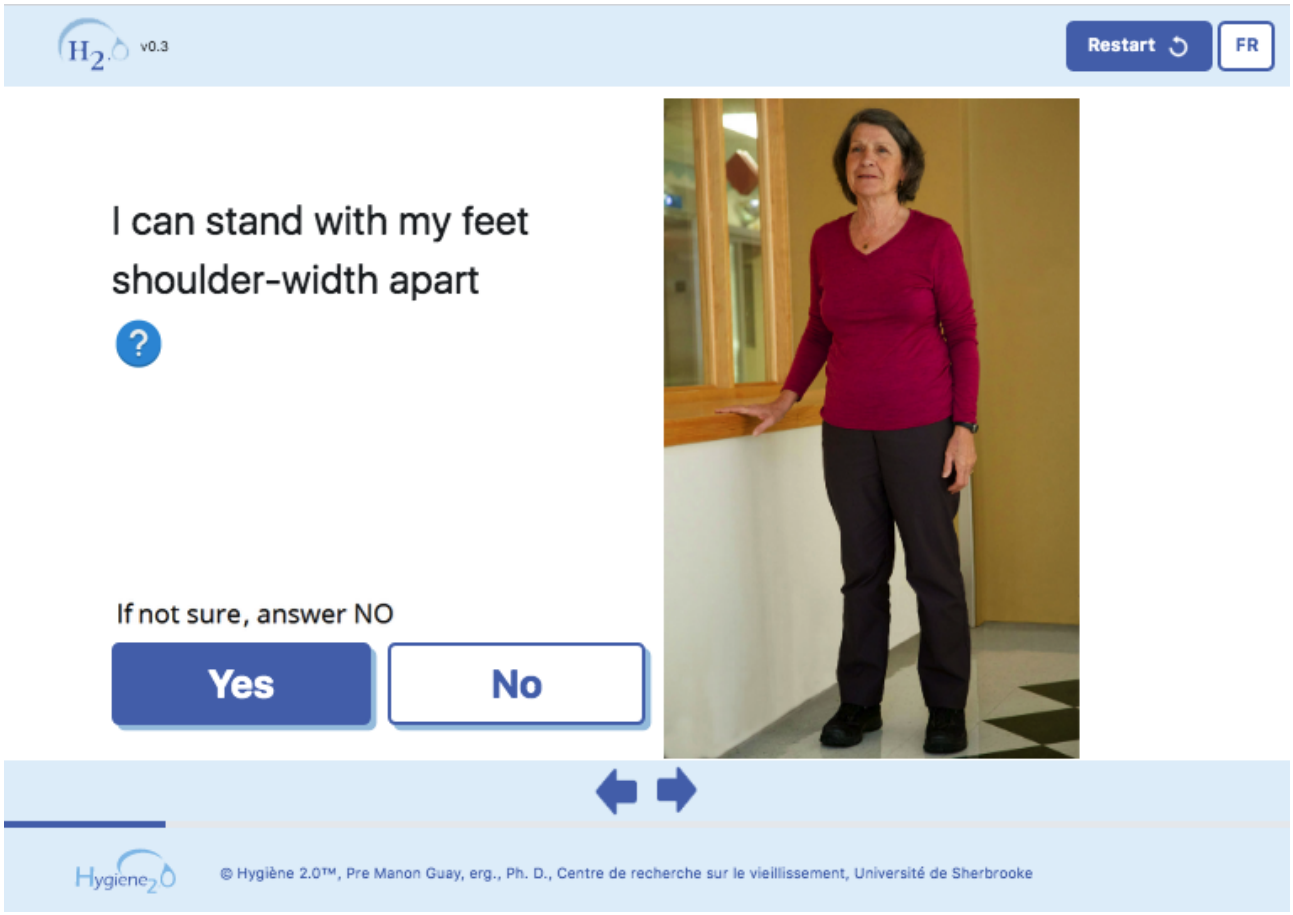
Every participant was invited to explore the prototype as if they were using it alone on a personal device of their choosing while enunciating their thoughts out loud [32]. The interviewer would ask the reason for a certain reaction or facial expression. Thorough note taking and recording of interviews allowed us to compile and analyze comments, from which the designers and programmer evaluated the modifications deemed essential before further testing. The decision to modify the prototype following a comment was made by considering its relation to the themes drawn from the previous research on user needs; its influence on the user's comprehension; the resources available; and the different, possibly contradicting, comments. All modifications applied to the prototype would always help the user's comprehension to maintain the adequacy of the consequent recommendations. The comments with the highest impact on usability were prioritized to make corresponding changes (Figure 5). Modifications were made to the prototype following each group of the 3 tests described earlier.

Figure 5. Excerpt from the French version of the full prototype (translation: I can stand with my feet shoulder-width apart; if not sure, answer No; Options: Yes, No).

Translation and Graphic Identity

French was the working language throughout the study. Therefore, the prototype was developed and tested in French. However, English being the language used by most Canadians, and a language for scientific communications, an English translation (Figure 6) was performed by the designers and reviewed by a certified translator. The designers and the programmer collaborated with a graphic designer to create a graphic identity for H_2.0.

Figure 6. Excerpt from the English version of the full prototype.



Ethical Considerations

This research project was approved by the ethics committee of the CIUSSS de l'Estrie-CHUS. Sociodemographic data were not gathered during the second, third, and fourth iterations. Indeed, the second and fourth iterations involved an expert who helped design and review the structure of the e-DSS and not interface use. The guerrilla testing on the third iteration values quick and numerous comments on a few key interface aspects, and therefore should not involve sociodemographic data gathering. Confidentiality during navigation on H_2.0 was ensured by not asking for identity-related information (eg, no profile creation before use) and the computer's internet protocol address was not collected. The website is hosted locally on the research center server.

Results

Participant Characteristics

A total of 47 different participants were recruited during the 5 iterations. For iterations 1 and 5, 24 (75%) of the participants were women. Tables 1 and 2 present the sociodemographic characteristics. Interviews lasted on average 48 min (range 30-82 mins), focus groups lasted on average 105 min (range 98-116 mins), and testing lasted on average 53 min (range 30-93 mins).

Table 3 presents the different devices and browsers used by the participants during the fifth iteration. Researchers and OTs preferred using their smartphones or laptops, whereas older adults and caregivers were generally more comfortable using their desktop computer or iPad. Tests performed on a variety of brands and internet browsers allowed verification of website functioning.

Table 1. Number and profile of the participants by iteration.

Participant profile	Iterations, n (%)					Total, n (%)
	First	Second	Third	Fourth	Fifth	
Older adults	8 (17)	N/A[a]	N/A	N/A	3 (6.4)	11 (23.4)
Caregivers	5 (10.6)	N/A	N/A	N/A	3 (6.4)	8 (17)
Occupational therapists	3 (6.4)	1 (2.1)	N/A	1[b] (2.1)	3 (6.4)	7[b] (14.9)
Researchers	N/A	N/A	N/A	N/A	3 (6.4)	3 (6.4)
Other stakeholders	4[c] (8.5)	N/A	14[d] (29.8)	N/A	N/A	18 (38.3)
Total	20 (42.5)	1 (2.1)	14 (29.8)	1[b] (2.1)	12 (25.6)	47 (100)

[a]N/A: not applicable.

[b]The clinical expert was the same on the second and fourth iterations.

[c]Other stakeholders were members of health care interdisciplinary teams (n=1), members of community resources (n=2), and a bathing assistive technology provider (n=1).

[d]Passersby in rehabilitation or geriatric scientific conferences could include caregivers, clinical experts, researchers, and occupational therapists.

Table 2. Sociodemographic characteristics of participants for the first and last iterations (n=32).

Iteration and participants	Sample, n (%)	Age (years), mean (SD); (range)	Education (years), mean (SD); (range)	Use of internet (years), mean (SD); (range)	OT[a] experience (years since graduation), mean (SD); (range)
First iteration					
Interviewees	7 (22)	41 (7.4); (30-51)	17 (1); (16-18)	—[b]	N/A[c]
Focus group members	13 (41)	69 (7.1); (52-84)	14 (3.2); (8-18)	14 (9.7); (0-30)	N/A
Fifth iteration					
Prototype reviewers	3 (9)	49 (4); (45-53)	22 (0); (22-22)	23 (2.5); (20-25)	26 (4.1); (21-29)
Testers	9 (28)	61 (13.8); (41-76)	16 (3.2); (12-22)	24 (6.9); (12-30)	22 (6); (18-29)[d]

[a]OT: occupational therapist.

[b]Not available.

[c]N/A: not applicable.

[d]Values only for the 3 testers who were OTs.

Table 3. Devices and browsers used during individual testing (fifth iteration).

Participants	Device used for testing	Brand	Internet browser used for testing
Researchers			
1	Smartphone	Samsung	Chrome
2	Smartphone	Samsung	Chrome
3	Smartphone	Samsung Galaxy	Safari
Older adults			
1	Desktop	iMac Apple	Safari
2	Desktop	ASUS	Chrome
3	iPad	Apple	Chrome
Caregivers			
1	Desktop	ASUS	Firefox
2	Laptop and smartphone	Macbook pro and Samsung Galaxy A5	Chrome
3	iPad	Apple	Chrome
Occupational therapists			
1	Laptop	HP	Chrome
2	Smartphone	iPhone 6S	Safari
3	Smartphone	Samsung Note 3	Chrome

Identification and Integration of Stakeholders' Perspective During Prototype Design

Overall, 4 emerging themes (and 7 subthemes) of an e-DSS for older adults having difficulty performing personal hygiene, and their caregivers, were identified from interviews, focus groups, and tests. These are focus (safety, confidentiality, autonomy, and well-being), engage, facilitate (simplify, clarify, and illustrate), and access. Multimedia Appendix 3 presents a summary of how these themes and subthemes were translated into features of the prototype to adapt Algo's paper version into an e-DSS using the participants' perspective. The following paragraphs describe the themes and subthemes, which are included in each title. The integration of these themes in the resulting website is also illustrated in a video [33].

Focus on Underlying Values

The e-DSS for persons having difficulty performing personal hygiene should focus on 4 purposes: safety, confidentiality, well-being, and autonomy.

Promote Safety During Navigation and After

The e-DSS should promote the physical safety of seniors with diminishing autonomy. Physical safety must be ensured not only during the use of the system but also during the implementation of its recommendations:

> *No, but it's because you can see that there is no non-slip surface, you can see that there is no grab bar to hold on to. […] For me, it is not safe.* [Focus group–caregiver]

> *Q: You look at her position and you think: "Would I be able to take that position exactly?" A: Yes, that's it.* [Focus group–caregiver]

> *Q: What do you imagine when you read this question (I can hold onto a grab bar with each hand) […]? A: Well, I think of safety.* [Individual testing–older adult]

In addition, OTs indicated that having users enter both feet in the bathtub could be unsafe for some of them. The e-DSS would need to indicate only lifting the feet up to the bath rim to test one's capacity.

Ensure Confidentiality of Personal Information

The e-DSS should ensure the confidentiality of personal information. As the domestic help service director of a community resource explained in an interview, "… maybe there could be some reluctance from clients to enter their information into the client file… be careful with sensitive information…"

Enhance People's Well-Being

The e-DSS should enhance people's well-being. Sensible questions (eg, is the person at the end of his or her life?) should be rephrased and asked only if needed, and at the end of the navigation. Navigation and questions must also be straightforward to minimize the risk of confusion:

> *[…] caregivers who call us are in a higher age range […] so starting to explore each item when they already have some confusion about other things I think could be difficult for them.* [Interview−advisor in a caregiver support organization]

Promote Functional Autonomy of Users

The e-DSS should promote the autonomy of users seeking information about bathing difficulties and how to mitigate them:

> *It becomes humiliating, I find [to not be able to wash ourselves]. We can't get along on our own, we feel diminished, we don't feel like ourselves. It's as simple*

> *as that. It isn't much, but that's the way it is.* [Focus group–older adult]
>
> *"I am finding it hard to wash myself" [opening question], that summarizes a little what we would need; give us tips […] because our mobility is reduced. I mean, my back hurts. It's out of the question for me to wash my back or to wash below my knees. I can't bend. […] Maybe I will never be able to do that again, but at least if they suggested some tips.* [Focus group–older adult]

Engage Users Throughout Navigation

The e-DSS should engage users from the very beginning, opening the navigation with their worries. The main worry is hygiene, as an older adult said during a focus group, *"Yes, I really found hygiene was the most difficult."* Focus groups and interviews highlighted that in addition to difficulties in performing hygiene, two other major concerns must be addressed as opening questions. First, fear of falling while performing hygiene is a common reason to initiate bathroom adaptations:

> *[…] it's not rare for the child to say "my mother fell in the bathroom on the weekend, and now I want to make the bathroom safer, what options are available?* [Interview−director in an assistive technology store]
>
> *Even today, getting in and out of the bathtub, I'm always afraid of falling. That has remained. So, you see, it's been almost eight years. The fear has remained.* [Focus group–older adult]

Second, acquiring knowledge about available assistive technologies would also be a hook for potential users:

> *Sometimes caregivers would tell me "I bought this"… and they would ask me "is what I bought adequate?" […] Often the aids […] would ask "what do you think would be the safest to use as equipment?"… "is there another piece of equipment that could be safer, more adapted to the person?* [Interview–OT]
>
> *They give us tips like buying a sponge (the ones with the long handle) […], but to use it to reach our toes, that's another story.* [Focus group–older adult]

Facilitate Navigation

The e-DSS should make navigation as easy as possible. This implies simplifying, clarifying, and illustrating.

Simplify the Navigation Experience

First, navigation on the e-DSS should be simple and therefore short in time, requiring a minimal number of steps:

> *Because on Internet, that's the danger. We go over things quickly.* [Focus group–older adult]
>
> *You have to read a lot. I mean, nothing can be understood quickly in the image. You really must take your time to look at the […] different steps or different recommendations. You have to read. [Is it a lot of text for you?] Yes.* [Focus group—older adult]
>
> *[Did you think it was possible to […] go back only one screen?] Oh no, the thought did not cross my mind.* [Individual testing–caregiver]

Clarify Information Requested or Provided

Second, the information should be clear enough to facilitate the user's general comprehension and ensure the accuracy of the answers given to the e-DSS's questions. For example, users need to know what to answer if they are unsure of their capacities:

> *There is something about the closed question […] sometimes it's as if they [elderly people] are uncomfortable, […] it creates some hesitation, raises some anxiety, it's as if, well I don't know if it's yes, I don't know if it's no, well sometimes, so I don't know what to answer.* [Interview advisor in a caregiver support organization]
>
> *It would have been yes and no. If it had been in cycles, it would have been no when I just got out, and, a month later, yes. [Would you have answered yes or no according to now, the moment when you responded to the questionnaire, is that it?] Exactly. Today, yes, but a month ago, no. Oh no. [Would it be better to say: "Today, can you…"] It might be better to explain the moment when you have to spread your legs 12 inches. That might be better.* [Focus group–older adult]

Consquently, to provide clarity to the user, the sentence "If not sure, answer No" was added above the Yes and No buttons.

Another example of clarity is that the user must know exactly what to consider when answering the question:

> *Maybe have a note saying if it is for someone else, consider the other person's capacities.* [Individual testing–caregiver]
>
> *[…] I don't understand the context […] if I was a lady who did not have the notion of transfer capacity, I would wonder "why should I do that?" […] I would tend to word this question "I am able to move from one chair to the other" […]* [Individual testing–OT]

Clarity also means that the user must be aware of his progress while navigating. An OT said during individual testing, *"Well, it took me some time before I realized it was the end."*

The e-DSS also ensures clarity by offering two navigation pathways, one for older adults and one for caregivers. The caregiver path gives the option of being accompanied and answering on behalf of someone else.

> *Maybe encouraging people to navigate with another person could be interesting also [...]* [Quebec College of OTs (OEQ) staff member]

Illustrate What Is Communicated

Specifically, the e-DSS should use images to promote understanding and simplification of the questions while still considering the user's perceptions and preferences. Indeed, users rely on images to further understand the question and its context:

> *That image, for me, I find it very confusing, because I'm looking […] OK, a bath seat, I read the first sentence, "anti-slip mat". But then I don't see the relevance. […] I would move on to something else.* [Focus group–older adult]
>
> *Me, I am very visual, so I prefer with images.* [Focus group–older adult]
>
> *It should be as explicit as possible regarding what a standard environment is or what the particular characteristics can bring in terms of difficulty in installing the equipment so that what is usually done through the healthcare professional's judgement can be done by the client.* [Interview executive advisor for OT services]
>
> *According to the picture, I would say that you have to use your hands [to get up]. [Individual testing–caregiver]*
>
> *[What do you think of this picture compared to the drawing in the beginning?]. It's [the picture] easier, we can really see what it is.* [Individual testing–older adult]

Ensure Ready Access to Users

The e-DSS should be readily accessible to users. Potential users said that they would be happy to use an electronic tool that would be provided to them via a web link by their health professionals:

> *I think it would be wonderful. I mean, I leave the hospital […] and I don't have to go meet anyone at all, I just go on my computer. Advice, everything I need is there. I put them into action or I don't. I just have to type on my keyboard and I have it all. All my data is there.* [Focus group–older adult]
>
> *After an operation, when we return home, they say: "There's a computer program. We'll send it to you. Check it out."* [Focus group–older adult]
>
> *Had I known that a computer program like that one existed [H_2.0], I would have gone on Internet, yes, even if I'm not 100% computer-savvy. But I would have gone to check it out.* [Focus group–older adult]
>
> *Yes. I give my e-mail and it comes to my place. Hello.* [Focus group–older adult]

Furthermore, the use of an offline mobile app that can be downloaded at one time and used at another, when the internet is not available, is both practical and strategic. In fact, the internet is not always available to users for reasons related to cost or location (eg, cost of a data plan on a mobile device or availability of high-speed internet in rural areas). Users also have varying internet access depending on where they work and live:

> *I'm thinking of when I was working with a hospitalized clientele: it's one computer for five professionals, and there is no computer having an access to internet.* [Interview–OT]

A choice of either an offline mobile app or a responsive website, using any device, portable or fixed, could guarantee availability to most users. Through the focus groups, we gathered information that users prefer different devices for different activities (eg, consult bank account, play games, read an email). Through interviews, it appeared that users might also not be able to choose which device they use in their work context because of what is provided to them and what is restricted for confidentiality reasons:

> *Well, normally it's not well looked upon [to use a personal cell phone with a patient] on a confidentiality level and all.* [Interview–OT]
>
> *Well, if I had the choice, […] tablets are really an interesting technology because they are very easy to use.* [Interview–OT]

Users highlighted the importance of finding a name for the app that would contribute to accessing it. After brainstorming by the research team, the electronic prototype was named *H*ygiene *2.0*. This name is bilingual (French and English), short, and evocative of the purpose and the format.

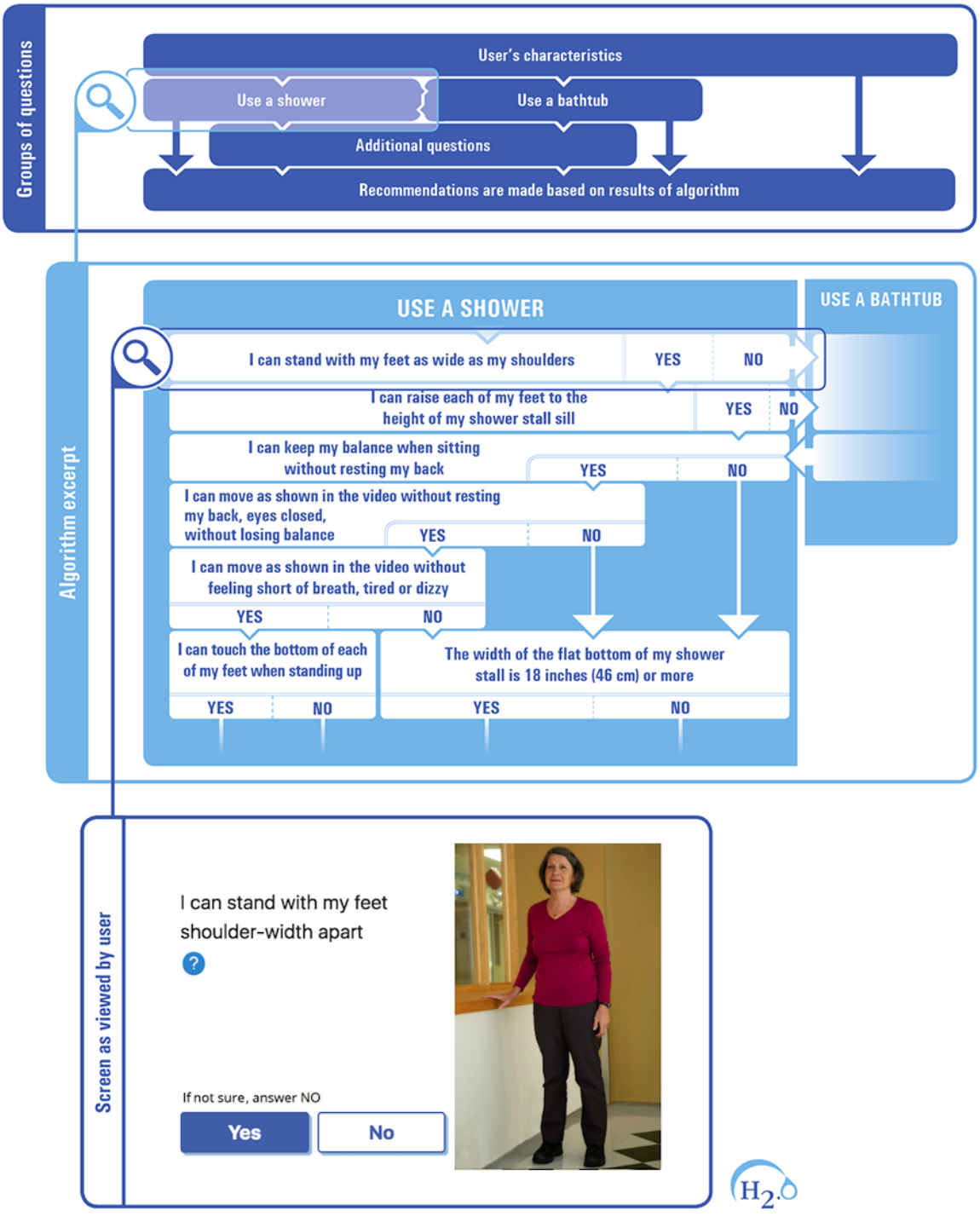
Hygiene 2.0 Full Electronic Prototype

Although H_2.0 is an adaptation of Algo's paper format decision support tool, some key differences exist between the 2. First, some questions are combined in the paper form (Figure 1), whereas the questions in the H_2.0 e-DSS are asked one at a time. This is meant to meet the facilitating criteria, a need expressed by older adults and caregivers. Even though it might add steps, it makes each question simpler and clearer for both pathways, the one designed for older adults facing hygiene challenges, and the one for their caregivers navigating with, or for them.

Second, although the whole algorithm is presented at once to a health care worker using Algo's paper format, in the H_2.0 e-DSS, each question presented to the user will be determined by the response given to the previous question. This dynamic aspect of the electronic version makes the process different for each user.

Third, the items in Algo are structured in 4 sections. The first 2 sections are grouped according to 3 central concepts in occupational therapy: occupation, person, and environment (Figure 1). Indeed, components are addressed successively, starting with items related to the occupation, then the ones regarding the person, and finally, those related to his or her environment. This structure is no longer followed in H_2.0 (Figure 7). All aspects are still covered, but the order favored is in line with user needs; for example, minimizing the number of questions and presenting the more sensitive ones at the end. The theoretical occupational therapy background is not as obvious for a lay user in H_2.0 as it is in Algo.

Figure 7. Structure of Hygiene 2.0.



Fourth, the paper format takes into account the user's desire regarding the place where he or she wants to shower: *Wants to shower in a bathtub* (Algo) versus *I shower in a standard bathtub* (H_2.0) and *Wants to shower in a shower stall* (Algo) versus *I shower in a standard shower stall* (H_2.0). This was thoroughly discussed while comparing the algorithm structure of the paper form and the electronic form with an OT:

> *I find it's at another level, not only preference. I can prefer washing myself in the bath but not be able to get into the bath […] so I don't do it in the bath but I would like to do it in the bath. […] There is preference and then there is capacity, the fact that the shower or the bath becomes an architectural barrier.* [Prototype review–OT]

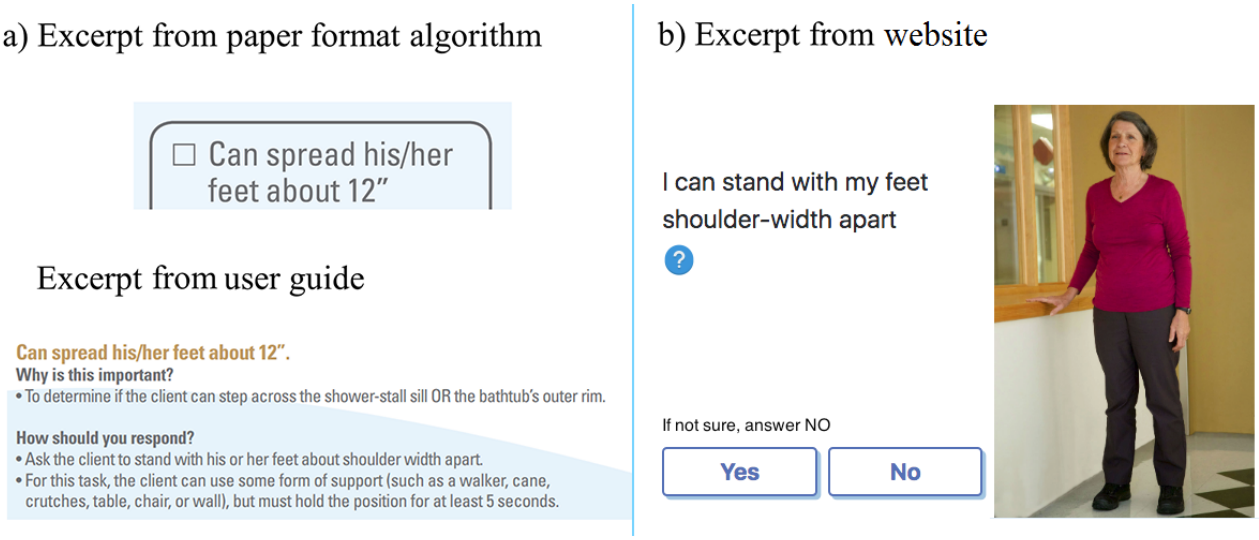
It is possible for a health care professional using Algo's paper format with a client to consider his or her preference while

taking into consideration his or her capacities as well as the presence of a standard shower or bathtub in the dwelling (Figure 1). This becomes more challenging when answering a series of questions presented one at a time, without the user's knowledge of the whole algorithm. To minimize the number of questions and maintain the yes or no answering pattern, the questions regarding the place where the user showers were modified in the e-DSS. This meets the user's need for facilitating. Although it does not consider the user's preference for showering in the shower stall or in the bathtub, it still evaluates the capacity and architectural barriers (standard shower stall or bathtub).

Fifth, to enhance facilitating while ensuring safety as desired by potential users, some details are not made explicit in the text but by combining text and picture, thereby minimizing the amount of text. For example, the question in the H_2.0 e-DSS does not explicitly state that the client can use some form of support as it is done in the Algo user guide (Figure 6). However, the picture shows a woman using a support with her right hand.

Sixth, to further enhance facilitating while answering questions, some formulations were adapted. Instead of asking the user to spread his or her feet 12 inches apart as in Algo, the shoulder-width reference is used in H_2.0 (Figure 8).

Figure 8. Spreading feet about 12 inches. (a) Paper format Algo and (b) website Hygiene 2.0.



Discussion

Principal Findings

This study aims to adapt Algo's paper form into an e-DSS to allow self-selection of bathroom adaptation by community-dwelling older adults. On the basis of a user-centered design method, e-DSS H_2.0 was created by taking into consideration the perspective of older adults having difficulty performing personal hygiene in their home, caregivers, OTs, assistive technology providers, and community resources. To enhance usability, H_2.0 focuses on 4 purposes (safety, confidentiality, well-being, and autonomy); engages users from the beginning by listing potential major concerns on the first page; and eases navigation as much as possible by simplifying, clarifying, and illustrating. For older adults having difficulty bathing (or their caregivers), navigating on such an e-DSS would enable them to, for straightforward cases, self-select common assistive technologies fitting their needs to promote autonomy and safety. The more complex cases identified will be redirected to occupational therapy services.

One might wonder about the choice of opting for a digital tool for people aged 65 years and older. Seniors, particularly those aged between 65 and 74 years, are increasingly using the internet as a source of information for their health. According to a CEFRIO (Centre facilitant la recherche et l'innovation dans les organisations) survey, 80% of people living in the province of Quebec, Canada, aged 65 years now have an internet connection at home, 60% use the internet every day, and 76% of internet users evaluate their personal internet use skills at an average or high level [34]. Digital technologies are, therefore, becoming major potential vectors to help older people take care of their health. However, it is important not to exclude those who do not use technologies; hence, the importance of caregivers being involved in the development of H_2.0 and the e-DSS giving an option to answer on behalf of someone else. Community pharmacists that sell assistive technology for hygiene could also be users of H_2.0 as well as other formal resources available in the community. This would allow people who are unfamiliar with the internet to receive guidance when using the website. Besides including the possibility of receiving help, the website is also easy to use for people with lower digital literacy. Nonetheless, as the older adults and caregivers who tested the website during the fifth iteration had been using the internet for many years, future usability tests should involve people with lower digital literacy.

According to the World Health Organization (WHO) guidelines, H_2.0 has reached the mid-demonstration stage [35] for monitoring and evaluating digital health interventions throughout their development process from the prototype stage to full implementation of the technology. The WHO defines 6 stages of an intervention maturity life cycle (preprototype,

prototype, pilot, demonstration, scale-up, integrated and sustained program) grouped into 3 larger categories: early, mid, and advanced stages. The H_2.0 e-DSS has been tested with users successively in an increasingly realistic environment, the latest of which is the older adults and caregiver's homes, navigating with their personal history in mind on their personal device. Therefore, realistic insights have been gathered to customize H_2.0 for a near future scaling-up of this electronic health intervention.

Nevertheless, H_2.0 can be considered as an intervention applied in controlled conditions, limited in terms of population and geography testing. It has yet to be deployed on a larger scale to reach its full implementation, which will be the subject of further studies as it will bring forth new challenges. Among other things, this will mean adapting the recommendations to the implementation context and creating the different features that were suggested during the focus groups and interviews but have not yet been applied (Multimedia Appendix 3). For example, more information regarding the questions and recommendations should be made available with links and help buttons. Moreover, a small convenience sample did not allow the capture of a variety of digital literacy and bathing difficulties, which should be pursued through additional field testing. The psychometric qualities of an e-DSS are also important aspects to investigate before full-scale implementation [36].

This research was conducted in French as it is the primary language in Quebec, Canada; the English version of H_2.0 lacks transcultural adaptation [36], a process that could be conducted further on. Despite 2 bilingual content experts and a translator revising the last iteration, the prototype might still employ field-specific terms that need to be adapted to make them understandable to the general French- and English-speaking public while conveying the correct meaning.

Conclusions

On the basis of knowledge embedded in the clinical *algo*rithm Algo, H_2.0 [33] is an e-DSS conceived for individuals having difficulty performing personal hygiene or people concerned about or assisting a community-dwelling older adult. H_2.0 focuses on safety, confidentiality, well-being, and autonomy to support the self-selection of common assistive technologies. Facilitating assistive technology provision for hygiene using H_2.0 implies simplifying, clarifying, and illustrating yes or no simple questions initially designed for non-OTs as well as developing an accessible format such as a responsive website and an offline mobile application. H_2.0 has reached the demonstration stage and could be integrated in a continuum of care to enhance the personal hygiene of community-dwelling older adults.

Acknowledgments

The authors received financial support from the *Réseau Québécois de recherche sur le vieillissement*, Research Center on Aging, and Faculty of Medicine and Health Sciences of *Université de Sherbrooke*. MG and CA received salary support from the *Fonds de Recherche en Santé du Québec*. The authors are thankful to Mr Guy Leblanc for his contribution in conducting the user-centered approach.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide for stakeholders.
[PDF File (Adobe PDF File), 82 KB - jmir_v22i8e16175_app1.pdf]

Multimedia Appendix 2
Interview guide for focus groups.
[PDF File (Adobe PDF File), 92 KB - jmir_v22i8e16175_app2.pdf]

Multimedia Appendix 3
Integration of stakeholders' perspective while designing an electronic prototype (H_2.0).
[DOCX File , 19 KB - jmir_v22i8e16175_app3.docx]

References

1. Guay M, Dubois M, Corrada M, Lapointe-Garant M, Kawas C. Exponential increases in the prevalence of disability in the oldest old: a Canadian national survey. Gerontology 2014;60(5):395-401 [FREE Full text] [doi: 10.1159/000358059] [Medline: 24818716]
2. King EC, Novak AC. Effect of bathroom aids and age on balance control during bathing transfers. Am J Occup Ther 2017;71(6):7106165030p1. [doi: 10.5014/ajot.2017.027136] [Medline: 29135427]
3. Zingmark M, Bernspång B. Meeting the needs of elderly with bathing disability. Aust Occup Ther J 2011 Jun;58(3):164-171. [doi: 10.1111/j.1440-1630.2010.00904.x] [Medline: 21599681]

4. Guay M. Algo. Outil Algo. 2012. URL: http://ergotherapie-outil-algo.ca/en/algo/ [accessed 2019-09-02]
5. Guay M, Dubois M, Desrosiers J, Robitaille J. Identifying characteristics of 'straightforward cases' for which support personnel could recommend home bathing equipment. Brit J Occup Ther 2012 Dec;75(12):563-569. [doi: 10.4276/030802212x13548955545576]
6. Guay M, Dubois M, Robitaille J, Desrosiers J. Development of Algo, a clinical algorithm for non-occupational therapists selecting bathing equipment. Can J Occup Ther 2014 Oct;81(4):237-246. [doi: 10.1177/0008417414539643] [Medline: 29898503]
7. Guay M, Levasseur M, Turgeon-Londeï S, Dubois M, Desrosiers J. Exploring support needed by home health aides in choosing bathing equipment: new challenges for occupational therapy collaboration. Work 2013 Jan 1;46(3):263-271. [doi: 10.3233/WOR-121527] [Medline: 23324694]
8. Guay M, Gagnon M, Ruest M, Bourget A. Interrater reliability of Algo used by non-occupational therapist members of homecare interdisciplinary teams. Disabil Rehabil 2017 May;39(9):883-888. [doi: 10.3109/09638288.2016.1168488] [Medline: 27111712]
9. Guay M, Dubois M, Desrosiers J. Can home health aids using the clinical algorithm Algo choose the right bath seat for clients having a straightforward problem? Clin Rehabil 2014 Feb;28(2):172-182. [doi: 10.1177/0269215513494027] [Medline: 23897948]
10. Ruest M, Bourget A, Delli-Colli N, Guay M. Algo used by homecare nonoccupational therapists selecting bathing assistive technology: enhancing standardization by exploring clinical reasoning. Occup Ther Health Care 2017 Jan;31(1):20-33. [doi: 10.1080/07380577.2016.1270481] [Medline: 28094591]
11. Ruest M, Bourget A, Delli-Colli N, Guay M. Recommandations pour favoriser la standardisation de la passation de l'Algo par le personnel non-ergothérapeute. Occup Ther Now 2019:- [FREE Full text]
12. Guay M. Le point sur les travaux de développement de l'Algo, un outil d?aide à la décision pour la sélection des aides techniques pour l'hygiène corporelle des québécois vivant à domicile. Revue Francophone de Recherche en Ergothérapie 2017;3(1):-. [doi: 10.18356/2333723a-fr]
13. Ruest M, Léonard G, Thomas A, Contandriopoulos D, Guay M. Canadian Knowledge Mobilization Forum 2016. Institute for Knowledge Mobilization. 2016. URL: http://www.knowledgemobilization.net/wp-content/uploads/2016/11/Canadian-Knowledge-Mobilization-Forum-2016-REPORT.pdf [accessed 2019-09-02]
14. Guay M, Ruest M, Contandriopoulos D. Deimplementing untested practices in homecare services: a preobservational-postobservational design. Occup Ther Int 2019;2019:5638939 [FREE Full text] [doi: 10.1155/2019/5638939] [Medline: 31015826]
15. Tucker S, Brand C, O'Shea S, Abendstern M, Clarkson P, Hughes J, et al. An evaluation of the use of self-assessment for the provision of community equipment and adaptations in English local authorities. Brit J Occup Ther 2011 Mar;74(3):119-128. [doi: 10.4276/030802211x12996065859201]
16. Berner E. Clinical Decision Support Systems: State of the Art. Agency for Healthcare Research and Quality. 2009. URL: https://healthit.ahrq.gov/sites/default/files/docs/page/09-0069-EF_1.pdf [accessed 2020-07-13]
17. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
18. Hebert R, Carrier R, Bilodeau A. The functional autonomy measurement system (SMAF): description and validation of an instrument for the measurement of handicaps. Age Ageing 1988 Sep;17(5):293-302. [doi: 10.1093/ageing/17.5.293] [Medline: 2976575]
19. Mind Map. Cambridge Dictionary. 2018. URL: https://dictionary.cambridge.org/dictionary/english/mind-map?q=mind+map [accessed 2019-09-02]
20. Chaffin AJ, Maddux CD. Accessibility accommodations for older adults seeking e-health information. J Gerontol Nurs 2007 Mar;33(3):6-12. [doi: 10.3928/00989134-20070301-03] [Medline: 17378186]
21. Czaja S, Lee C. Information technology and older adults. In: Sears A, Jacko JA, editors. Human-Computer Interaction: Designing for Diverse Users and Domains. New York, USA: CRC Press; 2009.
22. Fisk A, Rogers W, Charness N, Czaja SJ, Sharit J. Designing for Older Adults: Principles and Creative Human Factors Approaches, Second Edition. Boca Raton, FL: CRC Press; 2009.
23. Nielsen J. Usability Metrics. Nielsen Norman Group. 2001. URL: https://www.nngroup.com/articles/usability-metrics/ [accessed 2019-09-02]
24. National Institutes of Health. Making Your Website Senior Friendly. Senior Social Isolation. 2009. URL: https://seniorsocialisolation.ca/wp-content/uploads/2019/05/Making-Your-Website-Senior-Friendly-Tip-Sheet.pdf [accessed 2019-09-02]
25. International Organization for Standardization. Ergonomics of Human-system Interaction — Part 210: Human-Centred Design for Interactive Systems. International Organization for Standardization. 2010. URL: https://www.iso.org/standard/52075.html [accessed 2020-07-13]

26. International Organization for Standardization. Ergonomics of Human-System Interaction — Part 11: Usability: Definitions and Concepts. International Organization for Standardization. 2018. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2020-07-13]
27. Government of the United States of America. Federal Plain Language Guidelines. Plain Language. 2011. URL: https://plainlanguage.gov/media/FederalPLGuidelines.pdf [accessed 2019-09-02]
28. Lemieux V. Pour Qu'on Se Comprenne! Précautions Et Littératie en Santé. Réseau Santécom Koha. 2014. URL: http://www.santecom.qc.ca/bibliothequevirtuelle/Montreal/9782896733231.pdf [accessed 2019-09-02]
29. Hou S. Health literacy online: a guide to writing and designing easy-to-use health web sites. Health Promot Pract 2012 Sep;13(5):577-580. [doi: 10.1177/1524839912446480] [Medline: 22763891]
30. Simon D. The Art of Guerrilla Usability Testing. UX Booth. 2017. URL: https://www.uxbooth.com/articles/the-art-of-guerrilla-usability-testing/ [accessed 2019-09-02]
31. Nielsen J. Why You Only Need to Test with 5 Users. Nielsen Norman Group. 2000. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2019-09-02]
32. Nielsen J. Thinking aloud: The #1 usability tool. Nielsen Norman Group. 2012. URL: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/ [accessed 2019-09-02]
33. Hygiene 2.0 : Helps You or a Caregiver Select Bathroom Assistive Technology Suited to Your Needs. YouTube. 2020. URL: https://ergotherapie-outil-algo.espaceweb.usherbrooke.ca/demonstration [accessed 2020-06-22]
34. CEFRIO. Quelle Relation Les aînés Entretiennent-ils Avec Le Numérique en 2019? CEFRIO. 2018. URL: https://cefrio.qc.ca/fr/enquetes-et-donnees/netendances2018-vieillir-ere-numerique/ [accessed 2019-09-02]
35. World Health Organization. Monitoring and Evaluating Digital Health Interventions: A Practical Guide to Conducting Research and Assessment. Geneva, Switzerland: World Health Organization; 2016.
36. Chenel V, Mortenson WB, Guay M, Jutai JW, Auger C. Cultural adaptation and validation of patient decision aids: a scoping review. Patient Prefer Adherence 2018;12:321-332 [FREE Full text] [doi: 10.2147/PPA.S151833] [Medline: 29535507]

Abbreviations

e-DSS: electronic decision support system
H_2.0: Hygiene 2.0 (electronic prototype)
ISO: International Organization for Standardization
OT: occupational therapist
WHO: World Health Organization

Edited by G Eysenbach; submitted 13.09.19; peer-reviewed by H Wittink, H Young; comments to author 13.01.20; revised version received 28.04.20; accepted 03.06.20; published 10.08.20.

Please cite as:
Guay M, Latulippe K, Auger C, Giroux D, Séguin-Tremblay N, Gauthier J, Genest C, Morales E, Vincent C
Self-Selection of Bathroom-Assistive Technology: Development of an Electronic Decision Support System (Hygiene 2.0)
J Med Internet Res 2020;22(8):e16175
URL: https://www.jmir.org/2020/8/e16175
doi:10.2196/16175
PMID:32773380

Original Paper

A Federated Online Search Tool for Biospecimens (Sample Locator): Usability Study

Christina Schüttler[1], MSc; Verena Huth[2], MA; Magdaléna von Jagwitz-Biegnitz[2], MSc; Martin Lablans[3,4], PhD; Hans-Ulrich Prokosch[1], PhD; Lena Griebel[1], PhD

[1]Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[2]German Biobank Node, Charité – Universitätsmedizin Berlin, Berlin, Germany

[3]Federated Information Systems, German Cancer Research Center, Heidelberg, Germany

[4]University Medical Center Mannheim, Mannheim, Germany

Corresponding Author:
Christina Schüttler, MSc
Chair of Medical Informatics
Friedrich-Alexander-Universität Erlangen-Nürnberg
Wetterkreuz 13
Erlangen, 91058
Germany
Phone: 49 9131 85 26785
Email: christina.schuettler@fau.de

Abstract

Background: The German Biobank Alliance (GBA) aims to establish a cross-site biobank network. For this endeavor, the so-called Sample Locator, a federated search tool for biospecimens and related data, has been developed, forming the heart of its information technology (IT) infrastructure.

Objective: To ensure the sustainable use of such a tool, we included researchers as participants in an end user–based usability evaluation.

Methods: To develop a prototype ready for evaluation, we needed input from GBA IT experts. Thus, we conducted a 2-day workshop with 8 GBA IT team members. The focus was on the respective steps of a user-centered design process. With the acquired knowledge, the participants designed low-fidelity mock-ups. The main ideas of these mock-ups were discussed, extracted, and summarized into a comprehensive prototype using Microsoft PowerPoint. Furthermore, we created a questionnaire concerning the usability of the prototype, including the System Usability Scale (SUS), questions on negative and positive aspects, and typical tasks to be fulfilled with the tool. Subsequently, the prototype was pretested on the basis of this questionnaire with researchers who have a biobank background. Based on this preliminary work, the usability analysis was ultimately carried out with researchers and the results were evaluated.

Results: Altogether, 27 researchers familiar with sample requests evaluated the prototype. The analysis of the feedback certified a good usability, given that the Sample Locator prototype was seen as intuitive and user-friendly by 74% (20/27) of the participants. The total SUS score by the 25 persons that completed the questionnaire was 80.4, indicating good system usability. Still, the evaluation provided useful advice on optimization potential (eg, offering a help function).

Conclusions: The findings of this usability analysis indicate that the considerations regarding a user-friendly application that have been made in the development process so far strongly coincide with the perception of the study participants. Nevertheless, it was important to engage prospective end users to ensure that the previous development is going in the desired direction and that the Sample Locator will be used in the future. The user comments and suggestions for improvement will be considered in upcoming iterations for refinement.

(J Med Internet Res 2020;22(8):e17739) doi:10.2196/17739

KEYWORDS

software tools; biological specimen banks; user interface; evaluation; research

Introduction

To align with the overarching goal of improving patient care by strengthening medical research, increased efforts have recently been made to support the secondary use of data generated in the treatment context [1-4]. In Germany, initiatives such as the Medical Informatics Initiative [5] or the German Biobank Alliance (GBA) [6] have emerged to establish an appropriate infrastructure for this endeavor. Coordinated by the German Biobank Node (GBN), the GBA has taken up the challenge of creating a cross-location biobank network to support biospecimen-based research projects. To this end, it is essential not only to ensure the quality of biosamples and associated data, but also to provide tools to improve the sample request process. It is currently customary for biosamples to be requested directly at a biobank location or via a mailing list. The request for fitting samples is usually carried out using a paper form. In addition, the negotiation regarding the sample distribution usually takes place by telephone or via email. GBA strives to bundle these heterogeneous steps into a single information technology (IT) application and thus align the individual process steps of the biobanks. The resulting harmonization should enhance the sample request for the researcher by providing access to the biosamples available in the biobanks via a single point of contact by means of one inquiry request. For this purpose, the so-called Sample Locator forms the core of the federated search for biospecimens and associated data [7,8]. The Sample Locator is intended to enable researchers to send a request to all connected biobanks via a central web application. As a first step, the researcher can check the potentially available number of samples across all locations in a feasibility query. After a positive response, the user can log in and access further functions. This includes the detailed breakdown of the number of samples per biobank, the management of queries, and the establishment of a direct electronic contact to the relevant biobanks. Crucial factors for sustainable use include not only the technical aspects, but also the user-friendliness of such a tool. In order to ensure usability and meet users' needs, it is essential to involve end users in the development process. The reliance on the user-centered design (UCD) process has meanwhile proven itself in the development of IT applications in the medical (research) field [9]. A key aspect in the medical context is the avoidance of treatment errors due to lack of usability [10]. Even though the Sample Locator is not intended for application in the treatment context, a missing user-oriented approach can lead to the rejection of the tool. Therefore, its acceptance by the end users plays an influential role in our considerations.

The main objective of this paper is consequently to describe the evaluation of a midfidelity prototype in terms of its fitness for use. For the purpose of comprehensiveness, the paper also includes the necessary preparatory work, which covers the development of the prototype with the required functions and all steps of the request process. Since the focus of our work was to create a user-friendly search interface for the Sample Locator, we tested and assessed the usability of the prototype by end users, researchers familiar with sample requests. The resulting feedback will support the final tool development. An end user–based usability evaluation was used to ensure that the end users' opinions on the Sample Locator was taken into account in potential further development steps.

Methods

Preliminary Work

A rough sketch with the required functions of the query process served as a starting point for the development of an interactive prototype. We first examined the search interfaces of already existing tools for a similar scope of application in order to get a clearer picture of how such a search tool could look in its final version. Moreover, to elaborate the sketch towards a prototype ready for evaluation, the input of GBA IT experts was needed. Thus, we planned and conducted a 2-day workshop with 8 GBA IT members, 7 men and 1 woman. Three researchers with expertise in usability moderated the workshop, in which individual steps of a UCD process were pursued in 2 groups of 4 people each [11]. First, the participants discussed the typical context of use of the Sample Locator, resulting in the design of 2 hypothetical users (hereafter referred to as personas) representing potential end users. Second, the participants defined 2 example interaction designs. For this purpose, the groups were each assigned a specific use case to describe a typical usage sequence to be executed with the tool. The use cases were derived from real requests from researchers to biobanks, which were previously collected for a requirements analysis to determine the scope of performance of the Sample Locator. The use cases (adapted from the German version) can be found in Multimedia Appendix 1. With the acquired knowledge, the participants designed low-fidelity mock-ups using Balsamiq (Balsamiq Studios) [12]. The main ideas of these mock-ups were discussed, extracted, and summarized into a comprehensive midfidelity prototype. It was modeled using Microsoft PowerPoint (Microsoft Corp) [13]. The individual slides were linked together with hyperlinks, achieving an interactive navigation through the prototypical tool. Regarding the presentation of the user interface, the website layout of GBN [14] served as a template to harmonize the look with GBN's corporate design. Subsequently, the first draft of the prototype was pretested by 3 male researchers with a biobank background who are affiliated with GBA. In total, 3 versions of the Sample Locator prototype were created in the course of this work. Following the iterative nature of the design process, the versions are based on each other. The second version, which the questionnaire described below refers to, was used for the usability tests. This in turn resulted in the third and final prototype, which was enriched by feedback from the evaluation.

Study Design

Based on this preliminary work, the evaluation study was carried out. It was conducted as a usability analysis with female and male researchers aged between 18 and 67 years who work in a scientific institution in Germany. Prior to the start of the study, approval from the ethics committee of the Charité – Universitätsmedizin Berlin was obtained.

Recruitment

The study plan envisaged the recruitment of between 30 and 50 respondents to the survey over a period of 6 weeks. With this number of participants, a detection of 95% to 98% of problems within an application can be expected [15]. In order to address suitable participants, one contact person per GBN partner biobank (n=13) was determined in advance to personally approach potential end users with information material. The study team chose this intermediary approach to ensure the anonymity of the participants. Except for the required activity in the research environment, there were no further inclusion or exclusion criteria. Subsequently, the identified participants received an email from the contact person with a link to the online survey tool, LamaPoll (Langner/Maibaum/Notev GbR) [16]. A first introductory page served to briefly inform the participants about the study and to provide the prototype via a download link. After agreeing to participate by clicking a consent button, the participant was forwarded to the study questionnaire.

Instruments

The study questionnaire for the prototype evaluation consisted of 3 parts: (1) Sample Locator tasks and related questions on feasibility, (2) questions concerning the usability, and (3) general information.

In the first part, 4 tasks needed to be solved with the help of the provided prototype as a basis for subsequent questions. The first task was to search for samples of male patients with lung cancer. The others were to register to the Sample Locator, set project information, and refine the search query (lung metastasis samples plus excluding PAXgene-fixed tissue). The last task was to start negotiations with 2 selected biobanks. The tasks were designed to guide the test users through the prototypical system so they received insights into several possible functionalities. As with formulating the use cases for the workshop, these tasks were conceived based on the previously collected real inquiries to the biobanks. The first 6 questions enabled the researchers to evaluate the prototype's general intuitiveness and comprehensibility using a 5-point Likert scale (1=strongly disagree, 3=neutral, 5=strongly agree). The following 10 open questions aimed to elicit opinions on content and appearance of the individual steps and further comments on the prototype.

The second part contained the System Usability Scale (SUS), a widely applied and validated score for the quantitative measurement of the usability of an IT application [17]. The wording was slightly adapted by changing "system" to "application" to facilitate the comprehension of the scale.

The third part collected information on age and computer-handling characteristics with 3 questions.

The final questionnaire (adapted from the German version) can be found in Multimedia Appendix 2.

Data Analysis

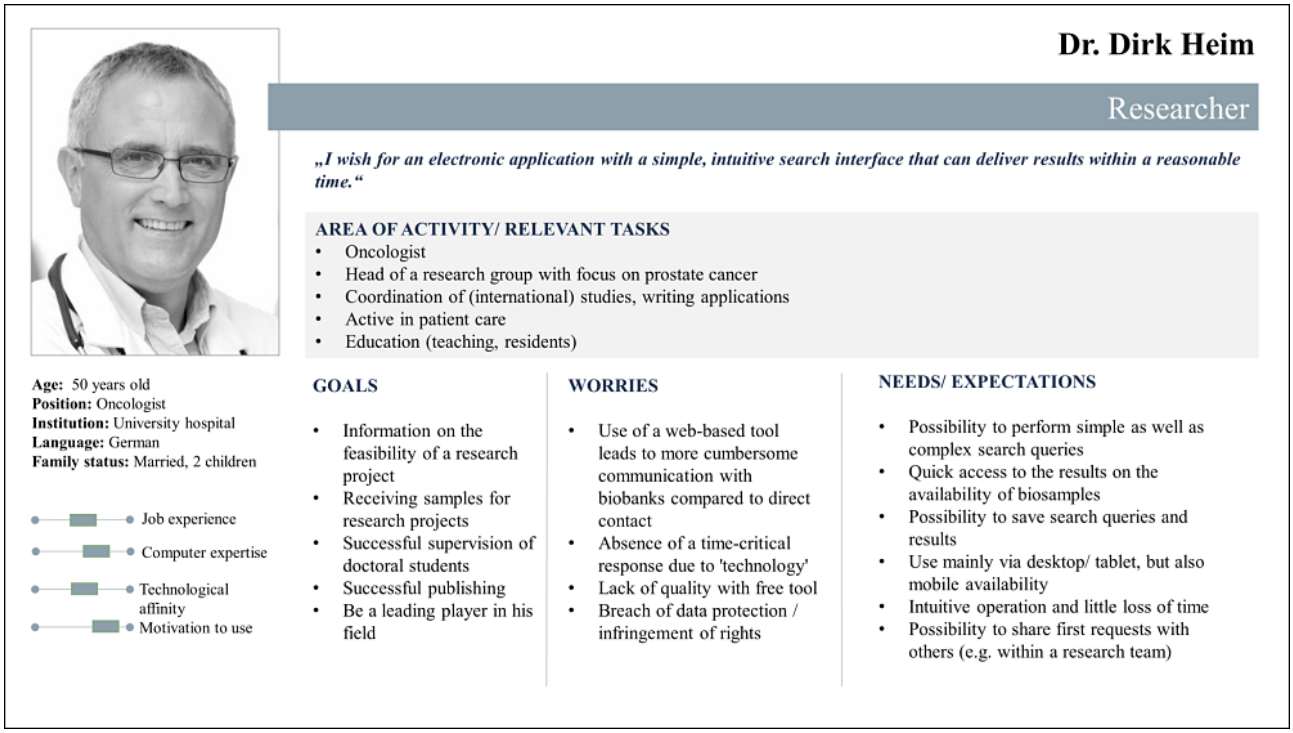
The data from the pretest were not subjected to a sophisticated analysis due to the small amount of data, so they were immediately examined and implemented by a scientist. Two scientists documented and analyzed the collected data for the usability assessment. The applied data analysis methods include a quantitative evaluation for the SUS score, the calculation of the average rating and the standard deviations for all closed questions, and a descriptive qualitative content analysis for open questions. The result of the qualitative content analysis was a categorized list of comments and suggestions for improvement by topic, which was then prioritized according to its practicability. This means that changes to the user interface can be directly incorporated into the final prototype, while technical aspects must first be discussed with the developers.

Results

Results From Preliminary Work

The workshop laid the groundwork for the following usability analysis. Initially, the 2 groups participating in the workshop each created one persona of a researcher. Since the results overlapped to a high degree, the 2 outcomes were combined into 1 model persona (Figure 1) for a better overview throughout the course of the workshop. The next step—the development of the interaction designs—resulted in 2 sketched workflows of different application scenarios according to the use cases provided to the workshop participants. Here, the participants' aim was to understand the essential steps of the process that were necessary for a successful trial of the respective use case and to visualize them in a flowchart (for the results, see Multimedia Appendix 3). Finally, the groups designed Sample Locator models by building on the previous steps, enabling the reproduction of the given use cases.

Figure 1. Persona of a researcher representing a future user of the Sample Locator.



The first version of the prototype was consequently created considering the graphically sketched specifications of the functionalities and the mock-ups developed during the workshop (for some impressions, see Multimedia Appendix 4). The second version was a revision based on the feedback from biobank-based reviewers from the pretest. However, the revisions were mainly individual imprecise formulations that were corrected for better comprehensibility.

Evaluation of the Prototype

A total of 27 participants complied with the call to respond to the survey and completed the questionnaire. Due to the recruitment setting, all participants were researchers who had already collaborated with biobanks and had experience in requesting samples. Thus, they represented typical end users of the Sample Locator. There were 6 persons aged 25 to 34 years, 18 persons aged 35 to 50 years, and 3 persons older than 50 years. Gender was not specifically surveyed in the study.

Sample Locator Tasks and Related Feasibility Questions

Most of the participants (20/27, 74%) agreed or fully agreed that the tasks were intuitively solvable using the prototype. Even more participants agreed or fully agreed that the display of the search process and the results were clearly arranged (22/27, 81%). The overall navigation was perceived as intuitive by approximately 74% (20/27) of the participants.

Participants particularly liked the absence of information overload and the prototype's clear structure, which was "reduced to the essential." Most testers confirmed this perception of clarity and structure for all of the process steps. Some mentioned that the mere prompt for International Classification of Diseases (ICD) codes for diagnoses was not usable "and it would be more convenient if the diagnoses...were behind the codes or if the list consisted entirely of plain text." Additionally, one participant stated that the AND/OR/NOT conjunctions to refine the search criteria were not self-explanatory. Another critical voice requested that the search interface should allow defining dependencies between sample parameters (eg, whether the removal date is before or after therapy). The intuitive and simple handling of the tool was highlighted as positive. However, it was also suggested to offer a kind of "query builder, where [users] can still customize the generated query script as power user," which allows more complex queries. Lastly, help functions were desired.

Questions About the Usability

Of the 27 participants, 25 (93%) completed the system usability questionnaire. The total SUS score of these 25 persons was 80.4, indicating good system usability.

The majority of the participants indicated that they would like to use the Sample Locator frequently (20/25, 80%). Additionally, most participants rated it easy to use (21/25, 84%). The complexity of the system was not seen as too high (21/25, 84% disagreed or fully disagreed that the complexity was too high). Regarding the need for technical support when using the Sample Locator, the results indicated a perceived high usability, as 23 of the 25 respondents (88%) disagreed or fully disagreed with the need for assistance.

The interface was seen as consistent (21/25, 84% disagreed or fully disagreed that the system was inconsistent). Most participants thought that their colleagues would easily learn to handle the Sample Locator (23/25, 92%), while 1 test person found the system hard to handle. Of the 25 participants, 21 indicated that they would feel confident using the Sample Locator. Nevertheless, 3 persons disagreed or fully disagreed with that. No participant indicated that he or she would need a lot of instruction before knowing how to use the Sample Locator.

General Information

All participants reported that they used a computer for fulfilling work tasks at least daily. Almost all persons (26/27, 96%) used a computer several times a day for their work. Of the 27 respondents, 10 persons (37%) indicated that they understood computers and computer technology very well, 15 persons (56%) understood computers well, and 2 (7%) self-assessed their understanding as sufficient.

Implementation of the Feedback in the Final Prototype

The third and final prototype was modified in accordance with the responses of the usability analysis with researchers. Figure 2 illustrates the initial search page with the basic functions of the prototype's third version. The search tool is divided into 2 sections. Search parameters can be selected in the left area of the screen. A distinction is made between sample-related criteria and donor-related criteria to obtain the most specific query possible. At this point, a help function has been added, as requested, which assists the user in finding the search parameters under the respective subheadings. Once the search criteria have been selected (eg, sex=male, as seen in Figure 3), the search can be performed by clicking the "Search" button. Here, a help function can now also be accessed to inform users about the format that can be searched for. A further implementation based on the test feedback when selecting the search parameters is the more detailed specification of the diagnostic codes. In addition to the simple code, the textual description has been added. In addition, the "direct entry of the ICD code (if already known) for input" is now possible via a free text field, as desired (Figure 4). The results will then be displayed in the right area of the screen. The feasibility query merely produces an aggregated number of potential samples and donor matches. As soon as the user has logged in to the tool in the next step, the allocation of suitable samples per biobank can be viewed (see Figure 5). In addition to an example query with detailed results for biobanks, Figure 5 also shows further functions. On the left "Search" side, the "Clear" button is used to completely restart a search, while the "Edit" button can be used to adjust the previous search. In order to have a better overview of which section you can currently interact with, a feature has been added that grays out and prevents editing within the inactive page. After the search has been carried out, one or more relevant biobanks can be selected on the right-hand side and contacted using the "Negotiate" button. The individual queries are saved and stored in a succeeding component. These can then be viewed and managed via the "Project Overview" button. In addition, general design changes have been introduced (eg, adjusting the font size and color to improve readability). In view of the intended international use, the content was also translated into English, per recommendation. For extensive screenshots of the Sample Locator prototype, see Multimedia Appendix 5.

Figure 2. Initial search page with the basic functions of the prototype.

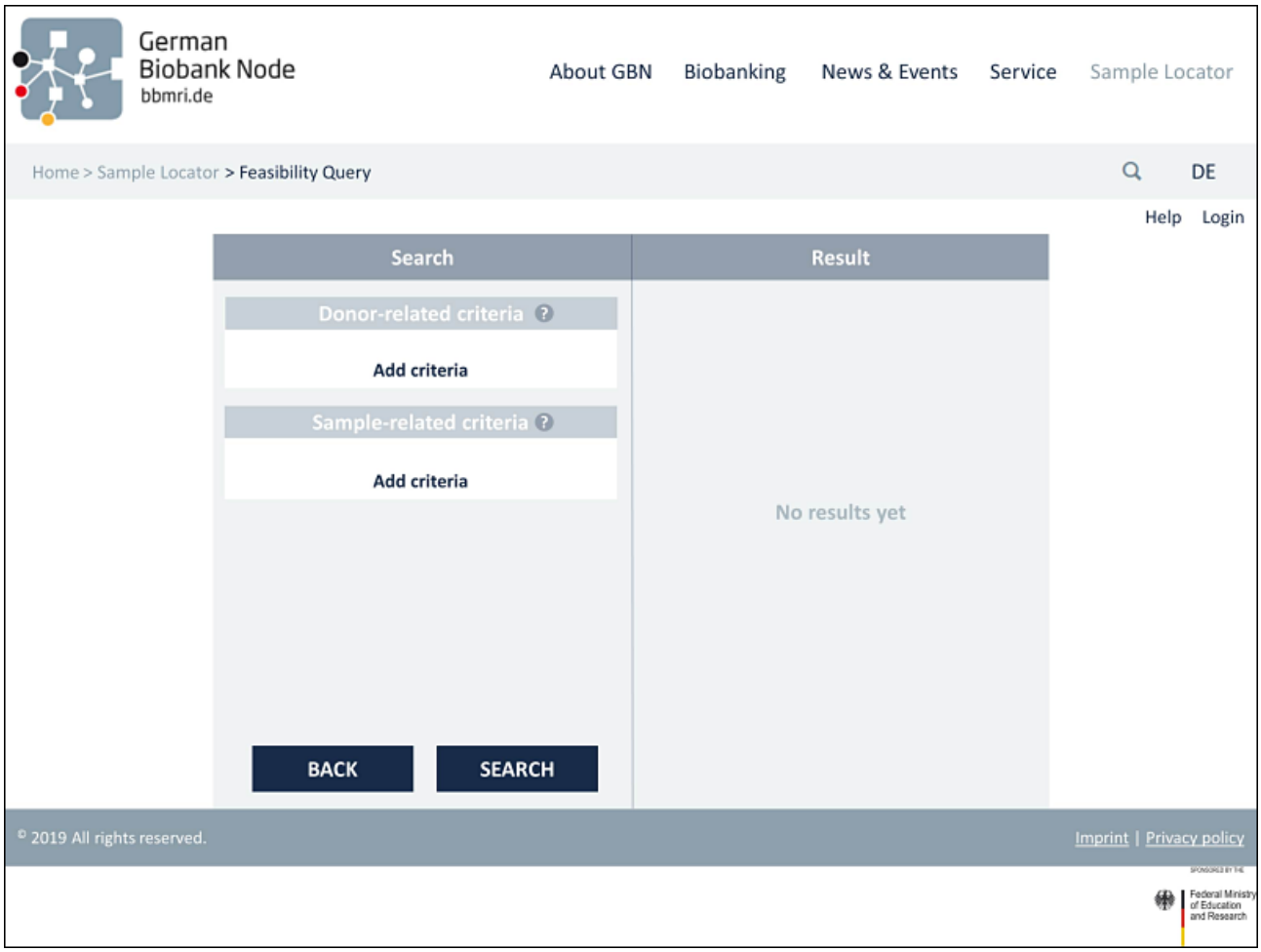


Figure 3. Illustration of the selection filters using the example of a patient with the sex parameter set to male.

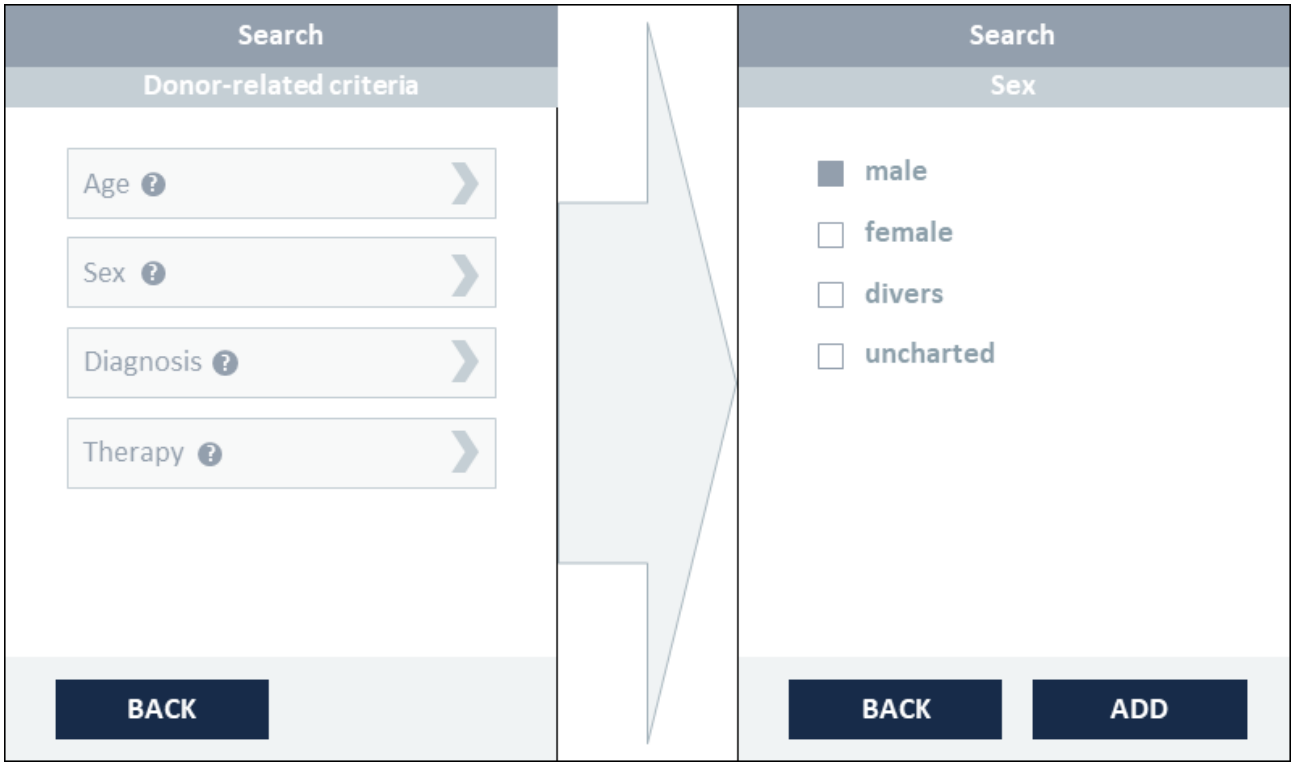


Figure 4. Continuation of the search for diagnoses by diagnosis classification and free text search.

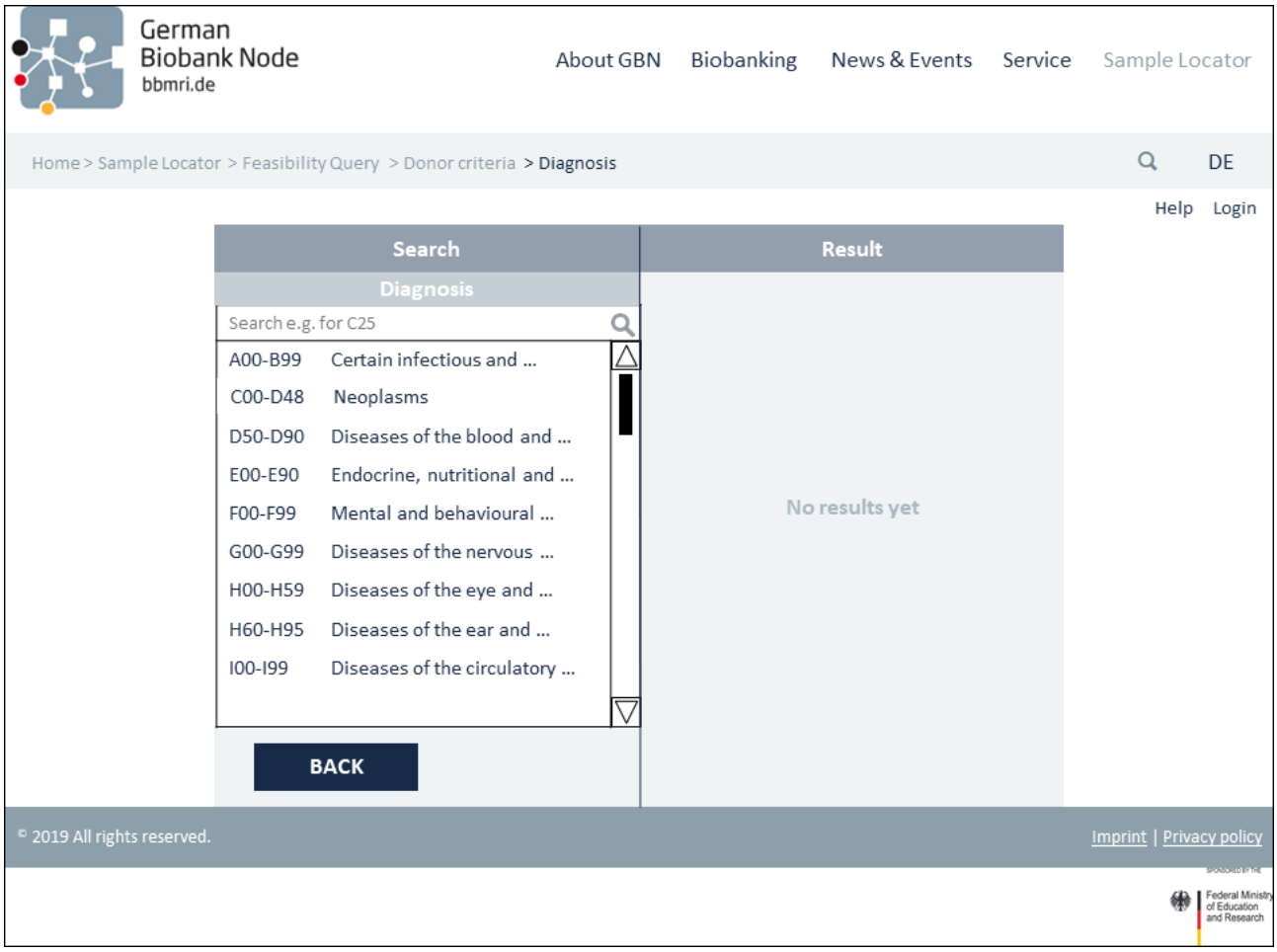
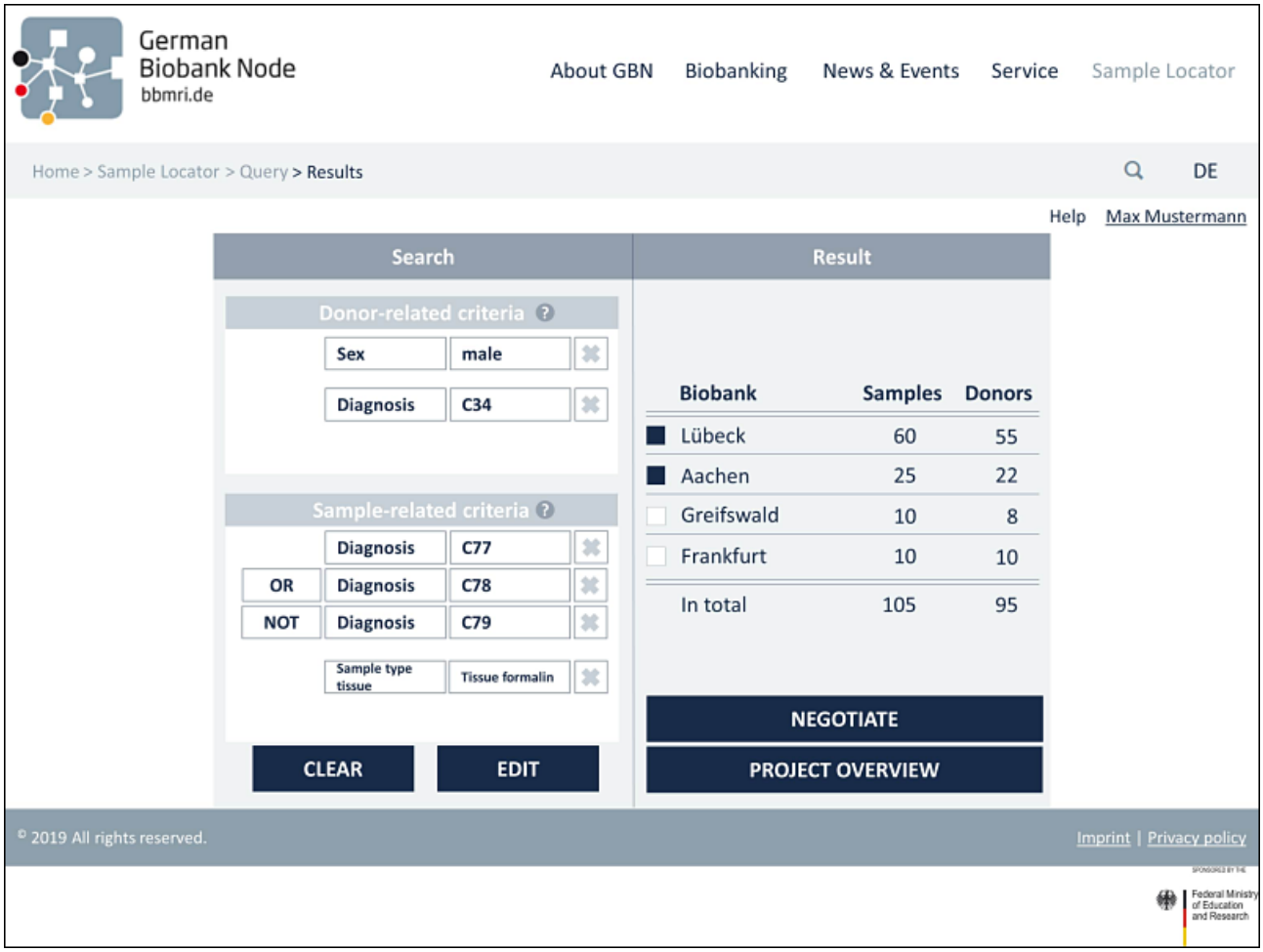


Figure 5. View of an example search query (left) with the corresponding detailed results per biobank (right).



Discussion

Overview

The main goal of our work was to evaluate the usability of a federated search tool for biosamples and associated data, the Sample Locator. This paper also illustrates the work that was accomplished in advance of the usability analysis. However, the evaluation of the tool by the end users was of particular importance. The predominantly positive feedback of the usability analysis indicated an intuitive and user-friendly operability. The answers to the open questions provided useful advice on optimization potential.

Comparison With Prior Work

In the area of medical research, especially with regard to cohort identification, several search tools are available. Commercial representatives include, among others, TriNetX [18], BC Platforms [19], and Clinerion [20]. Platforms or projects that make their work available as open source resources are i2b2 [21], tranSMART [22], and ConQuery [23]. The German Centre for Cardiovascular Research provides an in-house development for the feasibility query of biospecimens and data called the Feasibility Explorer [24]. The German Cancer Consortium (Deutsches Konsortium für Translationale Krebsforschung [DKTK]) has also implemented an in-house development (Searchbroker) as part of their comprehensive “bridgehead” architecture [8]. After reviewing these already existing tools, the deliberate decision was made to build on the DKTK codebase and their general architecture, but develop a customized, proprietary GBA Sample Locator solution. There were several reasons for this decision. Due to their developer power, industrial suppliers may have several advantages, particularly in the areas of speed and user interface design. However, especially when dealing with sensitive patient data, a solution that would be made available as an open source tool at the end of the project was preferred. In this case, all components can be developed, operated, and sustained by academic players, safeguarding the data sovereignty of biobanks. In this context, the Sample Locator, unlike tranSMART and the Feasibility Locator, follows a decentralized search approach. Consequently, the data remain locally at the biobank location, but can be queried centrally. Furthermore, we consider this open source approach to be a sensible way to ensure the continued existence of the Sample Locator beyond the end of the project [25]. In terms of further development, support, and bug fixing, the community resulting from this project (namely the Samply Community [26]) can continue to make a valuable contribution in the future. An additional functionality that distinguishes the Sample Locator from the search applications under review is that it limits the results visualization to the aggregated sample count from the whole GBA network for anonymous researchers. Only after a researcher’s authentication and log-in, the counts per biobank location are additionally displayed [6]. Further, after the search, the Sample Locator forwards the user directly

to the Negotiator, a dedicated communication tool to help request material from relevant biobanks and negotiate the terms involved based on the actual planned research project. Thus, based on the considerations outlined above, none of the existing tools were suitable for direct application and deployment in GBA. However, the various approaches provided valuable input for GBA's own search tool development in our incremental user-centered approach.

Discussion of Results

Preliminary Work

As already mentioned, the procedure described in this paper for preparing and conducting the evaluation study of the Sample Locator was oriented towards the UCD process outlined in ISO 9241-210 [11]. However, due to time constraints inherent to the project and the fact that work had already begun on a development version to drive progress on the technical back end of the search tool, the focus was on the steps most relevant to our objective of providing a user-friendly application in avoidance of undue delay. This comprised the creation of a persona, the identification of an interaction design, the creation of an interactive prototype, and the evaluation of the prototype. In this shortened procedure, the thorough analysis of the user group in particular was neglected. Instead, our work was based on the project specification that the main users of the Sample Locator will primarily be researchers. The requirements of the users were derived from the real biobank queries that were previously collected. The persona and the interaction design not only provided an important input for the creation of the prototype but also served as orientation over the entire development process in order to visualize the user and the process of the product. The preparation of the prototype and its evaluation were aimed specifically at the engagement of those end users. This abbreviated approach still allows us to incorporate the users' assessments in one of the following iteration steps in the further course of development.

Nonetheless, a limitation of our work might be that we used Microsoft PowerPoint as a tool for our prototype. Prior to creating the prototype, we also considered Balsamiq [12] and Axure [27] as alternative tools. Although these are dedicated tools for prototyping, we decided to use PowerPoint. On one hand, we wanted to have access to a tool that was as barrier free as possible. This applies to the licensing, handling, and distribution on our side as well as the use by the tester. PowerPoint is a widespread application that meets all our requirements in this respect, while the other tools would have required at least licensing and training to implement the project in the planned way. On the other hand, we did not strive to design a functioning, more realistic high-fidelity prototype. As a result, only the previously defined pathways resulting from the evaluation tasks were completely clickable. Although this presents a restricted interaction with the tool, this was sufficient for our purpose. The only constraint that became apparent in the course of implementation was the fact that there was not enough space to include more text, as the prototype was designed in typical PowerPoint slides. For example, ICD codes were not stored with textual descriptions, an aspect criticized by several test users. Apart from this limitation that prevented the test users from accurately experiencing the future tool's fully featured performance, PowerPoint was a suitable tool for achieving our goal. However, we accepted this limitation, since the prototype should only represent the main functions and be the basis for further developments. With our approach, we were able to receive profound insight into the usability of the general design of the Sample Locator (eg, regarding button arrangements).

Evaluation of the Prototype

Al-Ageel et al [28] performed a literature review and found that bioinformatics tools should be engineered with usability in mind to allow the development of intuitive interactions in order to achieve learnability and ease of use in the interfaces. Usability problems such as inconsistent navigation may even prevent users from completing their tasks [29].

With our task-based approach, using a midfidelity prototype supplemented by the use of the SUS among bioinformatics experts, we found that the Sample Locator was intuitive and easy to use and that there was no information overload. Most participants wanted to use the tool frequently and felt that the interface was consistent.

Bolchini et al [29] identified usability issues that are potentially relevant for bioinformatics resources. One important usability issue with bioinformatics tools is the search function, which potentially leads to usability barriers [29]. For the Sample Locator, such usability problems were identified as well. For instance, the AND/OR/NOT conjunctions for refining the search criteria were not perceived as self-explanatory by 1 study participant. Another participant stated that the search interface should allow the definition of dependencies between different sample parameters (eg, whether the collection date of a biospecimen is before or after therapy). Some participants mentioned that it was not practical to use only ICD codes to search for diagnoses and that it would be more convenient if the diagnoses were behind the codes. From this, we can generally draw the lesson that we should present all information in an immediately understandable way so that even laypersons would comprehend it directly. However, the critical feedback from individual test persons should be considered as well, as this feedback could also be perceived as usability barriers for future end users.

The findings of the usability evaluation indicate that the prototype represented the main functions in a suitable manner. Nevertheless, several limitations have to be taken into account when interpreting the results.

First, the usability survey was performed at a relatively late development stage. A UCD process normally aims to include potential end users' feedback in an early design phase [30]. Our approach was to assess the usability at a stage in which other attempts of designing the Sample Locator already existed. This unusual deferred approach was chosen because the first drafts of the Sample Locator focused more on technical possibilities, whereas our prototype's aim was to measure potential end users' opinions.

A further limitation inherent in our study is that we could only include 27 test users in our usability evaluation. This relatively

small number of participants does not lead to statistically significant results, but merely provides insight into the participants' opinions. However, our participants were experienced biobank users and thus could indeed provide reliable feedback. Moreover, it is not uncommon to use the SUS with a small number of participants. For example, Kersting and Weltermann [31] tested the usability of a software prototype supporting the management of multimorbid patients with 18 physicians. Likewise, Nielsen [32] stated that using 5 people in a study is sufficient for identifying almost all usability problems. We were able to gain valuable insights into test users' expectations and opinions by using several open questions. Thus, we are confident that we have been able to identify the majority of usability problems of the prototypical Sample Locator.

By slightly changing the wording in one item of the SUS, we changed a validated scale. Therefore, the validity of the scale might not be a given anymore. Nevertheless, we have accepted this risk to ensure that the scale was understandable for our test users.

Conclusion

The current lack of a suitable tool for a national search for biosamples prompted GBA to strive to close this gap. Although the initial focus was to provide a technically functioning development version at an early stage of the project, ensuring the user-friendliness of the tool should not be neglected in the process. To meet this demand, a usability analysis within the framework of a UCD process, which was also the basis for the accompanying preparatory work, was conducted. The findings of this usability analysis indicate that the considerations regarding a user-friendly application that have been made in the development process so far strongly coincide with the perception of the study participants. Despite this overlap, it was important to address the steps of the UCD process carried out in the context of this work and to engage prospective end users to ensure that the previous development was going in the desired direction and that the Sample Locator will be used in the future. Moreover, the users' comments and suggestions for improvement that were received through the open questions will be considered in upcoming iterations to refine the Sample Locator. In this way, we are confident in the delivery of a final product that offers added technical value but is also intuitive for the user. These results will also have a positive impact on the aspired Europe-wide rollout. In Europe, the Biobanking and Biomolecular Resources Research Infrastructure – European Research Infrastructure Consortium Directory is currently commonly used as an online search tool, listing 720 European biobanks with their 1621 individual collections [33]. The Directory provides a detailed overview of the collections' contents and enables users to contact a collection's principal investigator. However, the Sample Locator would offer added value on a more granular level, enabling the search for specific samples and the contacting of the respective biobanks. While the biobank of Masaryk University in Brno is already connected, the further integration of European biobanks is planned.

Acknowledgments

We sincerely thank all study participants for taking the time to give their highly appreciated feedback. Furthermore, we thank all workshop participants for providing valuable input.

The present work was performed in fulfillment of the requirements for obtaining the degree "Dr. rer. biol. hum." from the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) (CS).

This work is supported by the German Federal Ministry of Education and Research under the funding codes 01EY1701, 01EY1713, and 01EY1714.

Authors' Contributions

CS planned and conducted the usability workshop, assisted by Stefanie Schild and LG. CS designed, conducted, and assessed the usability evaluation. LG assisted during the design and evaluation phases. MvJB and VH advised in the preparation phase of the study. MvJB assisted in the collection of the evaluation data. CS wrote the main part of the paper and coordinated the inputs by the coauthors. LG participated in writing the paper. HUP, VH, MvJB, and ML revised the first draft and provided valuable input and comments. All authors read and approved the final manuscript. HUP was mainly responsible for the GBA project at the Chair of Medical Informatics of the Friedrich-Alexander-University Erlangen-Nürnberg.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Workshop use cases.
[PDF File (Adobe PDF File), 292 KB - jmir_v22i8e17739_app1.pdf]

Multimedia Appendix 2
Questionaire evaluation study.
[PDF File (Adobe PDF File), 802 KB - jmir_v22i8e17739_app2.pdf]

Multimedia Appendix 3
Interaction designs based on the provided use cases.
[PDF File (Adobe PDF File), 395 KB - jmir_v22i8e17739_app3.pdf]

Multimedia Appendix 4
Preliminary work: sketched functionalities and workshop mock-ups.
[PDF File (Adobe PDF File), 603 KB - jmir_v22i8e17739_app4.pdf]

Multimedia Appendix 5
Screenshots of the prototype.
[PDF File (Adobe PDF File), 691 KB - jmir_v22i8e17739_app5.pdf]

References

1. European Health Data & Evidence Network. EHDEN. URL: https://www.ehden.eu/
2. Observational Health Data Sciences and Informatics. OHDSI. URL: https://www.ohdsi.org/
3. Swiss Personalized Health Network. SPHN. URL: https://www.sphn.ch/
4. Health Research Infrastructure. Health RI. URL: http://www.health-ri.org/
5. Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med 2018 Jul;57(S 01):e50-e56 [FREE Full text] [doi: 10.3414/ME18-03-0003] [Medline: 30016818]
6. Schüttler C, Buschhüter N, Döllinger C, Ebert L, Hummel M, Linde J, et al. [Requirements for a cross-location biobank IT infrastructure : Survey of stakeholder input on the establishment of a biobank network of the German Biobank Alliance (GBA)]. Pathologe 2018 Jul;39(4):289-296 [in German]. [doi: 10.1007/s00292-018-0435-9] [Medline: 29691676]
7. Proynova R, Alexandre D, Lablans M, Van Enckevort D, Mate S, Eklund N, et al. A Decentralized IT Architecture for Locating and Negotiating Access to Biobank Samples. Stud Health Technol Inform 2017;243:75-79. [Medline: 28883174]
8. Lablans M, Kadioglu D, Muscholl M, Ückert F. Exploiting Distributed, Heterogeneous and Sensitive Data Stocks while Maintaining the Owner's Data Sovereignty. Methods Inf Med 2015;54(4):346-352. [doi: 10.3414/ME14-01-0137] [Medline: 26196653]
9. Schild S, Sedlmayr B, Schumacher A, Sedlmayr M, Prokosch HU, St Pierre M, German Cognitive Aid Working Group. A Digital Cognitive Aid for Anesthesia to Support Intraoperative Crisis Management: Results of the User-Centered Design Process. JMIR Mhealth Uhealth 2019 Apr 29;7(4):e13226 [FREE Full text] [doi: 10.2196/13226] [Medline: 31033445]
10. Hallbeck MS, Koneczny S, Büchel D, Matern U. Ergonomic usability testing of operating room devices. Stud Health Technol Inform 2008;132:147-152. [Medline: 18391276]
11. ISO 9241-210:2019 Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. Geneva, Switzerland: International Organization for Standardization; 2019. URL: https://www.iso.org/standard/77520.html [accessed 2019-03-27]
12. Balsamiq Cloud Web App. Balsamiq Studios. 2017. URL: https://balsamiq.com/wireframes/cloud/
13. Microsoft PowerPoint. Microsoft Corp. 2013. URL: https://products.office.com/powerpoint
14. German Biobank Node. URL: https://www.bbmri.de/?L=1 [accessed 2019-04-17]
15. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [Medline: 14587545]
16. LamaPoll [Software]. Lamano GmbH & Co. KG. URL: https://www.lamapoll.de/
17. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry. London, England: Taylor and Francis; 1996:189-194.
18. Topaloglu U, Palchuk MB. Using a Federated Network of Real-World Data to Optimize Clinical Trials Operations. JCO Clin Cancer Inform 2018 Dec;2:1-10 [FREE Full text] [doi: 10.1200/CCI.17.00067] [Medline: 30652541]
19. BC Platforms. BC Platforms Ltd. URL: https://www.bcplatforms.com/
20. Clinerion. Clinerion Ltd. URL: https://www.clinerion.com/
21. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 2010;17(2):124-130 [FREE Full text] [doi: 10.1136/jamia.2009.000893] [Medline: 20190053]
22. Athey BD, Braxenthaler M, Haas M, Guo Y. tranSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research. AMIA Jt Summits Transl Sci Proc 2013;2013:6-8 [FREE Full text] [Medline: 24303286]
23. Walker J, Hegner M, Priehn H. ConQuery - eine Open-Source Software zur einfachen und schnellen Analyse großer Datenbestände. Düsseldorf, Germany: German Medical Science GMS Publishing House; 2018 Presented at: 63rd annual meeting of the German Society for Medical Informatics, Biometry and Epidemiology eV (GMDS); Sep 2-6, 2018; Osnabrück, Germany p. 270.

24. Scheel H, Dathe H, Franke T, Scharfe T, Rottmann T. A Privacy Preserving Approach to Feasibility Analyses on Distributed Data Sources in Biomedical Research. Stud Health Technol Inform 2019 Sep;267:254-261. [doi: 10.3233/SHTI190835] [Medline: 31483280]
25. Đurković J, Vuković V, Raković L. Open Source Approach in Software Development - Advantages and Disadvantages. Management Information Systems 2008;3(2):029-033 [FREE Full text]
26. Samply Community. URL: https://github.com/samply [accessed 2020-03-10]
27. Axure [Software]. Axure Software Solutions. URL: https://www.axure.com/
28. Al-Ageel N, Al-Wabil A, Badr G, AlOmar N. Human Factors in the Design and Evaluation of Bioinformatics Tools. Procedia Manufacturing 2015;3:2003-2010 [FREE Full text] [doi: 10.1016/j.promfg.2015.07.247]
29. Bolchini D, Finkelstein A, Perrone V, Nagl S. Better bioinformatics through usability analysis. Bioinformatics 2009 Feb;25(3):406-412. [doi: 10.1093/bioinformatics/btn633] [Medline: 19073592]
30. Martin B, Hanington B. Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Beverly, MA: Rockport Publishers; 2012.
31. Kersting C, Weltermann B. Evaluating the Feasibility of a Software Prototype Supporting the Management of Multimorbid Seniors: Mixed Methods Study in General Practices. JMIR Hum Factors 2019 Jul;6(3):e12695 [FREE Full text] [doi: 10.2196/12695] [Medline: 31274115]
32. Nielsen J. How many test users in a usability study? Nielsen Norman Group. 2012 Jun 03. URL: https://www.nngroup.com/articles/how-many-test-users/
33. Holub P, Swertz M, Reihs R, van Enckevort D, Müller H, Litton J. BBMRI-ERIC Directory: 515 Biobanks with Over 60 Million Biological Samples. Biopreserv Biobank 2016 Dec;14(6):559-562 [FREE Full text] [doi: 10.1089/bio.2016.0088] [Medline: 27936342]

Abbreviations

DKTK: Deutsches Konsortium für Translationale Krebsforschung (German Cancer Consortium)
GBA: German Biobank Alliance
GBN: German Biobank Node
ICD: International Classification of Diseases
IT: information technology
SUS: System Usability Scale
UCD: user-centered design

Edited by G Eysenbach; submitted 13.01.20; peer-reviewed by M Tremblay, B Schreiweis; comments to author 23.03.20; revised version received 24.04.20; accepted 14.06.20; published 18.08.20.

Please cite as:
Schüttler C, Huth V, von Jagwitz-Biegnitz M, Lablans M, Prokosch HU, Griebel L
A Federated Online Search Tool for Biospecimens (Sample Locator): Usability Study
J Med Internet Res 2020;22(8):e17739
URL: http://www.jmir.org/2020/8/e17739/
doi:10.2196/17739
PMID:32663150

Original Paper

Using Web-Based Social Media to Recruit Heavy-Drinking Young Adults for Sleep Intervention: Prospective Observational Study

Garrett I Ash[1,2], PhD; David S Robledo[1], BS; Momoko Ishii[1,3], BS; Brian Pittman[1], MS; Kelly S DeMartini[1], PhD; Stephanie S O'Malley[1], PhD; Nancy S Redeker[4], PhD, RN; Lisa M Fucito[1,5,6], PhD

[1]Department of Psychiatry, Yale School of Medicine, New Haven, CT, United States

[2]Pain, Research, Informatics, Medical Comorbidities and Education Center (PRIME), Veterans Affairs Connecticut Healthcare System, West Haven, CT, United States

[3]Yale School of Public Health, New Haven, CT, United States

[4]Yale School of Nursing, New Haven, CT, United States

[5]Yale Cancer Center, New Haven, CT, United States

[6]Smilow Cancer Hospital, Yale-New Haven Hospital, New Haven, CT, United States

Corresponding Author:
Garrett I Ash, PhD
Department of Psychiatry
Yale School of Medicine
300 George Street, #901
New Haven, CT, 06511
United States
Phone: 1 2034443079
Email: garrett.ash@yale.edu

Abstract

Background: Novel alcohol prevention strategies are needed for heavy-drinking young adults. Sleep problems are common among young adults who drink heavily and are a risk factor for developing an alcohol use disorder (AUD). Young adults, interested in the connection between sleep and alcohol, are open to getting help with their sleep. Therefore, sleep interventions may offer an innovative solution. This study evaluates social media advertising for reaching young adults and recruiting them for a new alcohol prevention program focused on sleep.

Objective: This study aims to evaluate the effectiveness and cost of using Facebook, Instagram, and Snapchat advertising to reach young adults who drink heavily for a sleep intervention; characterize responders' sleep, alcohol use, and related concerns and interests; and identify the most appealing advertising content.

Methods: In study 1, advertisements targeting young adults with sleep concerns, heavy alcohol use, or interest in participating in a sleep program ran over 3 months. Advertisements directed volunteers to a brief web-based survey to determine initial sleep program eligibility and characterize the concerns or interests that attracted them to click the advertisement. In study 2, three advertisements ran simultaneously for 2 days to enable us to compare the effectiveness of specific advertising themes.

Results: In study 1, advertisements generated 13,638 clicks, 909 surveys, and 27 enrolled volunteers in 3 months across the social media platforms. Fees averaged US $0.27 per click, US $3.99 per completed survey, US $11.43 per volunteer meeting initial screening eligibility, and US $106.59 per study enrollee. On average, those who completed the web-based survey were 21.1 (SD 2.3) years of age, and 69.4% (631/909) were female. Most reported sleep concerns (725/909, 79.8%) and an interest in the connection between sleep and alcohol use (547/909, 60.2%), but few had drinking concerns (49/909, 5.4%). About one-third (317/909, 34.9%) were identified as being at risk for developing an AUD based on a validated alcohol screener. Among this subsample, 8.5% (27/317) met the final criteria and were enrolled in the trial. Some volunteers also referred additional volunteers by word of mouth. In study 2, advertisements targeting sleep yielded a higher response rate than advertisements targeting alcohol use (0.91% vs 0.56% click rate, respectively; $P<.001$).

Conclusions: Social media advertisements designed to target young adults with sleep concerns reached those who also drank alcohol heavily, despite few being concerned about their drinking. Moreover, advertisements focused on sleep were more effective than those focused on drinking. Compared with previous studies, cost-effectiveness was moderate for engagement (impressions

to clicks), excellent for conversion (clicks to survey completion), and reasonable for enrollment. These data demonstrate the utility of social media advertising focused on sleep to reach young adults who drink heavily and recruit them for intervention.

(J Med Internet Res 2020;22(8):e17449) doi:10.2196/17449

KEYWORDS

substance abuse; social media; alcohol drinking; sleep; mobile phone

Introduction

Background

Heavy alcohol use remains a problem among young adults. Alcohol use disorder (AUD) onset peaks during young adulthood (ie, 18-25 years) [1]. Compared with older adults, young adults report more frequent and heavier alcohol use that is linked to substantial negative consequences including the risk of accidental injury, the primary cause of death among young adults [2,3]. Current alcohol intervention strategies for young adults have modest effects [4-6], and young adults rarely self-identify for specialized alcohol treatment [7,8]. Thus, more research is needed to identify effective alcohol interventions and novel treatment engagement strategies to reduce this substantial public health burden.

Concurrent to their high rates of alcohol consumption, research suggests that many heavy-drinking young adults have poor sleep. In this age group, greater alcohol consumption and alcohol-related consequences are associated with shorter sleep duration, poorer sleep quality, and more delayed bed/wake times [9,10]. In addition, various sleep problems in adolescence predict earlier AUD onset and greater risk of heavy drinking, alcohol-related consequences, and AUD in young adulthood [11-13]. Furthermore, poor sleep in young adults predicts a greater future risk of alcohol-related problems [14].

Although most heavy-drinking young adults do not seek alcohol treatment [7,8], sleep interventions may be a means to identify those who drink heavily and engage them in treatment. Our previous formative work showed that young adults are interested in the connection between sleep and alcohol and are open to getting help for their sleep [15]. Thus, sleep could be a novel intervention target to engage young adults and promote alcohol behavior change [15,16].

The promise of this approach raises the question of the best vehicle for recruitment. Young adults visit health care providers in person less often [17] but go on the web for health information and use mobile health apps more often as compared to adolescents [18]. Among these mobile health apps, sleep is the third most popular topic, following fitness and nutrition [18]. Moreover, heavy-drinking young adults may prefer mobile sleep interventions over in-person models [15]. Thus, considering young adults' health practices and preferences, one possible recruitment strategy is to utilize technology.

Previous research suggests that social media is a cost-effective method for recruiting young adults [19-26] as well as hard-to-reach populations and those affected by specific physical or mental health conditions [27-29]. Urban women at high risk of HIV were recruited into an HIV prevention intervention [23], young adult smokers were recruited into a smoking cessation intervention [26], and African American women with elevated blood pressure were recruited into a physical activity and nutrition intervention [28]. The majority of studies evaluated Facebook for this purpose [19,20,23,25-29]. Other social media platforms, such as Instagram and Snapchat, have emerged and are highly popular among young adults [30], but few studies have researched them as recruitment tools [21,22].

This is an important research gap because the platforms differ both practically and theoretically. The most salient practical differences include the following: (1) Instagram and Snapchat are image based with minimal text, whereas Facebook uses both images and text, (2) Snapchat immediately deletes content after viewing, whereas Instagram and Facebook archive it, and (3) Snapchat restricts interaction with users with an existing relationship (ie, have exchanged screen names), whereas Instagram and Facebook allow public interactions. The uses and gratifications theory, which rests on the premise that consumers actively choose media with the intention of fulfilling specific needs, has unveiled some ramifications of these differences among young adults. First, image-based social media (Instagram and Snapchat) led to decreased loneliness and increased happiness and satisfaction with life, whereas Facebook and other text-based media did not [31]. Second, Snapchat led to deeper levels of personal disclosure than other platforms [32-34]. A qualitative analysis suggested this could be due to Snapchat's immediate deletion of communications assuaging user apprehension about their future ramifications and Snapchat's restriction of communication to existing relationships facilitating more trusting connections than the communications with strangers allowed on other platforms [32-34]. It is not known whether these differing uses and gratifications between the platforms lead to differing utility for research recruitment.

Objectives

The goal of this study is to evaluate social media advertising to recruit young adults who report heavy alcohol use for a mobile sleep intervention. The aims are to evaluate the effectiveness and cost of using Facebook, Instagram, and Snapchat advertising to reach young adults with sleep concerns and a subpopulation with heavy drinking behaviors; characterize sleep, alcohol use, and related concerns and interests among responders; and identify advertising content that yielded the highest response rate. These data may suggest new ideas for recruitment or innovative mobile interventions to address poor sleep and heavy drinking among young adults.

Methods

The methods described below are based on our previous investigation of social media advertising for a different population (heavy-drinking smokers) [35]. Our previous study

and this study shared the primary objective of evaluating the effectiveness and cost of using social media advertising to recruit from the population of interest.

Screening Process Overview

The target population of the web survey was young adult (aged 18-25 years) heavy drinkers at risk for AUD (defined below, Drinking Habits) who were interested in participating in a sleep intervention (trial registration: NCT036589); lived in the greater New Haven, Connecticut area (to complete in-person visits); owned a smartphone (to complete the mobile intervention); and were literate in English. Research staff contacted volunteers who met these preliminary criteria to verify survey-reported drinking patterns and screened out exclusion criteria: confounders of circadian rhythm (night shift work or >2 time zones travel in the past month or next 3 months), severe AUD, or unsafe to complete transdermal ankle alcohol monitoring (eg, peripheral vascular condition). Eligible and interested volunteers were invited to an intake visit at our research lab. Although the survey has been ongoing, we have restricted our analysis to (1) all responses collected for the first 3 months (January 14 to April 18, 2019) to describe the population of responders and the overall campaign effectiveness and (2) responses to specific advertisements over a short trial period (October 5 and 6, 2019) to compare the effectiveness of advertising themes. The study and screening process were approved by the Yale University Institutional Review Board.

Study 1: Description of the Population of Responders and Overall Campaign Effectiveness

Facebook, Instagram, and Snapchat Recruitment

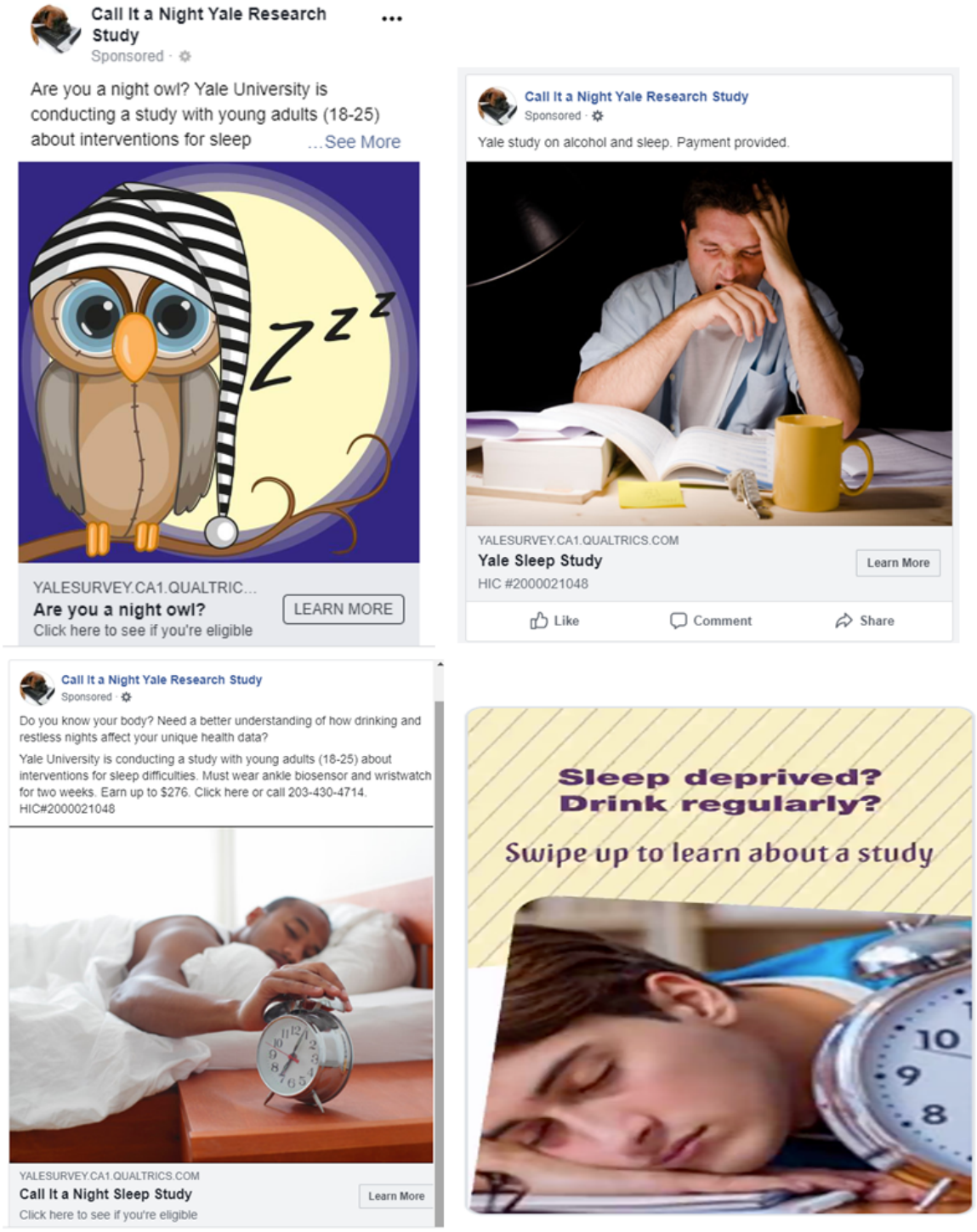
We ran advertisements through Facebook, Instagram, and Snapchat paid advertising services for 41 days over a 3-month period between January 14 and April 18, 2019. Our campaign included 4 advertisements appearing on the Facebook, Instagram (desktop and mobile), and Snapchat (mobile) interfaces of individuals in our age group (18-25 years). Advertisements were turned on, each for a period of days, until we enrolled our target number of volunteers for that study period (n=32). We specified a spending limit of US $25 per day for Facebook and Instagram (combined) and US $50 per day for Snapchat (Snapchat's minimum).

We restricted the geographic radius to a range that made travel to New Haven feasible without compromising the daily number of times the advertisement was displayed (ie, impressions). The advertising platforms reported a greater number of users within 25 miles of New Haven for Snapchat (383,000) than for Facebook and Instagram combined (220,000), which was counterbalanced by selecting a narrower radius for Snapchat (10 miles) than for Facebook and Instagram (25 miles). The 4 advertisements featured various combinations of the following 3 themes: (1) sleep concern (eg, "sleep deprived?"), (2) drinking behavior (eg, "drink regularly?") although not drinking concern, and (3) health behavior intersection (eg, "need a better understanding of how drinking and restless nights affect your unique health data?"; Figure 1). Images featuring only men, which have been previously reported to yield higher conversion rates than advertisements featuring women [36]. Wording style was taken from our previous successful social media campaign [16].

Facebook and Instagram run on a shared platform, and Snapchat runs on its own platform. Facebook and Snapchat allocate advertising space using an *auction* process based on the spending *bid* of the advertiser, relevance to the user (ie, web analytic estimated rate of the user acting upon the advertisement), and advertisement quality (ie, past user experience survey results) [37,38]. We used the bid-optimizing algorithms offered by Facebook and Snapchat targeting the lowest cost per click. Facebook's auctioning and bid optimization include the Instagram space.

Each platform monitored the number of impressions, total reach (ie, number of people seeing the advertisements), advertisement clicks, and total cost for all advertisements. These data allowed us to evaluate efficacy and cost-effectiveness.

Figure 1. Examples from advertising campaign (study #1) designed to reach young adults with heavy drinking for a sleep intervention.



Survey Procedures

The Yale Institutional Review Board approved this protocol. Study advertisements could be clicked on Facebook and Instagram or *swiped up* on Snapchat. By clicking on Facebook or Instagram advertisements, users were directed to a web-based survey to screen for initial eligibility. By *swiping up* on Snapchat, users were connected to our study website, which provided a description of the study [39] and a link to the same eligibility survey. Snapchat did not allow a direct link to the web-based survey because it found that our privacy policy lacked enough explanation to be directly linked from an advertisement. The eligibility survey was administered by a HIPAA-compliant web interface (Qualtrics). It began with an overview of the study: (1) we are looking for young adults who want to improve their sleep and drink alcohol, (2) we are testing different mobile health strategies for improving sleep, and (3) the procedures include web-based sleep education along with sleep and alcohol biosensors and diaries. It then provided information about study compensation (US $276), funding source, research team contact details, and confidentiality. Volunteers were informed that they had the option to complete screening by phone instead of using the web (none utilized).

Before completing the survey, volunteers were asked to affirm that they were older than 18 years of age, understood the information, and wished to proceed. Volunteers were then encouraged to provide their email addresses and phone numbers to be contacted if eligible.

Survey Items

The survey included 29 questions that took approximately 5-10 min to complete. "I choose not to answer" was an option for all multiple-choice items.

Study Interest and Referral Source

The initial questions asked how volunteers heard about the study and what made them interested in participating. Volunteers could select one or more options among the multiple options (concerned about sleep, concerned about drinking, interest in sleep, interest in alcohol drinking, interest in the connection between sleep and alcohol drinking, and others).

Demographics

Volunteers were asked about their age and sex.

Basic Exclusion Criteria

Volunteers were asked if they owned a smartphone and whether they could read and write in English.

Alcohol Quantity and Frequency

Volunteers were asked about their quantity of drinking (average standard drinks consumed on each day of the week) over the past 30 days [40] and frequency of any heavy drinking (≥5 drinks among men vs ≥4 drinks among women) over the past 14 days. They were advised to report in standard drinks (12 ounces of beer, 5 ounces of wine, and 1.5 ounces of hard liquor shot or mixed drink).

AUD Risk Status

Volunteers also completed the consumption questions of the Alcohol Use Disorders Identification Test (AUDIT-C): frequency of any drinking, typical drinks per occasion, and frequency of ≥6 drinks. The AUDIT-C scores of ≥7 for men and ≥5 for women have strongly distinguished at-risk drinking for young adults compared with reference standards (area under receiver operating characteristic curve=0.89) [41]. Volunteers who met the AUDIT-C threshold were deemed preliminarily eligible. Final eligibility was verified by telephone. Volunteers had to report ≥3 episodes of heavy drinking in the past 14 days.

Sleep

Volunteers rated their sleep quality over the past 7 days using a single-item Likert-type scale with anchors 1="very poor" and 5="very good." This single item correlates with the Patient-Reported Outcomes Measurement System Sleep Disturbance Bank (θ>0.85), which converges with the Pittsburgh Sleep Quality Index (r=0.85) [42]. They were also asked if they were concerned about their sleep (yes or no).

Statistical Analysis

Incomplete surveys or those with duplicate contact information were removed. Histograms of all variables were inspected. All drinking variables were negatively skewed with a high number of *0* cases, but repeating the analysis with these cases excluded did not alter the statistical outcomes.

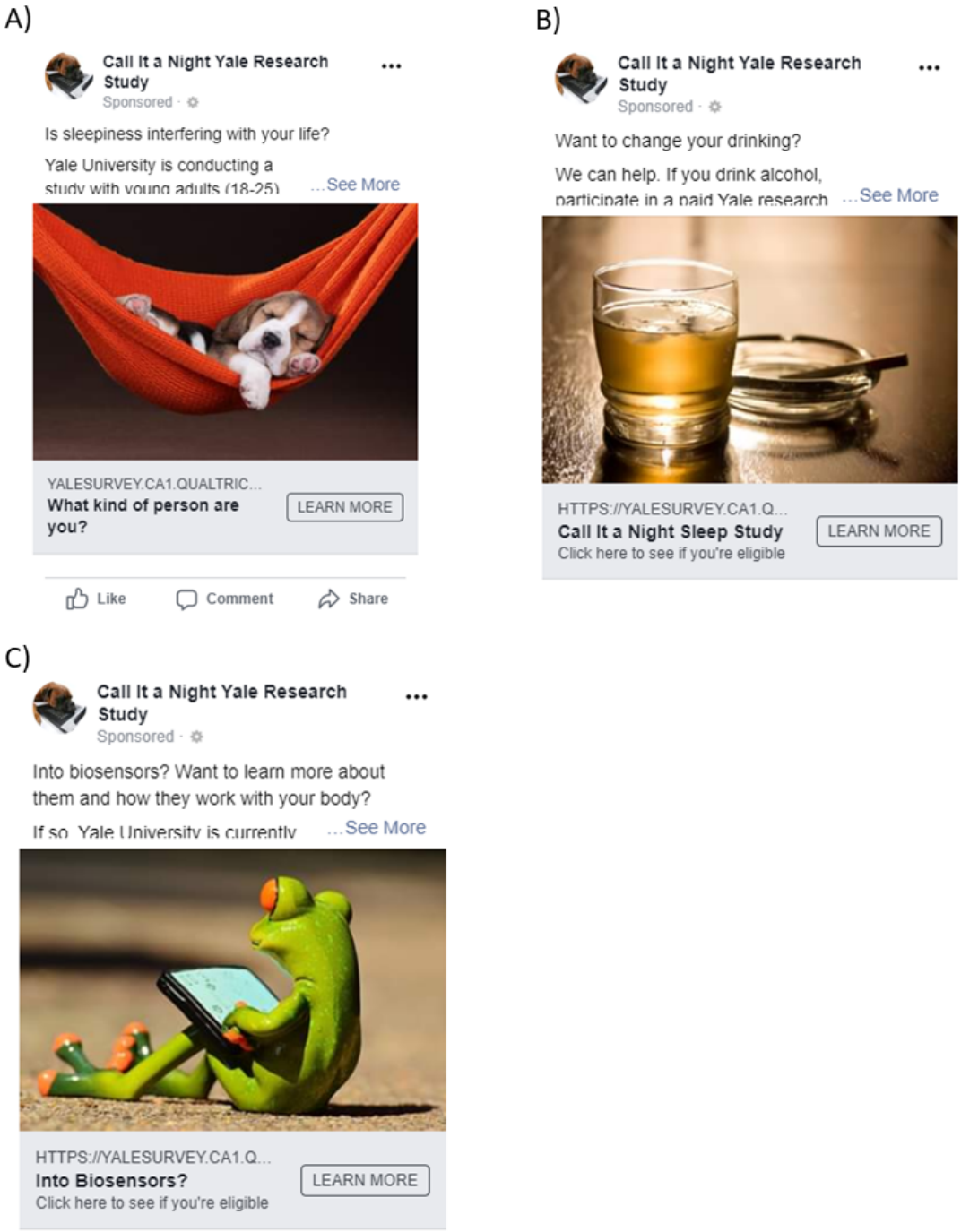
Descriptive statistics (frequencies, means, and standard deviations) were calculated for all completed surveys and then separately for (1) ineligible web surveys, (2) eligible web surveys not enrolled in the study (phone screen failures or not interested), and (3) eligible surveys that were enrolled in the study. These 3 groups were then compared using the analysis of covariance for continuous variables (ie, age, sleep quality, AUDIT-C, heavy drinking episodes per 14 days, drinks per week) and chi-square for categorical variables (ie, social media platform referral source, each possible reason for interest, sex, brand of smartphone). Similar analyses of covariance were also performed comparing participants with respect to (1) referral from Snapchat versus the other platforms (to investigate sample features possibly contributing to the lower advertising efficiency of Snapchat that we observed) and (2) sex (to investigate sample features affected by the higher preponderance of women than men in our final sample). Analyses were conducted using a significance level of α<.05, and adjusted for age and sex where significant. We applied the Bonferroni correction to post hoc pairwise comparisons. The analysis of AUDIT-C data did not include the ineligible web survey group, because they were all below the high-risk cutoffs (≥5 women and ≥7 men), and the other groups were all at or above these cutoffs.

Study 2: Comparing Effectiveness Between Advertising Themes

Advertising Experiment

We generated one content-specific advertising set for each theme represented in the advertisements in study 1 (see the section *Facebook, Instagram, and Snapchat Recruitment* from study 1) so that we could isolate what was most effective (Figure 2): (1) *sleep* advertisements focused solely on sleep concerns without mentioning alcohol (eg, "Is sleepiness interfering with your life?"), (2) *alcohol* advertisements focused solely on alcohol concerns without mentioning sleep (eg, "Want to change your drinking?"), and (3) *biosensor and health* advertisements focused on the concept of multiple health behavior intersections without mentioning sleep or alcohol (eg, "Into biosensors? Want to learn more about them and how they work with your body?"). The *biosensor and health* set also served as a control arm accounting for attraction to health information, thus isolating the impact of attraction to the specific health topics of sleep and alcohol consumption. We uploaded them to Facebook/Instagram for 48 hours (October 5 and 6, 2019). Each advertisement set was displayed to users by Facebook's auctioning and bid-optimizing process with an equal spending limit for each set (US $35 per day). Within each set, Facebook chose a specific advertisement for each impression by prioritizing those receiving the greatest number of clicks per impression (ie, a presentation-optimizing tool).

Figure 2. Examples from content-specific advertising sets to compare effectiveness between advertising themes (study #2). A) Sleep; B) Alcohol; C) Biosensor & health.



Statistical Analysis

To evaluate the relative success of different advertising content to attract interest, we compared the number of clicks generated per impression [43] among the 3 sets of advertisements tested for 2 days using a chi-square test with post hoc Bonferroni-adjusted pairwise comparisons.

Results

Study 1

Recruitment Results

The 4 Facebook/Instagram advertisements together were displayed 249,940 times (ie, impressions), clicked 4475 times, and yielded 1052 valid surveys in 3 months for a total cost of US $2013.28 (Figure 3). The number of unique users reached was 80,882 out of 220,000 Facebook/Instagram users in the set demographic (18-25 years of age in the New Haven area). One Snapchat advertisement, meanwhile, was displayed 659,366

times, swiped up 9126 times, and yielded 85 valid surveys in 3 months. The total cost was US $1610.70. The number of unique users reached was 140,142 out of 383,000 Snapchat users in the set demographic.

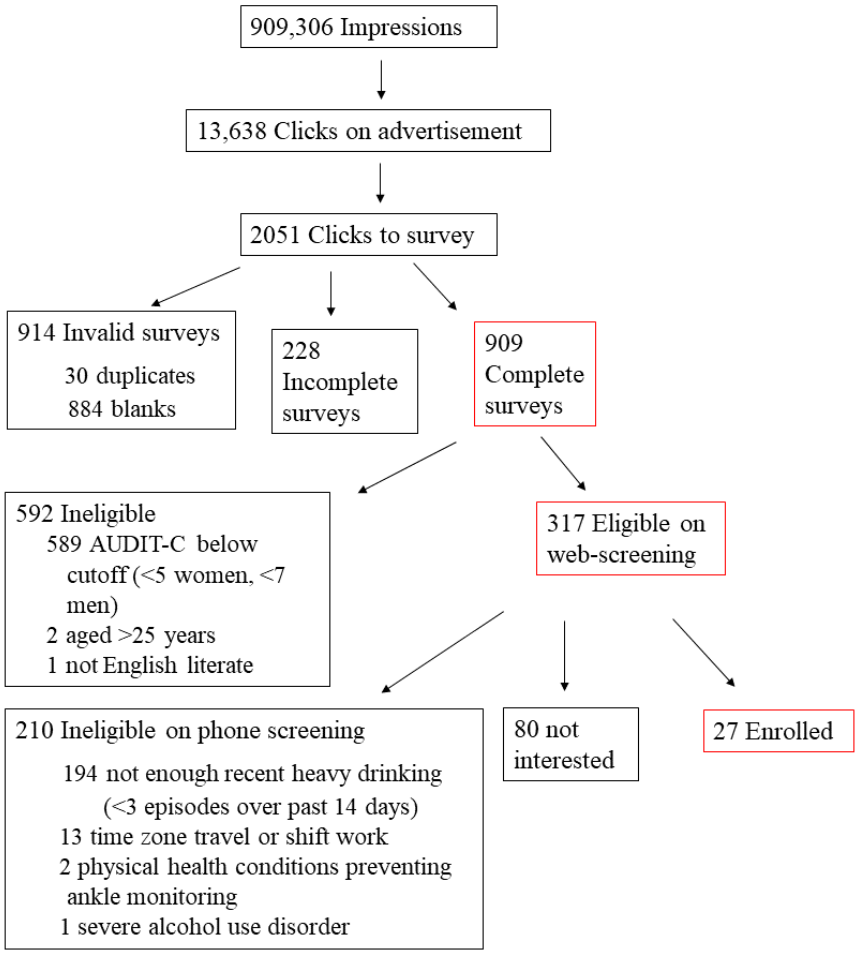
Facebook/Instagram advertisements were viewed mostly on mobile devices (4229/4475, 94.5% vs 215/4475, 4.8% on desktops and 27/4475, 0.6% on tablets), and Snapchat was delivered only on mobile devices. We removed blank (n=884) and duplicate contact information surveys (n=30), leaving 1137 valid entries for analysis. Among these, 228 (20.1%) answered initial questions about their source of referral and reason for interest in the study but did not complete demographic and drinking information. Compared with complete survey responders, these noncompleters expressed less interest (121/228, 53.1% vs 626/909, 68.9%) and concern about their sleep (142/228, 62.3% vs 725/909, 79.8%; *P*<.001) but similar interest (32/228, 14.0% vs 165/909, 18.2%; *P*=.14) and concern about alcohol (14/228, 6.1% vs 49/909, 5.4%; *P*=.66) and the connection between alcohol and sleep (135/228, 59.2% vs 547/909, 60.2%; *P*=.79). The source of study referral did not differ between survey completers and noncompleters (Facebook: 213/228, 93.4% vs 839/909, 92.3%; Snapchat: 15/228, 6.6% vs 70/909, 7.7%; *P*=.57).

Among the completed surveys (n=909), 317 volunteers met the preliminary drinking criteria (ie, AUDIT-C ≥5 women, ≥7 men) and were contacted by phone to verify their recent number of heavy drinking occasions (past 14 days) and final exclusion criteria. Many of these were excluded (n=210) or withdrew interest (n=80) during the screening process. The remaining 27 (ie, 8.5% of the 317 preliminarily eligible web screeners) attended an intake appointment and enrolled in the larger sleep intervention study.

Facebook/Instagram was more expensive per click than Snapchat (US $0.45 vs US $0.18), but less expensive per completed survey (US $2.40 vs US $23.01), initial positive eligibility screen (US $6.35 vs US $50.33), and enrollment (US $95.87 vs US $268.45). Thus, the average cost of enrolling 1 volunteer through the platforms combined was US $134.22. Some of these volunteers spread word-of-mouth referrals about the study, leading to the enrollment of 5 more volunteers (thus achieving the target of 32 volunteers) without further advertising. Counting this indirect return on investment, the average cost of enrolling 1 volunteer was US $113.25. The statistical comparisons reported below were unchanged with the 5 referred volunteers included.

Within the Facebook/Instagram platform, our advertisements were dramatically more successful on Facebook than on Instagram (US $0.43 vs US $0.79 per click) such that the bid-optimizing algorithm targeted 97.4% of the 249,940 impressions to the former.

Figure 3. Flow diagram showing response rates to advertisements and outcomes of screening among responders.



Attraction Mechanisms and Referral Source

Most survey completers reported sleep concerns (725/909, 79.8%), and the majority reported interest in sleep (626/909, 68.9%) and in the connection between sleep and alcohol (547/909, 60.2%; Table 1). However, few reported interest in alcohol (165/909, 18.2%) and still fewer reported concerns about alcohol (49/909, 5.4%) or any other reason for interest in the study (30/909, 3.3%). Sleep concerns were equally prevalent among those who were ineligible, eligible upon web screening but not enrolled, and enrolled. The prevalence of alcohol concerns, however, was greater among those meeting the preliminary drinking criteria (ie, AUDIT-C scores) and still greater among the subset that met full eligibility criteria and enrolled.

An overwhelming majority of survey completers were referred by the Facebook/Instagram platform (839/909, 92.3%) rather than Snapchat (70/909, 7.7%; Table 1). However, this difference was attenuated among volunteers who enrolled (21/27, 78% Facebook/Instagram vs 6/27, 22% Snapchat). Within the Facebook/Instagram platform, 803/839 (95.7%) of referrals came from Facebook and only 4.3% (36/839) came from Instagram, consistent with the greater proportion of impressions targeted to Facebook (see the section Recruitment Results).

Table 1. Referral and attraction mechanisms.

Characteristics	Ineligible upon web screening	Eligible upon web screening but not enrolled	Enrolled	Chi-square *(df)*	*P* value
Participants, n	592	290	27	N/A[a]	N/A
Referral source, n (%)					
Facebook or Instagram	554 (93.6)	264 (91.0)	21 (77.8)[b]	10.0 (2)	.01
Snapchat	38 (6.4)	26 (9.0)	6 (22.2)	N/A	N/A
Reason for interest, n (%)					
Sleep (interest)	438 (74.0)	174 (60.0)[b]	14 (59.1)[b]	21.5 (2)	<.001
Sleep (concern)	466 (78.7)	235 (81.0)	24 (88.9)	2.1 (2)	.35
Alcohol (interest)	73 (12.3)	83 (28.6)[b]	9 (33.3)[b]	39.1 (2)	<.001
Alcohol (concern)	14 (2.4)	29 (10.0)[b]	6 (22.2)[b]	37.7 (2)	<.001
Sleep-alcohol connection (interest)	294 (49.7)	232 (80.0)[b]	21 (77.8)[b]	78.4 (2)	<.001

[a]N/A: not applicable.

[b]Greater than ineligible upon web screening (Bonferonni-adjusted $Q<0.05$).

Demographic Characteristics, Sleep, and Drinking Characteristics

On average, those who completed the web-based survey were 21.1 (SD 2.3) years of age, and 69.4% (631/909) were female (Table 2). A substantial fraction (317/909, 34.9%) met the preliminary drinking criteria (ie, AUDIT-C score). Meeting the preliminary drinking criteria was associated with slightly worse subjective sleep quality. Those who met the preliminary drinking criteria but were later excluded or withdrew interest had lower AUDIT-C scores ($Q<.001$) and approximately 50% lower total drinks per week and frequency of heavy drinking than those who enrolled ($Q=.01$).

Table 2. Demographic, sleep, and drinking characteristics.

Characteristics	Ineligible upon web screening	Eligible upon web screening but not enrolled	Enrolled	Test statistic		*P* value
				F test (*df*)	Chi-square (*df*)	
Number of participants, n	592[a]	290	27	N/A[b]	N/A	N/A
Age (years), mean (SD)	21.1 (2.3)	21.2 (2.1)	20.3 (1.7)	2.352 (908)	N/A	.10
Sex (female), n (%)	392 (66.2)	223 (76.9)[c]	16 (59.3)[d]	N/A	11.8 (2)	.003
Smartphone, n (%)						
iPhone	509 (86.0)	264 (91.0)	24 (88.9)	N/A	4.6 (2)	.10
Android	83 (14.0)	26 (9.0)	3 (11.1)	N/A	N/A	N/A
Sleep quality on 1-5 scale, mean (SD)	2.8 (0.9)	2.7 (0.8)[c]	2.5 (0.9)	5.411 (827)	N/A	.005
AUDIT-C[e] on 0-12 scale, mean (SD)	All below high-risk cutoffs (<5 women and <7 men)	6.6 (1.5)	8.0 (1.6)[d]	19.032 (316)[f]	N/A	<.001
Heavy drinking episodes in the past 14 days, mean (SD)	0.5 (0.8)	2.6 (2.3)[c]	5.1 (1.9)[c,d]	299.243 (908)[f,g]	N/A	<.001
Drinks per week, mean (SD)	3.9 (4.5)	13.7 (8.6)[c]	28.1 (27.4)[c,d]	288.710 (908)[f,g]	N/A	<.001

[a]81 volunteers from this group were excluded from the analysis of sleep quality because they chose not to answer this question. They were nonetheless counted as completed surveys because sleep quality did not affect study eligibility.

[b]N/A: not applicable.

[c]Different than ineligible upon web screening (Bonferroni-adjusted $Q<.05$).

[d]Different than eligible upon web screening but not enrolled (Bonferroni-adjusted $Q<.05$).

[e]AUDIT-C: consumption questions of the Alcohol Use Disorders Identification Test.

[f]Adjusted for sex.

[g]Adjusted for age.

Characteristics by Referral Source

Survey completers referred by Snapchat versus Facebook were younger and more likely to be male (Table 3). They were more likely to report sleep concerns and high-risk drinking. They had a tendency to report lower sleep quality, but it did not reach statistical significance.

Table 3. Participants compared by referral source.

Characteristics	Facebook	Snapchat	Test statistic		*P* value
			Chi-square (*df*)	*F* test (*df*)	
Participants, n	839	70	N/A[a]	N/A	N/A
Reason for interest, n (%)					
Sleep (interest)	582 (69.4)	44 (62.9)	1.3 (1)	N/A	.26
Sleep (concern)	662 (78.9)	63 (90.0)	4.9 (1)	N/A	.03
Alcohol (interest)	154 (18.4)	11 (15.7)	0.3 (1)	N/A	.58
Alcohol (concern)	44 (5.2)	5 (7.1)	0.5 (1)	N/A	.50
Sleep-alcohol connection (interest)	501 (59.7)	46 (65.7)	1.0 (1)	N/A	.33
Age (years), mean (SD)	21.1 (2.2)	19.9 (2.2)	N/A	19.627 (908)	<.001
Sex (female), n (%)	596 (71.0)	35 (50.0)	N/A	13.468 (1)	<.001
Smartphone, n (%)					
iPhone	732 (87.2)	65 (92.9)	1.9 (1)	N/A	.17
Android	107 (12.8)	5 (7.1)	N/A	N/A	N/A
Sleep quality on 1-5 scale, mean (SD)[b]	2.8 (0.9)	2.6 (0.9)	N/A	3.714 (827)	.05
AUDIT-C[c] above high-risk cutoff (≥5 women and ≥7 men)	284 (33.9)	32 (45.7)[b]	4.0 (1)	N/A	.047

[a]N/A: not applicable.

[b]Excludes 9% (76/839) of Facebook volunteers and 7% (5/70) of Snapchat volunteers because they chose not to answer this question.

[c]AUDIT-C: consumption questions of the Alcohol Use Disorders Identification Test.

Characteristics by Sex

Compared with men, women had a less prevalent concern and interest in alcohol, but a greater prevalence of heavy drinking (Table 4). They were also more likely than men to use iPhones versus Android smartphones.

Table 4. Participants compared by sex.

Characteristics	Women	Men	Test statistic		*P* value
			Chi-square (*df*)	*t* test (*df*)	
Number of participants	631	278	N/A[a]	N/A	N/A
Reason for interest, n (%)					
Sleep (interest)	433 (68.6)	193 (69.4)	0.1 (1)	N/A	.81
Sleep (concern)	497 (78.8)	228 (82.0)	1.3 (1)	N/A	.26
Alcohol (interest)	101 (16.0)	64 (23.0)	6.4 (1)	N/A	.01
Alcohol (concern)	26 (4.1)	23 (8.3)	6.5 (1)	N/A	.01
Sleep-alcohol connection (interest)	368 (58.3)	179 (64.4)	3.0 (1)	N/A	.09
Age (years), mean (SD)	21.1 (2.3)	21.0 (2.2)	N/A	1.422 (908)	.23
Smartphone, n (%)					
iPhone	572 (90.6)	225 (80.9)	16.9 (2)	N/A	<.001
Android	59 (9.4)	53 (19.1)	N/A	N/A	N/A
Sleep quality on 1-5 scale, mean (SD)[b]	2.8 (0.9)	2.7 (0.9)	N/A	1.011 (827)	.315
AUDIT-C[c] above high-risk cutoff (≥5 women and ≥7 men)	238 (37.8)	78 (28.2)	7.9 (1)	N/A	.005

[a]N/A: not applicable.

[a]Excludes 9% (56/631) of women and 9% (25/278) of men because they chose not to answer this question.

[b]AUDIT-C: consumption questions of the Alcohol Use Disorders Identification Test.

Study 2

Over the 2 days that we ran content-specific advertising sets with the same platform settings and spending limits, each set received a similar number of impressions, indicating similar efficiency of dissemination (*sleep* 12,523; *alcohol* 11,629; *biosensor and health* 12,825). However, the advertisement set *sleep* generated more clicks per impression (114/12,523, 0.91%) than the other sets (*alcohol* 65/11,629, 0.56%; *biosensor and health* 64/12,825, 0.51%; $Q<.001$).

Discussion

Principal Findings

The results of this study provide overall evidence for the effectiveness of social media advertising for recruiting heavy-drinking young adults to engage in treatment and specifically for content focused on sleep. Poor sleep quality and heavy drinking behaviors both occurred for a substantial fraction of the sample. However, sleep concerns were significantly more common than drinking concerns and appeared to drive engagement in the web screening process. In particular, the sleep advertisement performed better than the drinking advertisement as an advertising hook to generate initial clicks, and sleep, but not drinking concerns was associated with completion of the survey beyond the first two pages. Among the heavy drinkers identified by the web screener, more than one-third passed the further eligibility criteria, and there was strong enrollment uptake among this finally eligible group. Many heavy-drinking young adults report being unconcerned about their drinking [7,8], but our findings indicate that this cohort may nonetheless be concerned about their sleep, which could be a novel on-ramp to engage them in drinking-related treatment. The most common reason for failing to meet the eligibility criteria after the web screening process was acute recent drinking patterns that were too low to meet the study criteria (<3 heavy episodes in the past 14 days) despite chronic drinking patterns that were high risk according to the AUDIT-C scores. The main study required ≥3 heavy drinking episodes in the past 14 days to test the effect of a mobile sleep intervention for the greater at-risk population of young adult heavy drinkers. Using this broader criterion (ie, AUDIT-C scores) would approximately triple the number of eligible volunteers that we could have enrolled.

Comparison With Previous Work

Previous research has demonstrated the cost-effectiveness of Facebook recruitment. A previous systematic review of 27 studies utilizing Facebook recruitment reported that the median cost of enrolling an eligible candidate was US $14.41 [43], which is substantially cheaper than our findings. The possible factors that can elevate cost are low engagement (clicks per impression), conversion (surveys completed per click), eligibility (surveys eligible per surveys completed), and enrollment (volunteers enrolled per eligible volunteers). Our rate of engagement outscored 57% (13/23) of the studies reviewed (ie, our cost per click was US $0.45 vs a median of US $0.51), and our rate of conversion outscored 90% (18/20) of the studies reviewed (ie, our cost per completed survey was US $2.40 vs a median of US $12.00). However, our rate of eligibility was lower than that of every study reviewed (ie, 13% vs a median of 61%). Once eligible candidates were identified, our rate of enrollment was strong for a behavioral intervention clinical trial (25% (27/107) of eligible volunteers enrolled). In summary, the driver of our high cost was the low proportion of survey respondents who met the eligibility criteria. However, as noted above, this would have tripled if we only screened for AUDIT-C scores without requiring a high concentration of heavy drinking in the past 14 days. In that instance, our cost per enrolled volunteer would have been approximately US $32.

Other important factors could have accounted for the cost-effectiveness of the results. For instance, our main study involved participation in an in-person intervention requiring a total of six contacts. In comparison, most previous studies required less volunteer commitment; most involved brief web-based assessments or interventions. These factors could have attenuated the engagement, conversion, and enrollment of our recruitment process and driven up costs.

Costs were higher for Snapchat than for Facebook, as also seen in the only previous study comparing the two platforms for recruitment of youth for research [21]. However, the reasons for this greater cost remain unclear. Snapchat is gaining popularity among young people relative to Facebook [30] and has the potential to be a valuable recruitment tool. Snapchat users were not inferior study candidates than Facebook users. In fact, survey completers referred by Snapchat versus Facebook were significantly more likely to report sleep concerns and high-risk drinking and had a nonsignificant tendency to report lower sleep quality (Table 3). One could speculate that Snapchat's overall greater uses and gratifications around personal disclosure [32-34] facilitated this personal disclosure about health behaviors and a subsequent interest in treating them through our study. Demographic differences also do not explain Snapchat's higher cost: survey completers referred by Snapchat versus Facebook were significantly younger and more likely to be male (Table 3), characteristics that were associated with a tendency for greater enrollment in our sample (Table 2). Differences in geographic radius are also unlikely to explain Snapchat's higher cost: Snapchat advertisements did not need to be extended as far as Facebook advertisements to attain a meaningful number of impressions (10 miles vs 25 miles), indicating that Snapchat users were in closer proximity to New Haven, thus even better positioned to attend the office visits.

Therefore, a more likely explanation for Snapchat's greater cost may be logistical differences between the platforms that we lacked resources to experimentally control given that directly comparing Facebook and Snapchat was not the primary aim of our study. The first such difference was that during analysis 1 (ie, January-April, 2019), Facebook was more restrictive than Snapchat regarding alcohol-related advertising content for users aged <21 years. Our Facebook advertisements during this window placed greater relative emphasis on sleep than alcohol, whereas the Snapchat advertisement had a more balanced emphasis between the two (Figure 1). This difference could have been a confounder, as study 2 later revealed the superiority of sleep content in attracting clicks. The second logistical difference between Facebook and Snapchat was that the latter required a greater number of informational landing pages

separating the advertisement from the survey questions. This requirement, imposed because Snapchat found our privacy policy lacked enough explanation to be directly linked from an advertisement, could have deterred survey completion. On the positive side, it filtered some users not engaged enough to follow through with enrollment, evidenced by the lesser cost difference when tabulating enrolled volunteers as opposed to completed surveys. However, this only partially rebalanced the cost difference between the platforms. Future studies could elicit specialized customer support from Facebook and Snapchat to control these logistical differences, thus making a more valid direct comparison. As for Instagram, it was outperformed by Facebook advertisement placement, but the shared platform bid-optimizing algorithm automatically solved this problem by targeting advertisement impressions away from Instagram to Facebook. Although image-based social media, such as Instagram, previously led to decreased loneliness and increased happiness and satisfaction with life compared with Facebook among young adults [31], we found it was less effective as a recruitment tool.

Limitations

This study had several limitations. First, the in-person nature of the study limited geographic representation. Second, social media advertising cannot ensure representative sampling because click rates are low. Furthermore, we acknowledge that these low click rates could reflect this being a hard-to-reach population that may be better reached by online community-based or respondent-driven sampling [44]. On the other hand, our click rates were above the median of previous studies that included easier-to-reach populations [43], suggesting this was also an appropriate population for which to explore social media advertising. In addition, we successfully oversampled heavy drinkers compared with survey data [45]. Third, alcohol use risk status was limited to self-report. Fourth, we studied predictors of advertisement clicking but not survey completion (eg, survey user experience). Fifth, a disproportionate number of respondents were women (69%). However, the women in our sample had significantly more prevalent heavy drinking and less prevalent concern about drinking than men (Table 4), thus their greater inclusion may be a natural consequence of our objective to target high-risk drinkers who are not concerned about their drinking. Sixth, although we designed advertising content to achieve the aims of both studies, style is a possible confounder. In study 1, although the text mentioned both sleep and alcohol, the visuals were more focused on sleep, which could have biased attention toward survey responses reflecting greater sleep concerns and lesser alcohol concerns. However, before accessing the survey, participants had to view several layers of text (advertising text and study overview text on the following page), which extensively referenced both sleep and alcohol as integral to the inclusion criteria and intervention. Thus, survey respondents had been exposed to information on both sleep and alcohol. We therefore maintain that our analysis demonstrates greater concerns about sleep than alcohol. In study 2, the sets had stylistic differences (visual placement, size of images, and wording) that could have confounded the content differences that we aimed to examine. However, the 27 enrolled participants were queried on poststudy interviews as to what they found appealing about the advertisements, and none mentioned image or stylistic aspects (full content of interviews to be published in a forthcoming manuscript). The final limitation was that we reached more college students than nonstudents.

Conclusions

Despite these limitations, these data demonstrate that social media advertisements targeting young adults with sleep concerns or an interest in sleep reach those who drink heavily and are more effective than advertisements focused on drinking. Many young adults do not seek help for their drinking because of several potential reasons, including low perceived need [46]. Thus, having another on-ramp for alcohol prevention strategies is important. In this case, targeting a coexisting health behavior that young adults may be more open to discuss (ie, sleep) may facilitate better engagement regarding their drinking, a behavior they may be open to discussing. We previously found that heavy-drinking young adults find sleep interventions appealing, are interested in personalized information about sleep and alcohol interactions [15], and that sleep interventions demonstrate promise as a gateway for intervening in alcohol use and engaging heavy-drinking young adults in treatment [16]. Data from this study lend insight into the scalability of this approach by demonstrating that it can be disseminated using social media. We could further infer from this study that a sleep intervention could be disseminated using other web-based venues that are becoming widely popular among young adults (eg, web-based information, mobile apps, and support groups [18]). In addition, among young adults, sleep and alcohol have been found to cluster not only with each other but also with other health behaviors such as smoking and diet. Furthermore, interventions changing one of these behaviors sometimes change others as well [47]. Thus, the proof-of-concept generated by our past work and this study bears potential for extension to these other behaviors. In addition, this line of research warrants further study across other popular social media platforms, such as YouTube and Twitter [48].

Acknowledgments

This study was supported by the National Institute on Alcohol Abuse and Alcoholism (R34AA026021). GA was supported by a fellowship from the Office of Academic Affiliations at the United States Veterans Health Administration. These sponsors had no role in the study design; collection, analysis, or interpretation of the data; writing the manuscript; or the decision to submit the paper for publication. The authors thank Madilyn Augustine, Megan Benlock, Anthony Cifre, and Marisa Rodgers for their essential assistance with data entry.

Authors' Contributions

KD, SO, NR, and LF contributed to the study concept and design. GA, DR, and LF contributed to advertisement and survey design. GA and DR collected the data. GA and MI organized the database. GA, DR, MI, BP, and LF formulated the analytic plan for this substudy. GA conducted the statistical analyses. GA and LF drafted the manuscript, and all authors provided input and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Falk D, Yi H, Hiller-Sturmhöfel S. An epidemiologic analysis of co-occurring alcohol and drug use and disorders: findings from the national epidemiologic survey of alcohol and related conditions (NESARC). Alcohol Res Health 2008;31(2):100-110 [FREE Full text] [Medline: 23584812]
2. Alcohol and Other Drugs. The National Institute on Alcohol Abuse and Alcoholism [NIAAA]. 2008. URL: https://pubs.niaaa.nih.gov/publications/AA76/AA76.htm [accessed 2019-10-15]
3. Hingson RW, Zha W, Weitzman ER. Magnitude of and trends in alcohol-related mortality and morbidity among US college students ages 18-24, 1998-2005. J Stud Alcohol Drugs Suppl 2009 Jul(16):12-20 [FREE Full text] [doi: 10.15288/jsads.2009.s16.12] [Medline: 19538908]
4. Carey KB, Scott-Sheldon LA, Carey MP, DeMartini KS. Individual-level interventions to reduce college student drinking: a meta-analytic review. Addict Behav 2007 Nov;32(11):2469-2494 [FREE Full text] [doi: 10.1016/j.addbeh.2007.05.004] [Medline: 17590277]
5. Carey KB, Scott-Sheldon LA, Elliott JC, Bolles JR, Carey MP. Computer-delivered interventions to reduce college student drinking: a meta-analysis. Addiction 2009 Nov;104(11):1807-1819 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02691.x] [Medline: 19744139]
6. Cronce JM, Larimer ME. Individual-focused approaches to the prevention of college student drinking. Alcohol Res Health 2011;34(2):210-221 [FREE Full text] [Medline: 22330220]
7. Buscemi J, Murphy JG, Martens MP, McDevitt-Murphy ME, Dennhardt AA, Skidmore JR. Help-seeking for alcohol-related problems in college students: correlates and preferred resources. Psychol Addict Behav 2010 Dec;24(4):571-580 [FREE Full text] [doi: 10.1037/a0021122] [Medline: 21198220]
8. Black DR, Coster DC. Interest in a stepped approach model (SAM): identification of recruitment strategies for university alcohol programs. Health Educ Q 1996 Feb;23(1):98-114. [doi: 10.1177/109019819602300107] [Medline: 8822404]
9. DeMartini KS, Fucito LM. Variations in sleep characteristics and sleep-related impairment in at-risk college drinkers: a latent profile analysis. Health Psychol 2014 Oct;33(10):1164-1173 [FREE Full text] [doi: 10.1037/hea0000115] [Medline: 25133844]
10. Singleton RA, Wolfson AR. Alcohol consumption, sleep, and academic performance among college students. J Stud Alcohol Drugs 2009 May;70(3):355-363. [doi: 10.15288/jsad.2009.70.355] [Medline: 19371486]
11. Hasler BP, Kirisci L, Clark DB. Restless sleep and variable sleep timing during late childhood accelerate the onset of alcohol and other drug involvement. J Stud Alcohol Drugs 2016 Jul;77(4):649-655 [FREE Full text] [doi: 10.15288/jsad.2016.77.649] [Medline: 27340970]
12. Wong MM, Robertson GC, Dyson RB. Prospective relationship between poor sleep and substance-related problems in a national sample of adolescents. Alcohol Clin Exp Res 2015 Feb;39(2):355-362 [FREE Full text] [doi: 10.1111/acer.12618] [Medline: 25598438]
13. Breslau N, Roth T, Rosenthal L, Andreski P. Sleep disturbance and psychiatric disorders: a longitudinal epidemiological study of young adults. Biol Psychiatry 1996 Mar 15;39(6):411-418. [doi: 10.1016/0006-3223(95)00188-3] [Medline: 8679786]
14. Miller MB, DiBello AM, Lust SA, Carey MP, Carey KB. Adequate sleep moderates the prospective association between alcohol use and consequences. Addict Behav 2016 Dec;63:23-28 [FREE Full text] [doi: 10.1016/j.addbeh.2016.05.005] [Medline: 27395437]
15. Fucito LM, DeMartini KS, Hanrahan TH, Whittemore R, Yaggi HK, Redeker NS. Perceptions of heavy-drinking college students about a sleep and alcohol health intervention. Behav Sleep Med 2015;13(5):395-411 [FREE Full text] [doi: 10.1080/15402002.2014.919919] [Medline: 24924956]
16. Fucito LM, DeMartini KS, Hanrahan TH, Yaggi HK, Heffern C, Redeker NS. Using sleep interventions to engage and treat heavy-drinking college students: a randomized pilot study. Alcohol Clin Exp Res 2017 Apr;41(4):798-809 [FREE Full text] [doi: 10.1111/acer.13342] [Medline: 28118486]
17. Ambulatory Care Use and Physician Office Visits. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/nchs/fastats/physician-visits.htm [accessed 2019-10-15]

18. Rideout V, Fox S. Digital Health Practices, Social Media Use, and Mental Well-Being Among Teens and Young Adults in the US. Hopelab. 2018. URL: https://hopelab.org/reports/pdf/a-national-survey-by-hopelab-and-well-being-trust-2018.pdf [accessed 2019-10-15]
19. Park BK, Calamaro C. A systematic review of social networking sites: innovative platforms for health research targeting adolescents and young adults. J Nurs Scholarsh 2013 Sep;45(3):256-264. [doi: 10.1111/jnu.12032] [Medline: 23676115]
20. Fazzino TL, Rose GL, Pollack SM, Helzer JE. Recruiting US and Canadian college students via social media for participation in a web-based brief intervention study. J Stud Alcohol Drugs 2015 Jan;76(1):127-132 [FREE Full text] [Medline: 25486401]
21. Ford KL, Albritton T, Dunn TA, Crawford K, Neuwirth J, Bull S. Youth study recruitment using paid advertising on Instagram, Snapchat, and Facebook: cross-sectional survey study. JMIR Public Health Surveill 2019 Oct 9;5(4):e14080 [FREE Full text] [doi: 10.2196/14080] [Medline: 31599739]
22. Guillory J, Wiant KF, Farrelly M, Fiacco L, Alam I, Hoffman L, et al. Recruiting hard-to-reach populations for survey research: using Facebook and Instagram advertisements and in-person intercept in LGBT bars and nightclubs to recruit LGBT young adults. J Med Internet Res 2018 Jun 18;20(6):e197 [FREE Full text] [doi: 10.2196/jmir.9461] [Medline: 29914861]
23. Jones R, Lacroix LJ, Porcher E. Facebook advertising to recruit young, urban women into an HIV prevention clinical trial. AIDS Behav 2017 Nov;21(11):3141-3153 [FREE Full text] [doi: 10.1007/s10461-017-1797-3] [Medline: 28528463]
24. Marshall BD, Green TC, Elston B, Yedinak JL, Hadland SE, Clark MA. The effectiveness of internet- and field-based methods to recruit young adults who use prescription opioids nonmedically. Subst Use Misuse 2018 Aug 24;53(10):1688-1699 [FREE Full text] [doi: 10.1080/10826084.2018.1425725] [Medline: 29364768]
25. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Recruiting young gay and bisexual men for a human papillomavirus vaccination intervention through social media: the effects of advertisement content. JMIR Public Health Surveill 2017 Jun 2;3(2):e33 [FREE Full text] [doi: 10.2196/publichealth.7545] [Medline: 28576758]
26. Ramo DE, Rodriguez TM, Chavez K, Sommer MJ, Prochaska JJ. Facebook recruitment of young adult smokers for a cessation trial: methods, metrics, and lessons learned. Internet Interv 2014 Apr;1(2):58-64 [FREE Full text] [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
27. Teo AR, Liebow SB, Chan B, Dobscha SK, Graham AL. Reaching those at risk for psychiatric disorders and suicidal ideation: Facebook advertisements to recruit military veterans. JMIR Ment Health 2018 Jul 5;5(3):e10078 [FREE Full text] [doi: 10.2196/10078] [Medline: 29980498]
28. Staffileno BA, Zschunke J, Weber M, Gross LE, Fogg L, Tangney CC. The feasibility of using Facebook, craigslist, and other online strategies to recruit young African American women for a web-based healthy lifestyle behavior change intervention. J Cardiovasc Nurs 2017;32(4):365-371. [doi: 10.1097/JCN.0000000000000360] [Medline: 27428356]
29. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. J Med Internet Res 2014 May 1;16(5):e117 [FREE Full text] [doi: 10.2196/jmir.3229] [Medline: 24784982]
30. Perrin A, Anderson M. Share of US Adults Using Social Media, Including Facebook, is Mostly Unchanged Since 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2020-10-15]
31. Pittman M, Reich B. Social media and loneliness: why an Instagram picture may be worth more than a thousand Twitter words. Comput Hum Behav 2016 Sep;62:155-167. [doi: 10.1016/j.chb.2016.03.084]
32. Punyanunt-Carter NM, de la Cruz J, Wrench JS. Investigating the relationships among college students' satisfaction, addiction, needs, communication apprehension, motives, and uses & gratifications with Snapchat. Comput Hum Behav 2017 Oct;75:870-875. [doi: 10.1016/j.chb.2017.06.034]
33. Vaterlaus JM, Barnett K, Roche C, Young JA. 'Snapchat is more personal': an exploratory study on Snapchat behaviors and young adult interpersonal relationships. Comput Hum Behav 2016 Sep;62:594-601. [doi: 10.1016/j.chb.2016.04.029]
34. Choi TR, Sung Y. Instagram versus Snapchat: self-expression and privacy concern on social media. Telemat Inform 2018 Dec;35(8):2289-2298. [doi: 10.1016/j.tele.2018.09.009]
35. Bold KW, Hanrahan TH, O'Malley SS, Fucito LM. Exploring the utility of web-based social media advertising to recruit adult heavy-drinking smokers for treatment. J Med Internet Res 2016 May 18;18(5):e107 [FREE Full text] [doi: 10.2196/jmir.5360] [Medline: 27194456]
36. The Definitive Guide to Creative Optimization. Rocket Fuel. 2015. URL: http://info.rocketfuel.com/rs/rocketfuel/images/RocketFuel_Guide_To_Creative_Optimization.pdf [accessed 2020-04-05]
37. About Ad Auctions. Facebook for Business. 2019. URL: https://www.facebook.com/business/help/430291176997542?id=561906377587030 [accessed 2019-08-10]
38. Single Image or Video Metrics Glossary. Snapchat's Business Center. URL: https://businesshelp.snapchat.com/en-US/a/metrics [accessed 2019-08-10]
39. Call It A Night. Yale School of Medicine. URL: https://medicine.yale.edu/psychiatry/sleepstudy/ [accessed 2019-08-10]
40. Collins RL, Parks GA, Marlatt GA. Social determinants of alcohol consumption: the effects of social interaction and model status on the self-administration of alcohol. J Consult Clin Psychol 1985 Apr;53(2):189-200. [doi: 10.1037//0022-006x.53.2.189] [Medline: 3998247]

41. Demartini KS, Carey KB. Optimizing the use of the AUDIT for alcohol screening in college students. Psychol Assess 2012 Dec;24(4):954-963 [FREE Full text] [doi: 10.1037/a0028519] [Medline: 22612646]
42. Yu L, Buysse DJ, Germain A, Moul DE, Stover A, Dodds NE, et al. Development of short forms from the PROMIS sleep disturbance and sleep-related impairment item banks. Behav Sleep Med 2011 Dec 28;10(1):6-24 [FREE Full text] [doi: 10.1080/15402002.2012.636266] [Medline: 22250775]
43. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
44. Johnston LG, Sabin K, Mai TH, Pham TH. Assessment of respondent driven sampling for recruiting female sex workers in two Vietnamese cities: reaching the unseen sex worker. J Urban Health 2006 Nov;83(6 Suppl):i16-i28 [FREE Full text] [doi: 10.1007/s11524-006-9099-5] [Medline: 17031567]
45. Center for Behavioral Health Statistics and Quality (CBHSQ), Substance Abuse and Mental Health Services Administration (SAMHSA), US Department of Health and Human Services (HHS), RTI International. Results from the 2017 National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration (SAMHSA). 2017. URL: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHDetailedTabs2017/NSDUHDetailedTabs2017.htm#tab2-20B [accessed 2019-10-15]
46. America's Need for and Receipt of Substance Use Treatment in 2015. The Substance Abuse and Mental Health Services Administration (SAMHSA). 2016. URL: https://www.samhsa.gov/data/sites/default/files/report_2716/ShortReport-2716.html [accessed 2019-10-15]
47. Ramo DE, Thrul J, Vogel EA, Delucchi K, Prochaska JJ. Multiple health risk behaviors in young adult smokers: stages of change and stability over time. Ann Behav Med 2020 Jan 24;54(2):75-86 [FREE Full text] [doi: 10.1093/abm/kaz025] [Medline: 31157881]
48. Teens' Social Media Use: How They Connect & What It Means for Health. The US Department of Health and Human Services (HHS). 2016. URL: https://www.hhs.gov/ash/oah/news/e-updates/february-2016-teens-social-media-use/index.html [accessed 2019-10-15]

Abbreviations

AUD: alcohol use disorder
AUDIT-C: consumption questions of the Alcohol Use Disorders Identification Test

Edited by G Eysenbach; submitted 22.12.19; peer-reviewed by Z Ma, M Li, T Muto, V Osadchiy; comments to author 30.03.20; revised version received 12.04.20; accepted 15.06.20; published 11.08.20.

Please cite as:
Ash GI, Robledo DS, Ishii M, Pittman B, DeMartini KS, O'Malley SS, Redeker NS, Fucito LM
Using Web-Based Social Media to Recruit Heavy-Drinking Young Adults for Sleep Intervention: Prospective Observational Study
J Med Internet Res 2020;22(8):e17449
URL: https://www.jmir.org/2020/8/e17449
doi:10.2196/17449
PMID:32780027

Original Paper

Comparing Methods of Recruiting Spanish-Preferring Smokers in the United States: Findings from a Randomized Controlled Trial

Patricia Medina-Ramirez[1], MPH; Patricia Calixte-Civil[1,2], MA; Lauren R Meltzer[1], MPH; Karen O Brandon[1], PhD; Ursula Martinez[1], PhD; Steven K Sutton[2,3], PhD; Cathy D Meade[4,5], PhD, RN, FAAN; Margaret M Byrne[4,6], PhD; Thomas H Brandon[1,2], PhD; Vani N Simmons[1,2], PhD

[1]Tobacco Research & Intervention Program, Department of Health Outcomes & Behavior, Moffitt Cancer Center & Research Institute, Tampa, FL, United States

[2]Department of Psychology, University of South Florida, Tampa, FL, United States

[3]Biostatistics and Bioinformatics Department, Moffitt Cancer Center & Research Institute, Tampa, FL, United States

[4]Department of Health Outcomes & Behavior, Moffitt Cancer Center & Research Institute, Tampa, FL, United States

[5]College of Nursing, University of South Florida, Tampa, FL, United States

[6]Department of Oncologic Sciences, University of South Florida, Tampa, FL, United States

Corresponding Author:
Vani N Simmons, PhD
Tobacco Research & Intervention Program
Department of Health Outcomes & Behavior
Moffitt Cancer Center & Research Institute
4115 E. Fowler Avenue
Tampa, FL, 33617
United States
Phone: 1 813 745 4673 ext 4816
Email: vani.simmons@moffitt.org

Abstract

Background: There is a pressing need to address the unacceptable disparities and underrepresentation of racial and ethnic minority groups, including Hispanics or Latinxs, in smoking cessation trials.

Objective: Given the lack of research on recruitment strategies for this population, this study aims to assess effective recruitment methods based on enrollment and cost.

Methods: Recruitment and enrollment data were collected from a nationwide randomized controlled trial (RCT) of a Spanish-language smoking cessation intervention (N=1417). The effectiveness of each recruitment strategy was evaluated by computing the cost per participant (CPP), which is the ratio of direct cost over the number enrolled. More effective strategies yielded lower CPPs. Demographic and smoking-related characteristics of participants recruited via the two most effective strategies were also compared (n=1307).

Results: Facebook was the most effective method (CPP=US $74.12), followed by TV advertisements (CPP=US $191.31), whereas public bus interior card advertising was the least effective method (CPP=US $642.50). Participants recruited via Facebook had lower average age (*P*=.008) and had spent fewer years in the United States (*P*<.001). Among the participants recruited via Facebook, a greater percentage of individuals had at least a high school education (*P*<.001) and an annual income above US $10,000 (*P*<.001). In addition, a greater percentage of individuals were employed (*P*<.001) and foreign born (*P*=.003). In terms of subethnicity, among the subjects recruited via Facebook, a lower percentage of individuals were of Mexican origin (*P*<.001) and a greater percentage of individuals were of Central American (*P*=.02), South American (*P*=.01), and Cuban (*P*<.001) origin.

Conclusions: Facebook was the most effective method for recruiting Hispanic or Latinx smokers in the United States for this RCT. However, using multiple methods was necessary to recruit a more diverse sample of Spanish-preferring Hispanic or Latinx smokers.

(J Med Internet Res 2020;22(8):e19389) doi:10.2196/19389

KEYWORDS

Hispanic; Latino; smoking cessation intervention; randomized controlled trial; tobacco cigarette; recruitment; social media; Facebook; web banner ad

Introduction

Tobacco smoking is the leading cause of preventable death in the United States [1] and is associated with 4 of the 5 leading causes of death among Hispanics or Latinxs (cancer, heart disease, stroke, and diabetes) [2]. According to the US Census Bureau, Hispanics or Latinxs represent 18% of the US population and are considered the largest and the second fastest growing racial ethnic minority group, making tobacco cessation among this population a public health priority [3,4]. US Hispanic or Latinx adults generally report lower smoking rates (10%) than non-Hispanic Whites (15%) [5]. In addition, Hispanics or Latinxs smoke fewer cigarettes per day (CPD) and are more often nondaily smokers [6]. However, smoking prevalence has been found to vary by ethnic subgroup, with the highest prevalence among those of Puerto Rican (19% men and 16% women) and Mexican (15% men and 7% women) origins and the lowest prevalence among those of Dominican origin (6% total) [7].

Despite the significant burden of smoking-related morbidity and mortality, Hispanics or Latinxs are often underrepresented in clinical trials, including smoking cessation trials [6,8,9]. More than 10 years after the publication by the US Department of Health and Human Services [10] called for the development and evaluation of smoking cessation interventions for racial/ethnic minority groups, only a handful of trials have been conducted with Hispanics or Latinxs [11-16]. There is a scarcity of Spanish-language and evidence-based smoking cessation interventions tailored to this population [17,18]. Yet, 16 million Hispanics or Latinxs report speaking English less than "very well" or "not at all," and another 11 million are bilingual but prefer to use Spanish [19,20].

Recruiting Hispanics or Latinxs for randomized trials can be challenging, especially among those who speak little or no English [21,22]. Additional factors that influence recruitment might also relate to acculturation responses, limited or no previous experience of participating in research, mistrust in study sponsors (eg, government sponsors), language and health literacy, and apprehension about deportation for undocumented immigrants [23]. Although several studies have examined recruitment strategies for enrolling smokers into randomized trials [24], there is a lack of research examining strategies for recruiting underrepresented smokers such as Hispanics or Latinxs [25-27]. Furthermore, few studies have examined the costs of different recruitment strategies [24]. Successfully recruiting Hispanic or Latinx smokers will allow researchers and public health organizations to develop and test more effective interventions and programs for this underserved population. This goal aligns with the national imperatives of inclusivity.

This paper describes the recruitment strategies, including number enrolled, direct cost per strategy, and cost per participant (CPP), for an ongoing randomized controlled trial (RCT) of Hispanic or Latinx smokers who prefer receiving their health information in Spanish [28]. Demographic and smoking-related characteristics of participants recruited via the two most effective strategies were also compared. These findings can be useful to maximize recruitment success for future studies and to inform resource allocation for reaching a large, diverse, and hard-to-reach group of smokers.

Methods

Overview of Randomized Trial

There is a lack of evidence-based, Spanish-language smoking cessation interventions available for Hispanics or Latinxs [17]. To address this gap, we developed a series of culturally relevant self-help booklets and pamphlets for Hispanic or Latinx smokers. This new, targeted intervention was developed by transcreating (translating + adapting) an existing English-language, validated self-help smoking cessation and relapse prevention intervention titled, *Forever Free®: Stop Smoking for Good* (SSFG) [29-31]. Using a multiphase qualitative approach, we successfully transcreated the *SSFG* intervention into a version tailored specifically for individuals who prefer health education materials in Spanish [31]. The Spanish-language version of *SSFG* is titled *Libre del cigarrillo, por mi familia y por mí: Guía para dejar de fumar* (*Free from Cigarettes, for my family and for me: Guide to quit smoking*). Details regarding booklet development and content are presented elsewhere [31].

In the ongoing RCT (trial registration: NCT02945787), the efficacy of the *Libre del Cigarrillo* intervention is compared with usual care, a Spanish-language smoking cessation booklet from the National Cancer Institute [28,32]. The 1417 Hispanic or Latinx smokers enrolled in the study were oversampled from Florida, 555 from Florida and 852 from the rest of the contiguous United States or Puerto Rico. Participants were randomly assigned to receive the *Libre del Cigarrillo* intervention or usual care and were asked to complete follow-up assessments every 6 months for 2 years. Participants received a US $20 gift card for each completed assessment. Details regarding procedures, baseline sample characteristics, and study design for the RCT are presented elsewhere [28].

Procedures

Recruitment for the RCT began in October 2016 and ended in June 2018. Study team members were bilingual, and all recruitment materials (eg, screening forms, advertisements) were in Spanish. Study advertisements included a toll-free number and a link to the study web page for potential participants to enter contact information. This study was reviewed and approved by the Chesapeake Institutional Review Board.

Measures

Eligibility Screening

Eligibility screenings were completed by phone and in Spanish by study staff located in Tampa, Florida. Inclusion criteria were age ≥18 years, smoking ≥5 tobacco cigarettes per week over the last year, not currently enrolled in a face-to-face smoking cessation program, and preference for receiving health education materials in Spanish. Smokers were excluded from participating in the study if they were unable to provide a valid US mailing address or if a member of their household was already enrolled in the study. During the screening, participants were asked how they heard about the study.

Baseline Assessment

The baseline survey was in Spanish and assessed sociodemographic characteristics (eg, race, Hispanic or Latinx subethnicity, education, etc), smoking history, nicotine dependence (Spanish version of the Fagerström Test for Nicotine Dependence) [33], and readiness to quit smoking (Contemplation Ladder) [34]. Other measures assessed the motivation to quit smoking, acculturation, familism, and affect.

Recruitment Strategies and Cost

We used the following recruitment strategies: Facebook, television (TV), website banners, bus signage, flyers, press releases, Craigslist posts, TV and radio interviews, and electronic mailing lists (Table 1). Table 2 presents a description of each recruitment method, the direct cost of implementation, the number of participants enrolled, and the CPP (the direct cost for implementing the strategy divided by the number of enrolled participants). The monetary value of staff time was not included in the cost calculations because the time spent on initiation and implementation was shared equally across the recruitment strategies. Similarly, any cost incurred by the Moffitt Cancer Center (MCC), such as materials (ie, printing supplies) and services (eg, shipping, public relations, and graphic design) were excluded from cost calculations. We only included expenses directly charged to the trial.

Table 1. Recruitment methods.

Methods	Descriptions	Geographic reach
Facebook advertisement	Paid advertisement of free smoking cessation materials containing smoking-related imagery, the study phone number, and the direct link to the web page. Advertisements targeted Hispanic or Latinx adults who showed interest in tobacco-related topics	• Nationwide
TV advertisement	Paid 30-second video advertising of free smoking cessation materials as a part of a research study in Spanish. Advertisements ran on weekdays via Entravision Spanish-language TV stations	• Tampa, Florida • Orlando, Florida • McAllen, Texas • El Paso, Texas • San Diego, California • Albuquerque, New Mexico • Santa Fe, New Mexico • Hartford, Connecticut
Word of mouth	Participants told other smokers in their social network about the study without encouragement or compensation	• N/A[a]
Website banner advertisement	Paid advertisement of free smoking cessation materials. The banners contained pictures of people smoking, the study phone number, and the direct link to the web page. Web-based advertisements were displayed on desktops and mobile devices and targeted Hispanic or Latinx adult smokers on multiple websites	• Miami, Florida • Puerto Rico
Public bus interior cards	Paid advertisement of free smoking cessation materials, the study phone number, and web page address. Advertisements were displayed on the interior of buses	• Orlando, Florida
Flyers	CAB[b] members distributed flyers at community clinics, health fairs, and recruitment events. Flyers publicized free smoking cessation materials and included the study phone number and web page address	• Florida • Texas • New Mexico • Puerto Rico • California • Arizona • Colorado • North Carolina • Ohio • Massachusetts
Radio and TV interviews	A team member was interviewed about the study. The show, called *Tu Salud Informa*, was hosted by a CAB member in Spanish	• Puerto Rico
TV interview (Univision)	A team member was interviewed about the study on the show, *Pregúntale al Médico*	• Tampa Bay area, Florida
Press releases	Stories featuring the RCT[c] were distributed by the cancer center to local newspapers for the Hispanic Heritage Month and Lung Cancer Awareness Month	• Tampa Bay area, Florida
Electronic mailing lists	An electronic newsletter including the study flyer was sent to bilingual community members in the Tampa Bay area and health care providers and community organizations in Puerto Rico	• Tampa Bay area, Florida • Puerto Rico
Craigslist	Free advertisement placed under the *community* section for 5 Florida counties	• Hillsborough • Lee • Polk • Miami-Dade • Orange

[a]N/A: not applicable.

[b]CAB: community advisory board.

[c]RCT: randomized controlled trial.

Table 2. Recruitment cost of strategies directly charged to the project.

Recruitment methods	Direct cost to project, US $	Number of enrolled participants, n	Cost per enrolled participants[a], US $	Duration, months	Number of enrolled participants per month[b], n
Facebook advertisement	76,339	1030	74	16	64
Telivision advertisement	52,994	277	191	6	46
Website banner advertisement	4260	11	387	2	6
Public bus interior cards	1285	2	643	1	2

[a]Cost per enrolled participant=direct cost divided by the enrolled participant.

[b]Enrolled per month=enrolled participant divided by the duration.

TV Advertisement

The first strategy was a TV advertisement campaign. Entravision, an affiliate of the Univision and UniMás TV networks, created two 30-second videos (with a male and female voice-over) advertising no-cost smoking cessation materials, toll-free number, and web page. The first campaign ran from November 2016 to February 2017 in Texas, Florida, California, and New Mexico. The second campaign ran from June 2017 to July 2017 in Texas, Connecticut, and Massachusetts. Markets were selected based on the prevalence of smoking among adult Hispanics or Latinxs in each state. The videos were broadcast 1441 times by at least 13 TV stations. The advertisements initially ran Monday through Friday (weekends were substantially more expensive) but eventually focused on Wednesday through Friday because of greater response volumes. A limitation of TV advertising with Entravision was the company's lack of affiliates in Miami and Puerto Rico, the 2 markets with a substantial Hispanic or Latinx population. The total cost for creating and placing TV advertisements was US $52,994. We screened 525 potential participants, 52.8% (277/525) of whom were enrolled.

Website Banner Advertising

To reach the Miami and Puerto Rico markets, we worked with Entravision's Pulpo digital advertising unit, the leading Hispanic/Latinx advertising platform. A total of 5 versions of desktop and mobile banners were created that included a direct link to the study web page and content similar to the TV advertisements. The campaigns began in November 2016, but they were discontinued after 1 month because of a very low response rate (11 inquiries). A second digital campaign was run in Miami for 1 month (June 2017). Website banner advertisements were charged by impressions or the number of times the advertisements were shown. The costs were US $1500 (216,667 viewer impressions) for the first Miami campaign, US $2000 (300,000 viewer impressions) for the Puerto Rico campaign, and US $760 (110,000 viewer impressions) for the second Miami campaign. The total direct cost for the digital campaign was US $4620 or about US $0.07 per impression. A total of 16 potential participants were screened, and 11 (68%) participants were enrolled.

Facebook

We conducted a nationwide paid Facebook advertisement campaign to reach a broader segment of the target population. An advertising agency managed the campaign. Initially, users who self-identified as primarily Spanish-speaking and living in the United States were targeted. The campaign was also modified to target Hispanic bilingual users to extend the reach. The advertisements targeted users who *liked* or showed interest in tobacco-related topics using keywords (eg, cigarettes, quitting smoking, smoking, and tobacco smoking).

Facebook offered 1 format for the advertisements that included the use of 1 image per advertisement, a limited word count, the study phone number, and a link to the study web page. The research team used a different image, but identical text, for the 2 advertisements. The advertisements were reviewed and approved by Facebook [35]. After 11 months of launch, Facebook paused the campaign because the text did not meet its advertisement policies of attribution (targeting characteristics) [35]. Facebook restarted the campaign after revising and resubmitting text modifications, removing any indication that the advertisement was targeting known smokers.

Concurrent with the nationwide Facebook campaign, which displayed our advertisements to users in any US state and Puerto Rico, we targeted metropolitan areas within the following regions: Southwest (California, Texas, Arizona, New Mexico, Nevada, and Colorado); Northeast (Massachusetts, New York, New Jersey, and Connecticut); Midwest (Illinois); and Southeast United States (District of Columbia, Virginia, Georgia, and Florida). The Facebook advertisements ran from February 2017 to June 2018. The average daily spending was US $35.80, and the average cost per click was US $0.46. The cost of the advertising agency for project management over the course of the campaign was US $6751. Costs for the entire recruitment period totaled US $76,338. Of the 1686 screened individuals, 1030 (61.1%) were enrolled.

Public Transit Advertising

Direct Media USA (now Vector Media) developed and launched a public transit campaign in the Orlando, Florida, metropolitan area, where 31% of the population is Hispanic or Latinx [36]. MCC's graphic design team created an 11"×28" color advertisement that was installed in the interior of 50 buses, reaching more than 100,000 passengers daily. The campaign ran from July to August 2017 and was discontinued because of a low response rate. The total cost was US $1285: a one-time setup fee of US $10.50 per display for 50 displays, a monthly media rate of US $12 per display, plus a shipping and processing charge of US $160. A total of 5 potential participants were screened, and 2 participants were enrolled.

Cultural Advisory Board and Community Partnerships

Recruitment was facilitated by a cultural advisory board (CAB) comprising researchers with experience in recruiting Hispanic or Latinx participants located throughout the Southwest United States and Puerto Rico. Full page color flyers publicizing no-cost smoking cessation materials were distributed at community clinics, health fairs, and recruitment events. The flyers included the study toll-free number and web page address. In addition to the CAB, 2 existing partnerships of academic and community-based organizations facilitated and supported the recruitment process: the Tampa Bay Community Cancer Network (TBCCN) and the Ponce School of Medicine-MCC Partnership in Puerto Rico. In the Tampa Bay area, flyers were distributed via a subset of the community-based organizations that comprise TBCCN. In Puerto Rico, in addition to flyers, information about the study was disseminated via TV (Telemundo/ABC affiliate) and radio (Radio Salud 1520 AM) interviews with a team member. One potential participant was screened from a TV interview. No participants were enrolled using flyers, TV, or radio interviews, and there was no direct cost to the study.

Television Interview

An interview was conducted in Univision Tampa Bay's *Pregúntale al Médico (Ask the Physician)* with a team member about the study. The interview aired on WVEA-TV, an Entravision Communications local station serving the cities of Tampa and Saint Petersburg, FL. There was no direct cost for this service, which resulted in the screening of 2 potential participants and enrollment of 1 participant.

Additional Strategies

The study included an overrepresentation of participants from Florida to facilitate, in part, a biochemical verification of smoking status. Advertisements were placed on Craigslist for a month under the *community* section for 5 Florida counties (Hillsborough, Lee, Miami-Dade, Orange, and Polk). Electronic mailing lists were used to reach bilingual community members from the MCC's catchment area and health care providers in the Tampa Bay area and Puerto Rico, respectively. In addition, press releases were distributed by MCC to newspapers in the Tampa Bay area for Hispanic Heritage Month and Lung Cancer Awareness Month in 2016. These strategies did not result in any enrolled participants, and there was no direct cost to the study.

Word of Mouth

A total of 62 enrolled participants indicated that they learned about the study through other people. Participants were not encouraged to tell other smokers about the study. Word of mouth was not a planned strategy, and there was no direct cost to the study.

Analysis

To further inform recruitment strategies in future smoking cessation trials, analyses aimed to identify differences in demographics and smoking-related variables between participants recruited via Facebook versus TV advertisements, given that these were the recruitment methods that yielded the greatest number of participants. Comparisons were performed using two-tailed *t* tests for continuous variables and chi-square tests for categorical variables, with an alpha of .05. All statistical analyses were conducted using SPSS Statistics Version 25.0 (IBM Corporation).

Results

Participant Sample

For the RCT, 2387 smokers were screened. Of those, 2056 met the inclusion criteria, consented to participate, and were sent a baseline assessment. Of the 1467 who completed the baseline, 1417 were eligible and were randomized and enrolled in the study. Most participants (1307) were recruited via Facebook and TV advertisements. The 1030 participants recruited via Facebook represented 40 states from all US regions: South (n=648), West (n=191), Northeast (n=94), Midwest (n=44), and Puerto Rico (n=53).

Enrollment and CPP

Facebook yielded the greatest number of enrolled participants with 72.69% (1030/1417) of the total sample, followed by TV with 19.54% (277/1417), and word of mouth with 4.38% (64/1417). Website banner advertisements and public bus interior cards yielded less than 1% (12/1417) of the enrolled participants, combined. Facebook had the lowest cost (CPP=US $74.12), followed by TV advertisements (CPP=US $191.31). Website banner advertisements (CPP=US $387.27) and public bus interior cards (CPP=US $642.50) were the most expensive recruitment methods. The computed CPPs for the strategies that were directly charged to the project are displayed in Table 2, along with the average number of enrolled participants per month.

Facebook Versus TV

Demographic Variables

The characteristics of participants by recruitment strategy are presented in Table 3. Participants recruited via TV advertisements had a higher average age (t_{1305}=−2.7; *P*=.008), and a higher percentage of them had less than a high school education (n=1270; χ^2_1=76.3; *P*<.001). In addition, a higher percentage of participants recruited via TV had an annual household income below US $10,000 (n=1227; χ^2_1=12.6; *P*<.001), and a lower percentage of them were employed (n=1274; χ^2_1=35.9; *P*<.001). Participants recruited via TV advertisements had a higher average number of years living in the United States (t_{1207}=−6.6; *P*<.001) and a lower percentage of foreign births (n=1299; χ^2_1=8.8; *P*=.003). In terms of subethnicity, a higher percentage of participants recruited via TV advertisements were of Mexican/Mexican American origin (n=1300; χ^2_1=54.4; *P*<.001) and a lower percentage of them were of Central American origin (n=1300; χ^2_1=5.6; *P*=.02), South American origin (n=1300; χ^2_1=6.6; *P*=.01), and Cuban origin (n=1300; χ^2_1=41.4; *P*<.001). Finally, participants

recruited via TV advertisements had a higher percentage of those not reporting race (n=1307; χ^2_1=16.7; *P*<.001).

Other Variables

There were no statistically significant differences in the recruitment method for acculturation or any of the smoking-related variables.

Table 3. Participant characteristics by the recruitment method (n=1307).

Characteristics	Facebook (n=1030)	Television (n=277)	*P* value
Age (years), mean (SD)	49.4 (11.6)	51.5 (12.0)	.008
Age group (years), n (%)			.12
18-29	51 (5.0)	12 (4.3)	
30-39	171 (16.6)	34 (12.3)	
40-49	249 (24.2)	67 (24.2)	
50-59	361 (35.0)	93 (33.6)	
≥60	198 (19.2)	71 (25.6)	
Sex (women), n (%)	490 (47.6)	143 (51.6)	.23
Education (<high school diploma), n (%)	251 (25.0)	141 (52.8)	<.001
Employed, n (%)	631 (62.7)	113 (42.3)	<.001
Annual income per household <US $10,000, n (%)	371 (38.1)	128 (50.4)	<.001
Marital status (married or cohabiting), n (%)	479 (46.9)	139 (50.2)	.33
Race, n (%)			.001
White	486 (47.2)	109 (39.4)	.10
Black or African American	34 (3.3)	2 (0.7)	.06
American Indian or Alaska Native	32 (3.1)	6 (2.2)	.82
Asian	1 (0.1)	1 (0.4)	.31
Native Hawaiian or Other Pacific Islander	24 (2.3)	4 (1.4)	.80
Multiple races	40 (3.9)	6 (2.2)	.44
Not reported	413 (40.1)	149 (53.8)	<.001
Subethnicity, n (%)			<.001
Puerto Rican	163 (15.9)	53 (19.3)	.17
Central American	69 (6.7)	8 (2.9)	.02
Mexican or Mexican American	315 (30.7)	150 (54.2)	<.001
South American	99 (9.6)	13 (4.7)	.01
Cuban	254 (24.8)	19 (6.9)	<.001
Dominican	28 (2.7)	6 (2.2)	.62
Other	14 (1.4)	3 (1.1)	.73
More than one subethnicity	84 (8.2)	22 (8.0)	.93
Born outside the United States, n (%)	812 (78.8)	195 (70.4)	.003
Years in the United States (for foreign born)[a], mean (SD)	14.1 (11.7)	20.4 (13.0)	<.001
Acculturation (SASH)[b], mean (SD)	19.7 (6.3)	19.0 (6.5)	.08
Smoking-related variables			
Smoke daily, n (%)	970 (94.2)	253 (91.3)	.12
E-cigarettes[c], n (%)	45 (4.4)	11 (4.0)	.80
Years of smoking, mean (SD)	27.8 (12.6)	29.4 (13.8)	.09
CPD[d], mean (SD)	14.9 (8.5)	14.0 (7.7)	.09
FTND[e], mean (SD)	4.9 (2.4)	4.9 (2.3)	.90
Contemplation ladder, mean (SD)	6.9 (2.8)	7.0 (2.9)	.49

[a]Only includes those who were not born in the United States or Puerto Rico.

[b]SASH: Short Acculturation Scale for Hispanics.

[c]e-cigarettes: electronic cigarettes.

[d]CPD: cigarettes per day.

[e]FTND: Fagerström Test for Nicotine Dependence.

Discussion

Principal Findings

A greater representation of Hispanic or Latinx individuals is needed in smoking cessation trials. Moreover, given the heterogeneity of the Hispanic or Latinx population, it may be necessary to analyze which recruitment strategies successfully contributed to the accrual of a diverse sample. This study aimed to evaluate the methods used to recruit Spanish-speaking Hispanic or Latinx smokers in terms of CPP and enrollment success. The results will inform the development and implementation of effective recruitment strategies for future studies.

The ongoing RCT used a combination of mass media (ie, TV advertisements, press releases, and TV and radio interviews); social media (ie, Facebook); and internet-based (ie, banner advertisements) and community outreach strategies. A total of 3 recruitment methods accounted for the majority of enrolled participants: Facebook, TV, and word of mouth. Facebook was ultimately the most effective recruitment method based on CPP (US $74 per participant), and it also yielded the highest average enrollees per month (n=64). TV advertisements, the original recruitment method, yielded the second highest number of participants at about 46 enrollees per month and were the second most effective strategy at US $190 CPP. There are several reasons we believe Facebook was most effective in reaching our intended population. Unlike TV advertisements, which can be very expensive in markets with a high density of Hispanics or Latinxs (eg, Miami, New York City, and Puerto Rico), Facebook advertising campaigns can have a national reach with an affordable cost regardless of the city, state, or Hispanic or Latinx population share. We were also able to further target our advertising using interest-based keywords (eg, cigarettes, quitting smoking, smoking, and tobacco smoking), which is a unique benefit of using this platform as compared with TV. Another reason Facebook was the most effective method for the current RCT is that this platform is a trusted media source among Hispanics or Latinxs. Evidence suggests that Hispanics or Latinxs are receptive to advertising via Facebook, and compared with the other cultural groups, Hispanics or Latinxs have been found to be more inclined to share content with their extended networks [37].

In comparison with mass media, social media/web-based advertisements can be highly targeted, thus increasing the chances of reaching the intended population, which should translate to more eligible participants. However, Facebook policies targeting personal attributes changed midstudy. The original advertisement stated, “Are *you* interested in quitting smoking?” with the assumption that the social media user was a smoker. Following the policy change, less emphasis was placed on the user’s personal attributes and more emphasis on the availability of materials to quit smoking at no cost. An advertisement introduced later stated, “Quit smoking help available at no cost!” Furthermore, the process of revision and approval for Facebook advertisements was lengthy and iterative, which contributed to unanticipated delays in recruitment. Future researchers may consider consulting with Facebook earlier in advertisement development to reduce potential delays.

Offline campaigns, such as TV advertisements, usually resulted in call volumes that peaked immediately following the release of the advertisement. Facebook advertisements typically resulted in a steady response rate. Another advantage of Facebook advertisements was that they linked interested individuals directly to the study web page.

It is important to consider that Hispanics or Latinxs are among the leading social media users and the fastest growing ethnic group on Facebook in the United States. In 2015, 75% of Hispanic or Latinx adult internet users were on Facebook (compared with 45% in 2010) [38,39]. Furthermore, a recent report on TV viewership found that engagement with Spanish-language TV and the household reach of Spanish networks in the United States have decreased in recent years [40]. However, mass media (ie, TV and radio) have shown success in reaching the Hispanic or Latinx audience [39,41]. Accordingly, TV advertisements still played an essential role in ensuring that we recruited a diverse sample. Participants recruited via TV advertisements were 2 years older on average and had lower socioeconomic attainment (ie, education, income, and employment) than those recruited via Facebook. As such, utilizing methods besides Facebook is an essential strategy to ensure that a wider diversity of socioeconomic status is included. Ensuring that this subgroup is captured in smoking cessation trials is of particular importance because research demonstrates a higher prevalence of smoking among lower socioeconomic status groups, and they greatly bear the burden of tobacco-related health disparities [42].

Furthermore, using Facebook in addition to TV was essential in enrolling a diverse sample of Hispanic or Latinx subgroups. A higher percentage of Mexican/Mexican American participants were recruited via TV advertisements, whereas a higher percentage of Central American, South American, and Cuban participants were recruited from Facebook. The greater representation of Mexican Americans is possibly a reflection of heavier TV advertising in the Southwest. Given the variations within the Hispanic or Latinx population in terms of lifestyle, socioeconomic characteristics, access to health care, and smoking patterns, it is important to develop recruitment strategies to ensure representation from all countries of origin [43-47]. A more representative sample allows for comparisons across groups and increases the potential generalizability of these findings. It is evident that Facebook and TV advertisements attracted participants with different countries of origin into the RCT.

Unexpectedly, word of mouth was the method that yielded the third highest number of participants, even though we did not encourage or compensate for referrals. Brodar et al [26] also found that word of mouth was a free method that was highly

effective in yielding enrolled smokers. Receiving free smoking cessation materials and compensation for completing assessments may have motivated participants to tell their family members and friends about the study. However, researchers should be cautious about participants within the same social network being randomized to different intervention arms because of the potential for cross-contamination. Another unexpected finding was the poor performance of the flyers. However, this is consistent with past research showing that Hispanic or Latinx smokers were less likely to be recruited by flyers [26,48].

Among paid recruitment methods, website banners and bus signs yielded very few participants at a very high cost, consistent with previous reports [26]. Website banner advertisements may be more effective than physical advertisements because they provide potential participants with easier and more immediate access to the study web page.

Limitations

This study has some limitations that should be noted. First, there is a possibility of misclassification of recruitment methods by participants. However, only a negligible number of screened individuals (n=48) and enrolled participants (n=34) endorsed methods that were not used in their geographic region. Second, given the broad reach of the mass media campaign, and the fact that we used a variety of methods, it is possible that participants were exposed to more than one advertising modality. Third, only a single social media platform (ie, Facebook) was used. Twitter and Instagram, with about 10 and 13 million Hispanic or Latinx adult users, respectively, are other social media sites that hold great potential for recruiting participants from this population [38]. However, as stated previously, Facebook is the most popular social media platform among Hispanic or Latinx adults with 28 million users, more than Instagram and Twitter combined [38]. Finally, the representativeness of this sample to the population of US Spanish-language preferred, treatment-seeking Hispanic or Latinx smokers is difficult to know. On average, this sample comprised older, heavier smoking, and more daily smokers than national samples of Hispanic or Latinx smokers [43]. However, these demographics are not unusual for a treatment-seeking subpopulation of smokers.

Conclusions

Overall, successfully recruiting a socioeconomically and ethnically diverse sample of Hispanic or Latinx smokers is necessary to increase the access to cessation trials and the generalizability of study findings, in response to national calls for more inclusivity in clinical research [49]. Using multiple strategies, we recruited a large national sample of Hispanic or Latinx smokers representing diverse ethnic subgroups from this population. The utilization of social media was an effective method to recruit participants into an RCT and improve the inclusion of a historically underrepresented population. To the best of our knowledge, no other smoking cessation trial with Hispanic or Latinx smokers has reported on using Facebook as a recruitment strategy [50]. Social media use is rapidly growing among Hispanic or Latinx adults and holds great potential for the recruitment of study participants. Facebook is a low-cost, effective method to recruit Hispanic or Latinx smokers. However, mass media campaigns, specifically via TV, are still needed to reach a segment of the Hispanic or Latinx population that is at a greater socioeconomic disadvantage. Researchers aiming to enroll a representative sample of Hispanic or Latinx smokers should consider both recruitment methods.

Acknowledgments

This work was supported by the National Cancer Institute of the National Institutes of Health (grant number R01CA199143), the James and Esther King Biomedical Research Program of the Florida Department of Health (grant number 5JK03), and in part by the Biostatistics and Bioinformatics Shared Resource and the Participant Research, Interventions, and Measure Shared Resource at MCC, a National Cancer Institute-designated Comprehensive Cancer Center (grant number P30CA76292). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors appreciate the efforts of the Public Relations and Marketing Department of the H. Lee MCC and Research Institute. The authors also thank the many undergraduate research assistants for their contribution to this project.

Authors' Contributions

PM, VS, and TB conceived the presented idea. PM, PC, LM, VS, and TB drafted the manuscript with support from KB. PM, PC, and KB conducted data cleaning, and PM conducted the statistical analyses. TB and VS acquired funding for the research project. CM, MB, SS and UM provided critical feedback and helped shape the research, analysis, and manuscript. VS supervised the project.

Conflicts of Interest

TB has received research support from Pfizer Inc and is on the Advisory Board of Hava Health, Inc. All other coauthors report no conflicts of interest.

References

1. 2014 Surgeon General's Report: The Health Consequences of Smoking—50 Years of Progress. Centers for Disease Control and Prevention. 2014. URL: https://www.cdc.gov/tobacco/data_statistics/sgr/50th-anniversary/index.htm [accessed 2020-07-24]

2. Leading Causes of Death and Numbers of Deaths, by Sex, Race, and Hispanic Origin: United States, 1980 and 2016. Centers for Disease Control and Prevention. 2017. URL: https://www.cdc.gov/nchs/data/hus/2016/019.pdf [accessed 2019-11-04]
3. Facts for Features: Hispanic Heritage Month 2017. United States Census. 2017. URL: https://www.census.gov/newsroom/facts-for-features/2017/hispanic-heritage.html [accessed 2019-11-04]
4. Flores A. How the US Hispanic Population is Changing. Pew Research Center. 2017. URL: https://www.pewresearch.org/fact-tank/2017/09/18/how-the-u-s-hispanic-population-is-changing/ [accessed 2019-11-21]
5. Wang TW, Asman K, Gentzke AS, Cullen KA, Holder-Hayes E, Reyes-Guzman C, et al. Tobacco product use among adults - United States, 2017. MMWR Morb Mortal Wkly Rep 2018 Nov 9;67(44):1225-1232 [FREE Full text] [doi: 10.15585/mmwr.mm6744a2] [Medline: 30408019]
6. Trinidad D, Pérez-Stable EJ, Emery S, White M, Grana R, Messer K. Intermittent and light daily smoking across racial/ethnic groups in the United States. Nicotine Tob Res 2009 Feb;11(2):203-210 [FREE Full text] [doi: 10.1093/ntr/ntn018] [Medline: 19246433]
7. Miller KD, Goding Sauer A, Ortiz AP, Fedewa SA, Pinheiro PS, Tortolero-Luna G, et al. Cancer statistics for Hispanics/Latinos, 2018. CA Cancer J Clin 2018 Nov;68(6):425-445 [FREE Full text] [doi: 10.3322/caac.21494] [Medline: 30285281]
8. Ford ME, Siminoff LA, Pickelsimer E, Mainous AG, Smith DW, Diaz VA, et al. Unequal burden of disease, unequal participation in clinical trials: solutions from African American and Latino community members. Health Soc Work 2013 Feb;38(1):29-38 [FREE Full text] [doi: 10.1093/hsw/hlt001] [Medline: 23539894]
9. George S, Duran N, Norris K. A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health 2014 Feb;104(2):e16-e31. [doi: 10.2105/AJPH.2013.301706] [Medline: 24328648]
10. Clinical Practice Guideline Treating Tobacco Use and Dependence 2008 Update Panel, Liaisons, and Staff. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A US public health service report. Am J Prev Med 2008 Aug;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]
11. Stanton CA, Papandonatos GD, Shuter J, Bicki A, Lloyd-Richardson EE, de Dios MA, et al. Outcomes of a tailored intervention for cigarette smoking cessation among Latinos living with HIV/AIDS. Nicotine Tob Res 2015 Aug;17(8):975-982 [FREE Full text] [doi: 10.1093/ntr/ntv014] [Medline: 26180222]
12. de Dios MA, Cano M, Vaughan E, Childress SD, McNeel MM, Harvey LM, et al. A pilot randomized trial examining the feasibility and acceptability of a culturally tailored and adherence-enhancing intervention for Latino smokers in the US. PLoS One 2019;14(1):e0210323 [FREE Full text] [doi: 10.1371/journal.pone.0210323] [Medline: 30633744]
13. Asfar T, Caban-Martinez AJ, McClure LA, Ruano-Herreria EC, Sierra D, Gilford Clark G, et al. A cluster randomized pilot trial of a tailored worksite smoking cessation intervention targeting Hispanic/Latino construction workers: intervention development and research design. Contemp Clin Trials 2018 Apr;67:47-55 [FREE Full text] [doi: 10.1016/j.cct.2018.02.007] [Medline: 29454141]
14. Borrelli B, McQuaid EL, Novak SP, Hammond SK, Becker B. Motivating Latino caregivers of children with asthma to quit smoking: a randomized trial. J Consult Clin Psychol 2010 Feb;78(1):34-43. [doi: 10.1037/a0016932] [Medline: 20099948]
15. Cabriales JA, Cooper TV, Salgado-Garcia F, Naylor N, Gonzalez E. A randomized trial of a brief smoking cessation intervention in a light and intermittent Hispanic sample. Exp Clin Psychopharmacol 2012 Oct;20(5):410-419. [doi: 10.1037/a0029039] [Medline: 22731733]
16. Pollak KI, Lyna P, Bilheimer AK, Gordon KC, Peterson BL, Gao X, et al. Efficacy of a couple-based randomized controlled trial to help Latino fathers quit smoking during pregnancy and postpartum: the Parejas trial. Cancer Epidemiol Biomarkers Prev 2015 Feb;24(2):379-385 [FREE Full text] [doi: 10.1158/1055-9965.EPI-14-0841] [Medline: 25406226]
17. Webb MS, Rodríguez-Esquivel D, Baker EA. Smoking cessation interventions among Hispanics in the United States: a systematic review and mini meta-analysis. Am J Health Promot 2010;25(2):109-118. [doi: 10.4278/ajhp.090123-LIT-25] [Medline: 21039291]
18. Báezconde-Garbanati L, Beebe LA, Pérez-Stable EJ. Building capacity to address tobacco-related disparities among American Indian and Hispanic/Latino communities: conceptual and systemic considerations. Addiction 2007 Oct;102(Suppl 2):112-122. [doi: 10.1111/j.1360-0443.2007.01962.x] [Medline: 17850621]
19. Krogstan J, Stepler R, Lopez M. English Proficiency on the Rise Among Latinos. Pew Research Center. 2015. URL: https://www.pewresearch.org/hispanic/2015/05/12/english-proficiency-on-the-rise-among-latinos/ [accessed 2020-01-13]
20. Krosgstad J, Gonzalez-Barrera A. A Majority of English-speaking Hispanics in the US Are Bilingual. Pew Research Center. 2015. URL: https://www.pewresearch.org/fact-tank/2015/03/24/a-majority-of-english-speaking-hispanics-in-the-u-s-are-bilingual/ [accessed 2020-01-13]
21. Aguirre TM, Koehler AE, Joshi A, Wilhelm SL. Recruitment and retention challenges and successes. Ethn Health 2018 Jan;23(1):111-119. [doi: 10.1080/13557858.2016.1246427] [Medline: 27764955]
22. Felsen CB, Shaw EK, Ferrante JM, Lacroix LJ, Crabtree BF. Strategies for in-person recruitment: lessons learned from a New Jersey primary care research network (NJPCRN) study. J Am Board Fam Med 2010;23(4):523-533 [FREE Full text] [doi: 10.3122/jabfm.2010.04.090096] [Medline: 20616295]

23. Sha M, McAvinchey G, Quiroz R, Moncada J. Successful techniques to recruit Hispanic and Latino research participants. Surv Pract 2017 Jun 1;10(3):1-9 [FREE Full text] [doi: 10.29115/sp-2017-0014] [Medline: 27114390]
24. Belisario JS, Bruggeling MN, Gunn LH, Brusamento S, Car J. Interventions for recruiting smokers into cessation programmes. Cochrane Database Syst Rev 2012 Dec 12;12:CD009187 [FREE Full text] [doi: 10.1002/14651858.CD009187.pub2] [Medline: 23235672]
25. Thompson TP, Greaves CJ, Ayres R, Aveyard P, Warren FC, Byng R, et al. Lessons learned from recruiting socioeconomically disadvantaged smokers into a pilot randomized controlled trial to explore the role of exercise assisted reduction then stop (EARS) smoking. Trials 2015 Feb 12;16:1 [FREE Full text] [doi: 10.1186/1745-6215-16-1] [Medline: 25971836]
26. Brodar KE, Hall MG, Butler EN, Parada H, Stein-Seroussi A, Hanley S, et al. Recruiting diverse smokers: enrollment yields and cost. Int J Environ Res Public Health 2016 Dec 16;13(12):- [FREE Full text] [doi: 10.3390/ijerph13121251] [Medline: 27999280]
27. Webb MS, Seigers D, Wood EA. Recruiting African American smokers into intervention research: relationships between recruitment strategies and participant characteristics. Res Nurs Health 2009 Feb;32(1):86-95. [doi: 10.1002/nur.20299] [Medline: 18767129]
28. Medina-Ramírez P, Sutton SK, Martínez U, Meade CD, Byrne MM, Brandon KO, et al. A randomized controlled trial of a smoking cessation self-help intervention for Spanish-speaking Hispanic/Latinx smokers: study design and baseline characteristics. Contemp Clin Trials 2019 Oct;85:105836. [doi: 10.1016/j.cct.2019.105836] [Medline: 31473331]
29. Brandon TH, Simmons VN, Sutton SK, Unrod M, Harrell PT, Meade CD, et al. Extended self-help for smoking cessation: a randomized controlled trial. Am J Prev Med 2016 Jul;51(1):54-62 [FREE Full text] [doi: 10.1016/j.amepre.2015.12.016] [Medline: 26868284]
30. Simmons VN, Sutton SK, Meltzer LR, Unrod M, Meade CD, Brandon TH. Long-term outcomes from a self-help smoking cessation randomized controlled trial. Psychol Addict Behav 2018 Nov;32(7):710-714 [FREE Full text] [doi: 10.1037/adb0000401] [Medline: 30284878]
31. Piñeiro B, Díaz DR, Monsalve LM, Martínez U, Meade CD, Meltzer LR, et al. Systematic transcreation of self-help smoking cessation materials for Hispanic/Latino smokers: improving cultural relevance and acceptability. J Health Commun 2018;23(4):350-359 [FREE Full text] [doi: 10.1080/10810730.2018.1448487] [Medline: 29533167]
32. Guía: Viva De Forma Más Saludable Para Usted Y Su Familia. SmokeFree. 2011. URL: http://smokefree.gov/sites/default/files/pdf/guia-para-dejar-de-fuma-2011.pdf [accessed 2019-09-12]
33. Becoña E, Vázquez FL. The Fagerström test for nicotine dependence in a Spanish sample. Psychol Rep 1998 Dec;83(3 Pt 2):1455-1458. [doi: 10.2466/pr0.1998.83.3f.1455] [Medline: 10079737]
34. Biener L, Abrams DB. The contemplation ladder: validation of a measure of readiness to consider smoking cessation. Health Psychol 1991;10(5):360-365. [doi: 10.1037//0278-6133.10.5.360] [Medline: 1935872]
35. Advertising Policies. Facebook. URL: https://www.facebook.com/policies/ads/# [accessed 2019-10-03]
36. Orlando-Kissimmee-Sanford, FL Metro Area. Census Reporter. URL: https://censusreporter.org/profiles/31000US36740-orlando-kissimmee-sanford-fl-metro-area/ [accessed 2018-01-01]
37. Digital Diversity: A Closer Look at US Hispanics. Facebook for Business. 2014. URL: https://www.facebook.com/business/news/insights/digital-diversity-a-closer-look-at-us-hispanics [accessed 2019-12-09]
38. Duggan M, Ellison N, Lampe C, Lenhart A, Madden M. Social Media Update 2014. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/01/09/social-media-update-2014/ [accessed 2019-12-09]
39. Cultural Insights; Communicating With Hispanics/latinos. CDC Stacks. 2012. URL: https://stacks.cdc.gov/view/cdc/13183 [accessed 2019-12-09]
40. Rice C. Are Hispanic Viewers Watching Less Spanish Language TV? Samba TV. 2018. URL: https://platform.samba.tv/spanish-language-viewers/ [accessed 2019-12-09]
41. Sonderup L. Hispanic Marketing: A Critical Market Segment. Advertising & Marketing Review. 2010. URL: http://www.admarketreview.com/public_html/docs/fs075.html [accessed 2019-12-09]
42. The Epidemiology of Tobacco-Related Health Disparities. Division of Cancer Control and Population Sciences: DCCPS. 2017. URL: https://cancercontrol.cancer.gov/brp/tcrb/monographs/22/docs/m22_2.pdf [accessed 2019-11-26]
43. Kaplan RC, Bangdiwala SI, Barnhart JM, Castañeda SF, Gellman MD, Lee DJ, et al. Smoking among US Hispanic/Latino adults: the Hispanic community health study/study of Latinos. Am J Prev Med 2014 May;46(5):496-506 [FREE Full text] [doi: 10.1016/j.amepre.2014.01.014] [Medline: 24745640]
44. Blanco L, Garcia R, Pérez-Stable EJ, White MM, Messer K, Pierce JP, et al. National trends in smoking behaviors among Mexican, Puerto Rican, and Cuban men and women in the United States. Am J Public Health 2014 May;104(5):896-903. [doi: 10.2105/AJPH.2013.301844] [Medline: 24625159]
45. Morales LS, Lara M, Kington RS, Valdez RO, Escarce JJ. Socioeconomic, cultural, and behavioral factors affecting Hispanic health outcomes. J Health Care Poor Underserved 2002 Nov;13(4):477-503 [FREE Full text] [doi: 10.1177/104920802237532] [Medline: 12407964]

46. Mothel S, Patten E. The 10 Largest Hispanic Origin Groups: Characteristics, Rankings, Top Counties. Pew Research Center. 2012. URL: https://www.pewresearch.org/hispanic/2012/06/27/the-10-largest-hispanic-origin-groups-characteristics-rankings-top-counties/ [accessed 2019-12-02]
47. Profile: Hispanic/Latino Americans. The Office of Minority Health (OMH). 2019 Aug 22. URL: https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=3&lvlid=64 [accessed 2019-12-02]
48. García AA, Zuñiga JA, Lagon C. A personal touch: the most important strategy for recruiting Latino research participants. J Transcult Nurs 2017 Jul;28(4):342-347 [FREE Full text] [doi: 10.1177/1043659616644958] [Medline: 27114390]
49. NIH Policy and Guidelines on The Inclusion of Women and Minorities as Subjects in Clinical Research. NIH Central Resource for Grants and Funding Information. 2017 Dec 6. URL: https://grants.nih.gov/policy/inclusion/women-and-minorities/guidelines.htm [accessed 2020-02-12]
50. Hudnut-Beumler J, Po'e E, Barkin S. The use of social media for health promotion in Hispanic populations: a scoping systematic review. JMIR Public Health Surveill 2016 Jul 11;2(2):e32 [FREE Full text] [doi: 10.2196/publichealth.5579] [Medline: 27400979]

Abbreviations

CAB: cultural advisory board
CPD: cigarettes per day
CPP: cost per participant
MCC: Moffitt Cancer Center
RCT: randomized controlled trial
SSFG: Stop Smoking for Good
TBCCN: Tampa Bay Community Cancer Network

Edited by G Eysenbach; submitted 17.04.20; peer-reviewed by P Sukumaran, K Aguirre, Y Rivera; comments to author 12.06.20; revised version received 23.06.20; accepted 25.06.20; published 14.08.20.

Please cite as:
Medina-Ramirez P, Calixte-Civil P, Meltzer LR, Brandon KO, Martinez U, Sutton SK, Meade CD, Byrne MM, Brandon TH, Simmons VN
Comparing Methods of Recruiting Spanish-Preferring Smokers in the United States: Findings from a Randomized Controlled Trial
J Med Internet Res 2020;22(8):e19389
URL: https://www.jmir.org/2020/8/e19389
doi:10.2196/19389
PMID:32795986

Original Paper

Wildfire-Like Effect of a WhatsApp Campaign to Mobilize a Group of Predominantly Health Professionals With a University Degree on a Health Issue: Infodemiology Study

Vanja Kopilaš[1], MA; Srećko Gajović[2], MD, PhD

[1]Faculty of Croatian Studies, University of Zagreb, Zagreb, Croatia

[2]Croatian Institute for Brain Research, University of Zagreb School of Medicine, Zagreb, Croatia

Corresponding Author:
Srećko Gajović, MD, PhD
Croatian Institute for Brain Research
University of Zagreb School of Medicine
Šalata 12
Zagreb, HR-10000
Croatia
Phone: 385 14566948
Email: srecko.gajovic@hiim.hr

Abstract

Background: Online interactions within a closed WhatsApp group can influence the attitudes and behaviors of the users in relation to health issues.

Objective: This study aimed to analyze the activity of the members of a WhatsApp group initiated to raise awareness of the possible health effects of 5G mobile networks and mobilize members to sign the related petition.

Methods: We retrospectively analyzed data from the WhatsApp group of 205 members that was active during 4 consecutive days in August 2019. The messages exchanged were collected, anonymized, and analyzed according to their timing and content.

Results: The WhatsApp group members were invited to the group from the administrator's contacts; 91% (187/205) had a university degree, 68% (140/205) were medical professionals, and 24% (50/205) held academic positions. Approximately a quarter of the members (47/205, 23%) declared in their messages they signed the corresponding petition. The intense message exchange had wildfire-like features, and the majority of messages (126/133, 95%) were exchanged during the first 26 hours. Despite the viral activity and high rate of members openly declaring that they signed the petition, only 8 (8/133, 6%) messages from the group members, excluding the administrator, referred to the health issue, which was the topic of the group. No member expressed an opposite opinion to those presented by the administrator, and there was no debate in the form of exchanging opposite opinions.

Conclusions: The wildfire-like activity of the WhatsApp group and open declaration of signing the petition as a result of the mobilization campaign were not accompanied by any form of a debate related to the corresponding health issue, although the group members were predominantly health professionals, with a quarter of holding academic positions.

(J Med Internet Res 2020;22(8):e17051) doi:10.2196/17051

KEYWORDS

instant messaging; rumor; 5G mobile networks; participatory web; virality; infodemiology; infodemic

Introduction

The participatory web (Web 2.0) allows users to not only access and use the content but also create and exchange content during active interactions. Users are prosumers, in that they are both consumers and producers of the content [1]. When searching for knowledge, in particular for health-related knowledge, digital capabilities create a sophisticated environment described by a recently introduced metaphor – knowledge landscapes [2]. Knowledge is accessible (or hidden) in knowledge landscapes and can be approached by various individualized pathways shaped by personal, cultural, and societal contexts [3].

The participatory nature of the web also implies that online interactions influence attitudes, raise awareness, and mobilize

users. With health, use of the digital environment, both intentional and unintentional, affects health-related behaviors and the health status of the interacting individuals [4]. The online environment immensely augments the number of participants, speed, and geographical reach of interactions. However, the dynamics and patterns of these interactions, which are at the core of the new digital society, are still mostly unknown. Subsequently, their health effects are not well understood and could go in both directions, either beneficial or harmful [5].

Here, we describe retrospective analyses of the content and temporal dynamics of a WhatsApp group that was created as a mobilizing campaign to raise awareness of the possible health effects of 5G mobile networks. The specific group topic (ie, the prospective effect of 5G mobile networks on human health or environment) will not be discussed here, as the topic was used just as a general paradigm of the incoming complex technology and health uncertainties due to its application.

Instant messaging services have become one of the most popular and commonly used communication tools [6]. Immediacy, privacy, and cost-free use are features that contribute to the popularity of instant messaging [6]. WhatsApp as a social media platform is a mobile-based instant messenger characterized by the exchange of messages in real-time, usually between two users or among a group of users. With over 1.5 billion users worldwide, WhatsApp has emerged as a leader in the instant messaging industry, outranking other services such as Facebook, WeChat, Viber, and Skype [6,7]. WhatsApp is already widely used for various health purposes, including health education [8], rapid consultations in surgery, obstetrics, or in case of stroke [9,10], and as a tool for support groups like smoking cessation or weight management [11,12]. Moreover, due to its private and controlled environment, WhatsApp has become an ideal platform for discussing current matters of interest spanning news, politics, and activism [6].

Since WhatsApp groups emerge spontaneously and are visible only to their members, they are mostly unavailable to be studied. The WhatsApp group described in this study was created, active, and completed prior to the conception of this project; therefore, the analyses presented here are retrospective. The health concerns and messages exchanged in the group created a profile of interactions, similar to the dynamics of disease (ie, epidemiology), but in the context of digital activity (ie, infodemiology) [13]. Subsequently, the presented WhatsApp group analyses provide new insight into health-relevant participatory digital activity.

Methods

The initiator of the campaign, who undertook the role of the WhatsApp group administrator, invited individuals to the group via their mobile numbers. The WhatsApp group administrator was a medical doctor who, due to the nature of her or his profession and employment, was well connected with academics and medical professionals at local hospitals. The group was initiated at 2 pm on Saturday and had a total of 205 members. The group was active for 4 consecutive days (Saturday to Tuesday), after which it remained silent. Although 2 isolated messages were posted weeks later, they were not included in the analysis. The group was created during a 3-day weekend, as that Monday was a public holiday in Croatia. Moreover, August is regularly a period of vacations in Croatia. Thus, it is to be assumed that most of the group members were not at work until Tuesday (until the last day of the 4 consecutive days of group activity).

The idea to analyze the activity of the group was conceived after the group messaging had ceased. The WhatsApp group administrator was the only one who had knowledge of all members' identities. The messages were collected and deidentified, and their content and timing were analyzed. The descriptive characteristics of the group members (eg, gender, profession, academic achievements) were obtained from the group administrator. We conducted a frequency analysis of the messages using the collected content and timing. All necessary precautions were taken to maintain member anonymity and avoid potential member recognition. Ethical approval for the study was obtained from the University of Zagreb School of Medicine.

External links in the messages were followed to verify their viability. Claims of forwarding the discussion to other social networks (ie, Facebook) were not verified. The daily number of petition signatures was collected from the corresponding petition website, but the identity of signatories was not matched to the WhatsApp group members due to their anonymity. The Google Trends analysis was performed for a 7-day period, to analyze searches performed in Croatia with the search term "5G."

Results

Characteristics of the WhatsApp Group Members

The group consisted of 205 members (Table 1), with 187 members holding a university degree (187/205, 91%), and there were 140 health professionals (140/205, 68%), of which 125 were medical doctors (125/205, 61%) covering 16 different medical specializations. Furthermore, 75 (75/205, 37%) members had obtained a PhD, 50 (50/205, 24%) of whom held academic positions (professors and assistant professors). Although members were predominantly from the health sector, 8 members were electronic engineers and assumed to be educated to understand mobile communication technologies. All of the members were older than 18 years, and only 9 members were unemployed: 5 students and 4 retired individuals.

Table 1. Characteristics of the 205 members of the WhatsApp group.

Member characteristics	n (%)
Gender	
Women	106 (52)
Men	99 (48)
Academic status	
Current student	5 (2)
University degree	187 (91)
PhD	75 (37)
Profession	
Medical doctor	125 (61)
Medical doctor with specialization	97 (47)
Engineer	12 (6)
Electronic engineer	8 (4)
Retired	4 (2)
PhD with an academic position at a university	50 (24)

The WhatsApp group was initiated in August 2019 as a mobilizing campaign aimed to encourage members to sign a petition related to an incoming change to the city of Zagreb's (Croatia) legislation regarding mobile network antennas. In the first message, the administrator described the selection criterion for members from her or his contacts as those who "do not think only about themselves, here and now" but also consider the "long-term wellbeing of our kids and our planet Earth." Together with the main aim, to sign the petition, it was clear upfront that the general aim was to raise awareness of the possible health effects of 5G mobile networks. The administrator immediately declared that the members were free to leave the group whenever they liked.

Results of the WhatsApp Group

Of the 205 group members, 81 (81/205, 40%) were active by posting messages (Figure 1, Table 2). From a total of 133 messages, 28 were generated by the group administrator, and 105 were generated by group members, which is 1.3 messages per active group member, not taking into account the administrator. This indicates that the group was active not only regarding the number of messages but also in particular by wide member participation.

Figure 1. Number and time distribution of the members entering the discussion, as indicated by their first message.

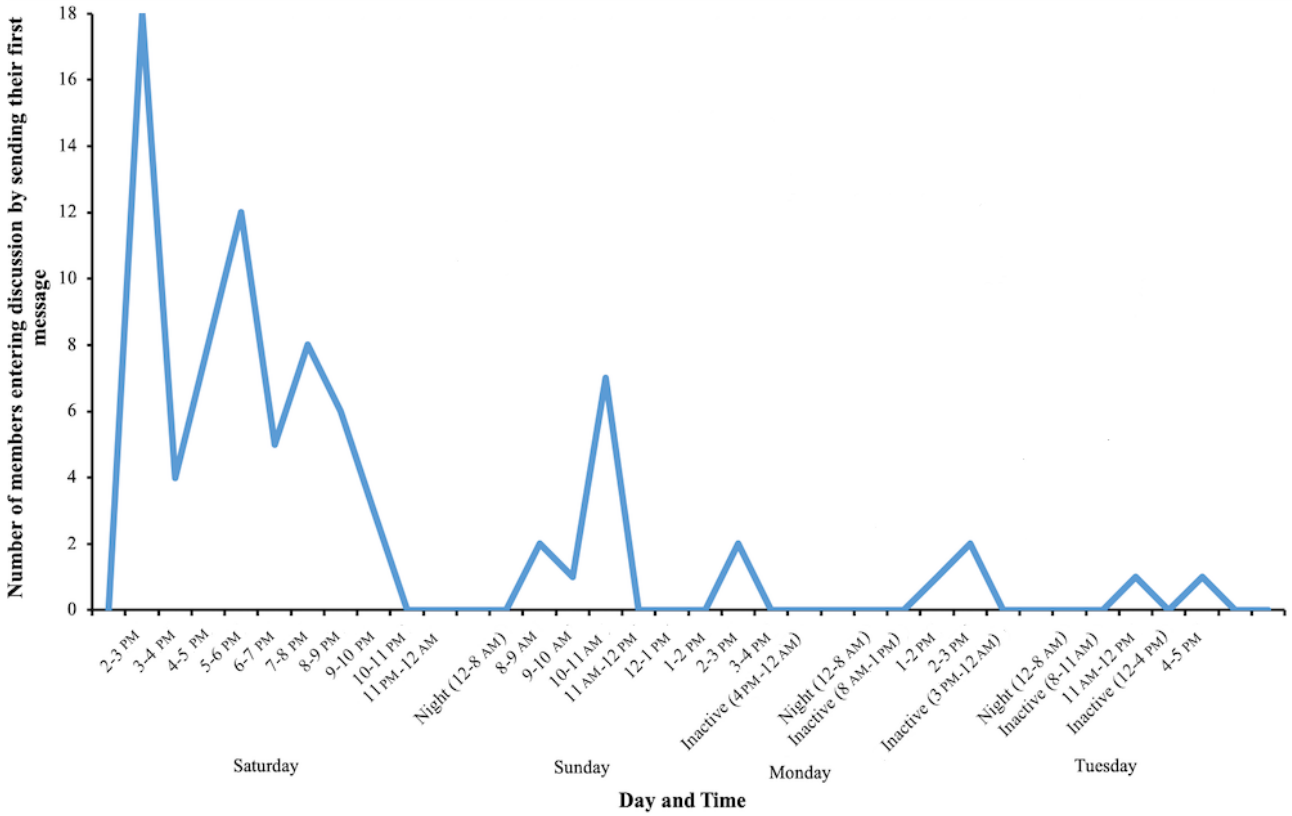
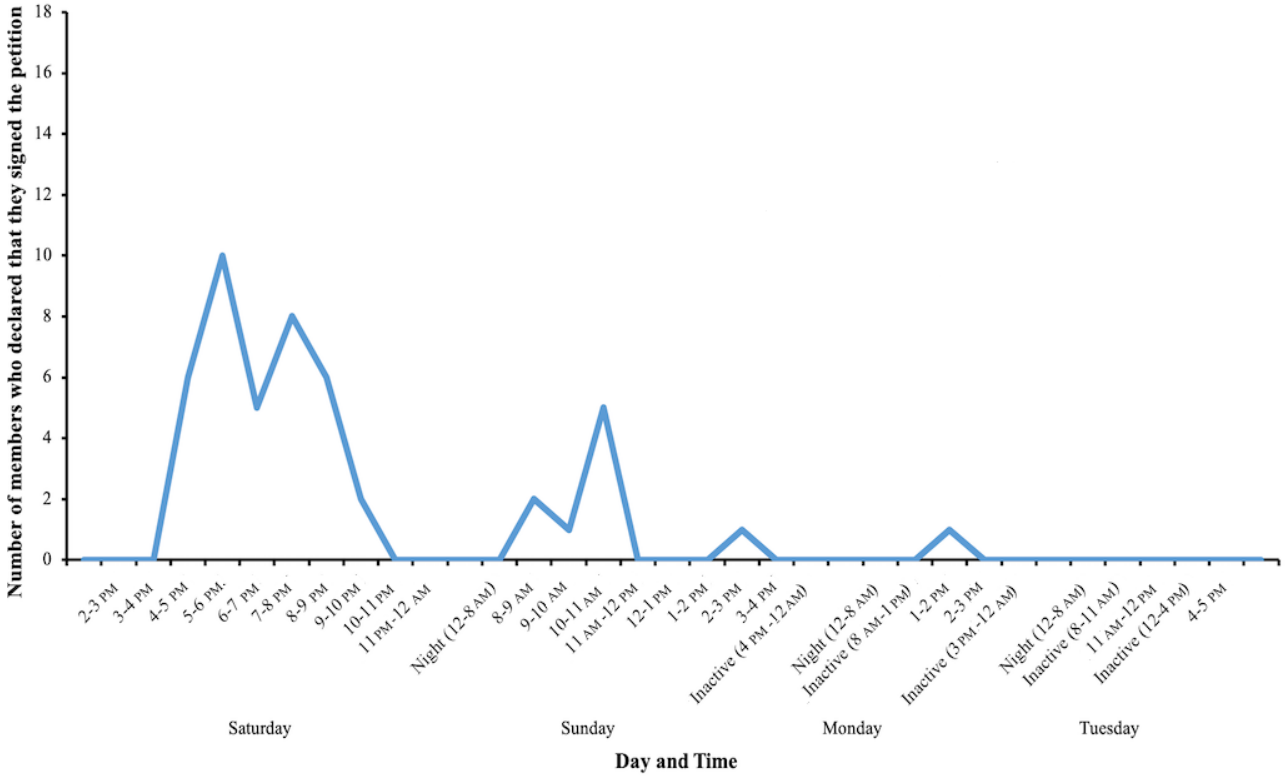


Table 2. Timing and extent of the different features of WhatsApp group events by the 205 members.

WhatsApp group events	1 hour	4 hours	10 hours (up until the first night)	26 hours	Total (75 hours)
Members who posted their first message, n	18	42	64	76	81
Messages posted, n					
All messages	30	72	103	126	133
Administrator messages	8	17	21	26	28
Members who declared that they signed the petition, n	0	16	37	46	47
Members who left the group, n					
Total	11	40	65	79	83
Left the group after signing the petition	0	7	14	17	17
Administrator responses to others, n					
Total	2	5	8	11	13
Positive responses	2	5	8	11	13
Negative responses	0	0	0	0	0
Emojis used, n					
Total	10	26	35	40	49
Positive feelings	9	24	33	35	44
Negative feelings	1	2	2	5	5
Messages that referred to 5G technology, n					
Total	5	12	14	19	19
Rumors	2	7	8	12	12
Not rumors	3	5	6	7	7

A notable result of the WhatsApp group was open declaration of signing the petition, which was the clear aim of the mobilizing campaign. The members of the group influenced their peers to sign the petition by declaring to others in their messages that they had signed it (Figure 2). A total of 47 (47/205, 23%) members openly declared signing the petition in their messages. From the petition website, it was possible to get insight into the daily statistics of new signatures. The day before the onset of the WhatsApp group, only 2 signatures were collected, which changed abruptly to 50 on the first day after the formation of the WhatsApp group and 42 more on the second day of the group activities. For the 3 subsequent days, the number of signatures was still above average (15-19 signatures/day), and 6 days after, it dropped to 4 signatures and stayed at this level in the days that followed. There were 137 total signatures collected during these 5 days coinciding with the WhatsApp group activity. Due to the anonymous analysis of the group activities, it was not possible to connect the names of the signatories published at the petition site with the identities of the WhatsApp group members. Therefore, it was not possible to know exactly how many signatures, beyond the 47 signatures that were openly declared in the WhatsApp group messages, were the effect of the mobilizing campaign. The group members themselves declared in 7 messages that they were sharing the discussion outside the WhatsApp group, and in 3 messages, Facebook was explicitly named as the platform for sharing. Facebook was the only other social network mentioned in the conversation.

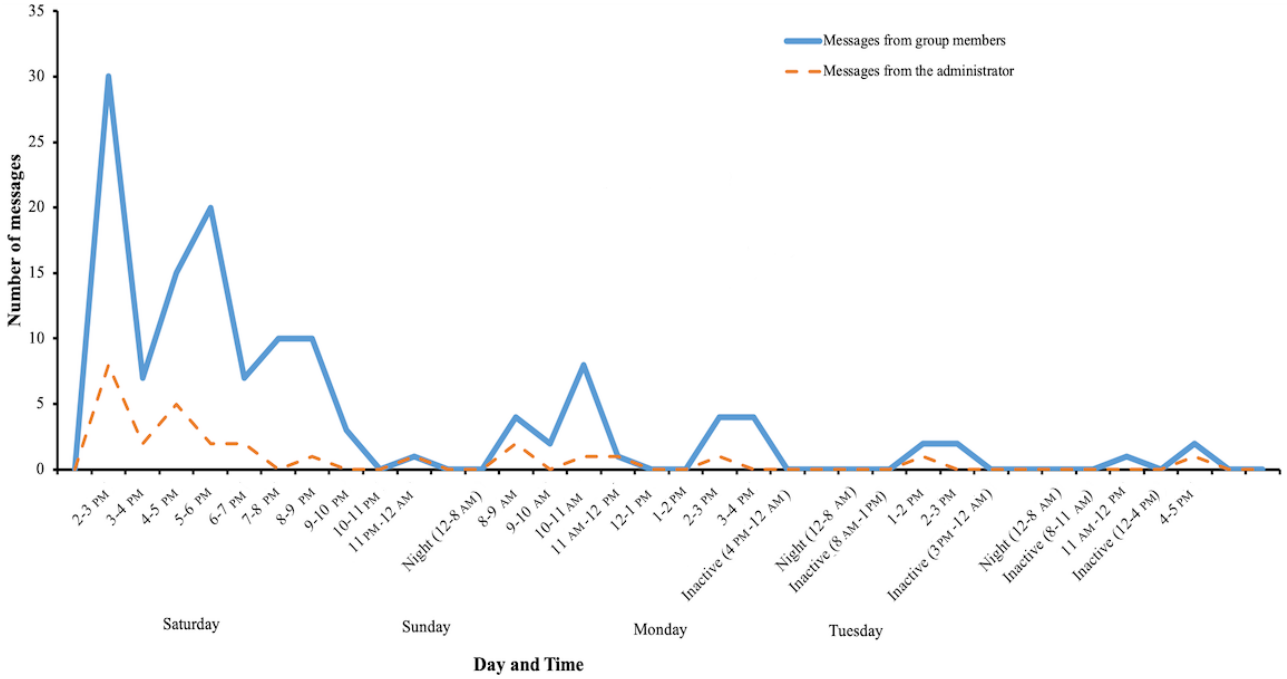
Figure 2. Number of members and the time distribution of their messages declaring they signed the petition.



Message Characteristics in the Group Activity

The actigraphy of the group showed that the total active time of the WhatsApp group was from Saturday 2 pm until 5 pm on Tuesday, or a total of 75 hours (Figure 3). This period included 4 periods of inactivity of 53 hours total (8 inactive hours Saturday night, 21 inactive hours Sunday night, 20 inactive hours Monday night, and 4 inactive hours during Tuesday daylight). This indicates that there were only 22 (29%) active hours from a total of 75 hours of continuous message exchanges.

Figure 3. Number and time distribution of the messages from the members versus those from the administrator.



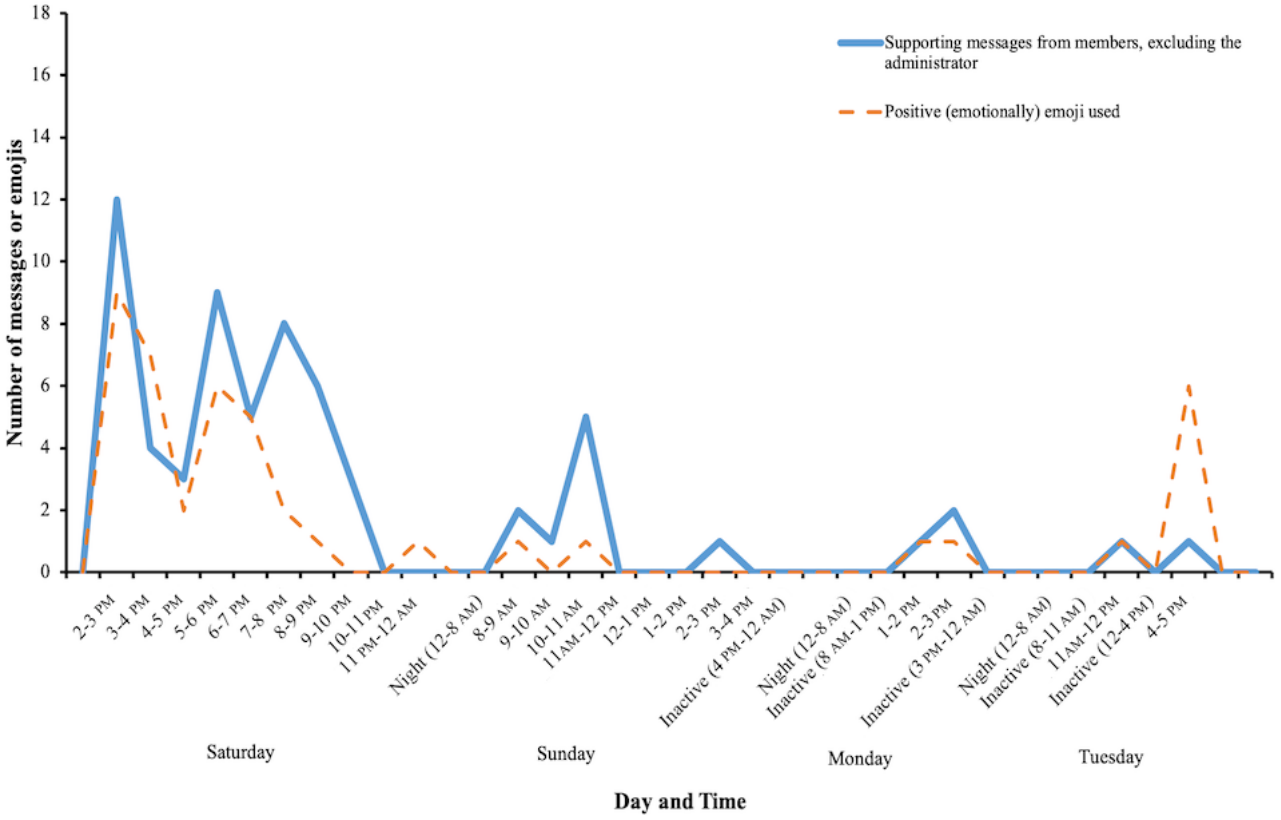
Most of the activity was concentrated in the first 26 hours of group existence (from Saturday 2 pm to Sunday 4 pm), during which time, 126 messages were exchanged (95% of the 133 total messages), and 76 members posted to the group (94% of 81 active members; Figures 1 and 3). Only 7 messages were exchanged, and 5 members became active in the next 49 hours. The 26-hour period included 8 hours of nighttime, from midnight to 8 am, when the group was inactive. Subsequently, regarding active hours, the first period contained 18 active hours to generate 126 messages, and the next period had only 4 active hours to generate 7 messages, although it extended through 3 calendar days or 49 hours (Figure 3).

The maximum activity was reached already in the first hour of the group's existence (30 messages, of which 8 were from the group administrator; 18 members posting messages; Figures 1 and 3). By the fourth hour of activity, 54% (72/133) of the total number of messages was reached, and during the same time, 52% (42/81) of active members posted their first message. The fourth hour was the second most active hour, with 20 messages exchanged. Before the night break (Saturday midnight, 10 hours of activity from the onset), 103 (103/133, 77%) messages were exchanged, and 64 (64/81, 79%) members produced a message. These data clearly showed that the onset of activity regarding both messaging and involvement of the group members was very rapid and concentrated at the very beginning of the activities.

Communication Content and Extent of the Debate

The role of the administrator was rather pronounced. The administrator generated 28 messages (21% of the total 133 messages), stirring the discussion and responding to the members. The administrator's responses to others were exclusively positive, confirming the group members' statements or praising the members personally. The members supported the administrator in 64 messages (64/133, 61%), and in 29 messages (29/133, 28%), they directly complimented the administrator (Figure 4). The emotional aspects of the messages were reflected by the use of emojis (total of 49), where 44 emojis were used depicting positive emotions, compared to only 5 negative emojis (Figure 4).

Figure 4. Number and time distribution of messages from the members supporting the administrator and of positive (emotionally) emojis posted.

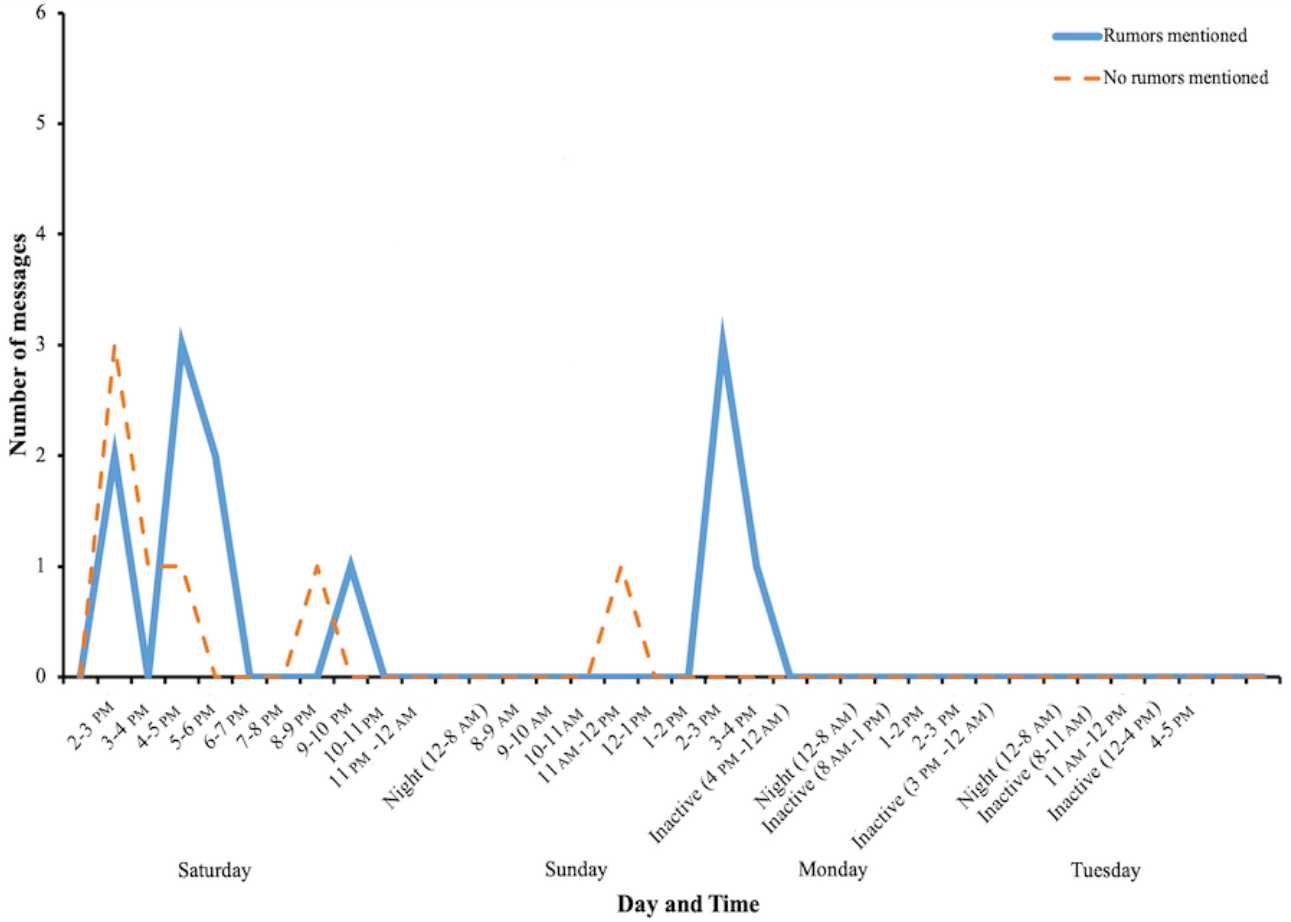


It should be noted that the administrator never asked for a comment or a response from a specific member, but only responded to the messages posted. Moreover, the administrator refrained from attempting to restart the activities after inactive periods, and the activities were restarted by the members themselves. Therefore, the final cessation of activities could be considered as a spontaneous turning off of the members' interest.

Despite the lively group activity, the specific topic of the WhatsApp group (ie, 5G mobile networks) was elaborated in only 19 (19/133, 14%) messages; of these, 11 were from the administrator. It should be emphasized that only 8 (8/133, 6%) messages from group members (excluding the administrator) referred to the specific topic of the group. This could be compared to the 7 messages posted, which were utterly unrelated to the exchanged WhatsApp group messages (eg, pictures of nature, mentioning excursions or bicycling together).

In the 19 messages dealing with the topic of 5G mobile networks, 10 outside resources were referenced. These included 1 link to a published research article, 1 link to a YouTube video created by Croatian national television, 1 link to a magazine article, and 7 links to various web pages. All of the outside references supported the viewpoints of the administrator. From the 19 messages discussing the topic of the group, 12 messages mentioned a rumor related to the topic in contrast to only 7 messages without rumors (Figure 5).

Figure 5. Number and time distribution of messages elaborating the group topic (influence of technology on health) that contained a rumor versus equivalent messages that did not contain a rumor.

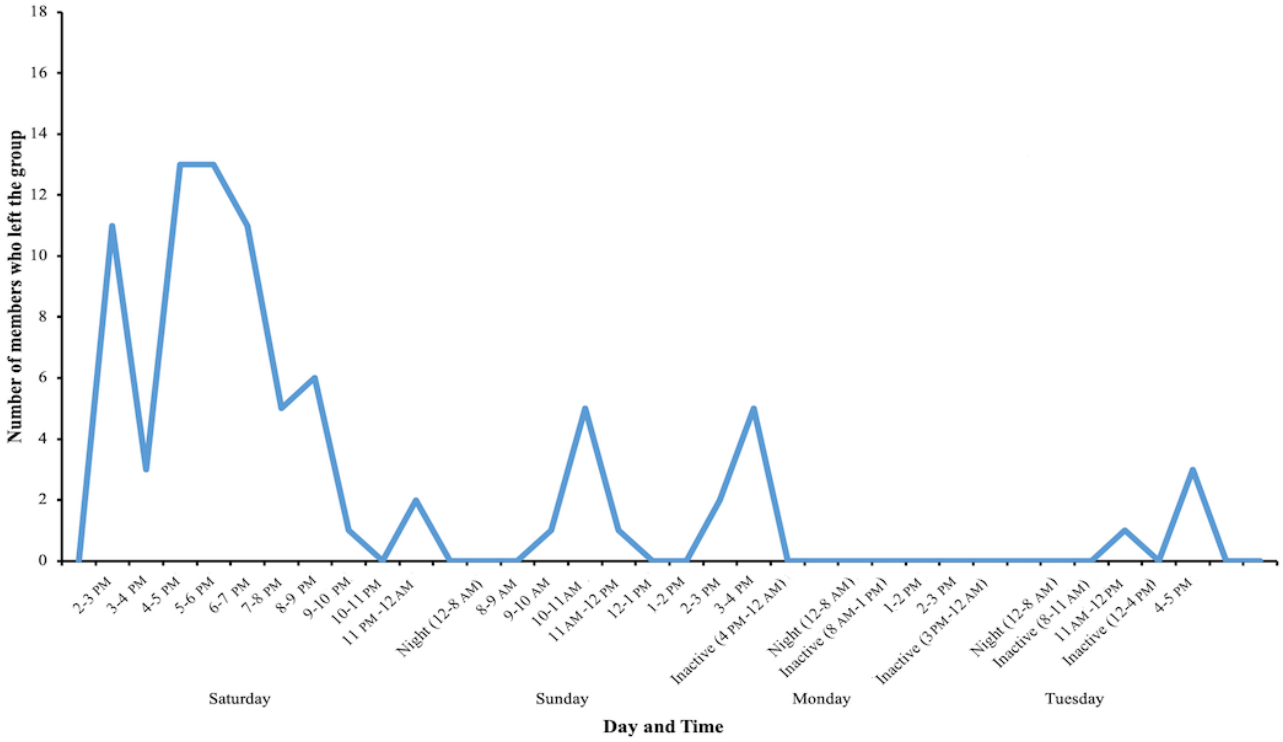


No message stated a counterargument or opposed the viewpoints of the administrator. Subsequently, there was no real debate in the form of a series of opposing messages exchanged. Only 2 messages could be considered vaguely opposite to those declared by the administrator: a message claiming "Sorry, this is not important," after which the member immediately left the group, and another message stating, "I consulted the friends from Faculty of Natural Sciences and Faculty of Engineering," after which the member immediately left the group. The administrator repeatedly urged in 8 messages that members should undertake their research and form their own opinion about the topic.

During the whole period of the WhatsApp group activity, 83 (83/205, 40%) members left the group (Figure 6); 17 members left the group after openly declaring to have signed the petition, and 9 members left even before the administrator sent the first message explaining the purpose of the group. Subsequently, if we could speculate on the silent expression of disagreement from the group members, 124 (124/205, 60%) members never produced a message, and 57 (57/205, 28%) left the group without letting the others know the reason they left.

The Google Trends results for "5G" as a topic in Croatia for this period were inconsistent. The results retrieved for the same period (7 days) but including the day before or the day after showed remarkably different results in relation to the days of WhatsApp group activity. Subsequently, the only resource to estimate whether there were any related Google searches by the group members was not effective.

Figure 6. Number and time distribution of members who left the group.



Discussion

Wildfire-Like Dynamic of the WhatsApp Group

The mobilizing campaign described here and executed through the creation of the WhatsApp group could be judged remarkably successful. The one measure of success was the fulfillment of the immediate aim of the campaign, signing the corresponding petition, which was openly declared by less than a quarter of the group members. The other measure of success was the involvement of the group members in lively digital activity (40% of the members), each posting 1.3 messages on average, one message per 2 minutes in the peak activity period. The fact that all members were acquaintances of the group administrator and had their telephone numbers in the administrator's contact list, indicated that possibly similar effects could be achieved without the help of the digital environment by merely contacting the members individually. However, what clearly distinguished the digital from "offline" scenarios was the speed and intensity of the group activities (ie, the wildfire-like dynamic) [14].

The description of wildfire-like dynamics is based on the rapid initiation of the message exchanges (with the first hour being the most active), involvement of the invitees in the messaging process (40% of the members), short duration of the intense activity (50% of messages were posted within the first 4 hours, and 95% were posted within the first 26 hours), and subsequent rapid stopping of the activities (after 4 days). Despite its short and intensive lifespan, the group achieved the initial purpose of mobilizing the members to sign the petition.

It could be argued whether the term wildfire is used correctly here, as it denotes both rapid and wide geographical spread of the message, considering that the group had only 205 members. However, as the WhatsApp group is closed, the "forest" of the members was limited, and when "burnt," the fire stopped. The digital wildfire was conceived originally as a term to depict the spread of rumors or malicious information causing significant harm [15]. Analyzing digital activity in the form of infodemiology has a similar connotation to the epidemiology of a disease, with diseases viewed negatively. We would like to argue against this normative approach to the phenomenon to be considered negative or bad *per se* (vs positive or good), as the same dynamics can be achieved regardless of the normative nature of the information. In this study, we intentionally avoided discussing the normative classification of the particular topic (ie, the relation between 5G mobile technologies and health). Whatever the topic, it deserves to be discussed in the digital environment of a social network. Considering that social networks represent a base for any grassroots movement, attaching a normative tag (bad vs good, harmful vs beneficial) could be at least controversial if not counterproductive in the sense of participatory democracy. Therefore, we claim here that wildfire-like dynamics are typical for the digital environment, as a specific feature of the digital society, and would happen in favorable circumstances regardless of the content, certainly not only if the content is malicious.

Another controversy related to using the term wildfire is its obvious overlap with the term virality, which refers to the viral-like spread of digital content. The observed dynamics visualized by the actigraphy in this study corresponds to viral event signatures [16]. Our choice to refer to the WhatsApp group activity as wildfire was based on the lack of specific content shared in the WhatsApp group being a "virus;" rather, the group shared a concern on a possible health issue more broadly. Subsequently, the term virality could be used to describe the spread of digital content, while wildfire could encompass overall mobilization and response of the users, which was initiated by viral content.

When discussing the circumstances contributing to the success of the mobilizing campaign and its wildfire-like attributes, four concepts could be considered: trust, motivation, situation, and narrative context [17]. All these concepts bring attention to the initiator and administrator of the WhatsApp group, who, in our opinion, should be praised for the group achievements. In the sense of "trust," the group members all knew the administrator personally, and there was a substantial number of messages supporting the administrator (64% of the total 205 messages) or directly complimenting the administrator (28% of the total). "Motivation" was stirred by the administrator being the most active group member in posting messages (21% of the messages) and providing the arguments, including external links to their own material, to support the signing of the petition. The enhancing contributor was the "situation" of the vacation period and long weekend, providing less distraction from daily routines. Finally, the predominant "narrative context" was the appealing fight of the individuals versus the governing elite (the parliament of the city of Zagreb) and versus corporate interests (through the introduction of new lucrative technology). The posteriori analyses showed that in every dimension analyzed — trust, motivation, situation, and narrative context — the group had all prerequisites to acquire wildfire-like features.

The analysis of the WhatsApp group dynamic presented the features of emotional contagion as well, where emotions were shared among the interacting netizens in the digital environment, influencing the spread of information and attitudes [18,19]. The emotions were specifically shared through emojis, which were abundantly used in composing the message (37% of the messages), showing predominantly positive emotions [20]. Interestingly, the dynamics of this study's WhatsApp group can be compared to the recently published emotional contagion simulation on rumor refuting, which indicates that after initial input, the group will stabilize in 4 days (same as in the group in this study), and that none of the subjects would act negatively to what has happened but rather positively (agreeing with the initial input) or neutrally (indifferent to the topic) [21]. Moreover, the success of the current WhatsApp group can be related to the fact that most of the members were naïve both to the digital platform and to the topic, while subsequent attempts for mobilization on some other topic can fail due to the cry-wolf phenomenon or neutral status of the members [22].

Absence of Debate

The analysis of the WhatsApp group dynamics indicated that its dynamics can be related to the previously suggested mechanisms of rumor distribution and emotional contagion. However, these mechanisms, although applicable to the analyzed WhatsApp group, are not relevant to the topic of discussion. Subsequently, we were rather keen to identify the comments that elaborated the specific topic of the group (ie, 5G mobile communications and health). The advantage of the group was that its members were predominantly health professionals with university degrees, and many held academic positions. Therefore, it is to be assumed that the members were educated and had above-average capabilities to grasp complex health-related and technology-related issues. However, although the dynamic of the messaging had a wildfire-like feature and some group members responded to the mobilizing campaign by signing the petition, the posted messages elaborating the topic of the group were surprisingly rare. What was surprising is that none of the messages opposed the administrator, and no sequence of messages was exchanged expressing opposing opinions that would represent the presence of a digital debate. The ambiguous sign of disapproval could be inferred by members leaving the group or "lurkers" not posting anything; however, they certainly did not act in the sense of debating the issue.

The absence of opposing opinions and debate is rather alarming, as we assume that the digital environment is an ideal forum for the exchange of divergent opinions and a place where an eventual social dispute could be settled [23]. As the digital environment poses no barriers, the societally relevant controversies and subsequent reconciliations are open to everyone interested. The digital capabilities allow the relevant discussions to be amplified and involve a wide range of participants (ie, total humanity in the idealistic sense). Subsequently, digital technology could serve as a basis for the all-embracing participatory democracy [24].

However, in the current example, this entire concept was missing, although the participants, according to their abilities (university education and academic positions) and professions (health and information technology professionals), were well suited to engage in a debate. The topic of the current study is rather specific to claim whether this was an exception or a rule, as we lacked the eventual comparison with similar situations; this is a major limitation of this study.

Several elements may indicate reasons why the debate was not present. First, the exchange of the messages was rather hierarchical, from the administrator toward the members and back to the administrator, with no messaging among the members. Subsequently, any opposing opinion would be a direct confrontation to the administrator, challenging the hierarchy and not just a dispute between some of the members. None of the members gained an "extra force" (as opinion leader, motivator, or central node in the network) during the message exchange. Not to be forgotten is that all members were the administrators' contacts in the offline world. Additionally, the wildfire-like effect of fast message exchanges left less time for individual research on the topic (despite the fact that the administrator urged the members to form their own opinion through their approach). Finally, the timing of the activities was during the weekend and vacation period, and we assume that most of the participants were at home or out of town, leaving them in a position to be unable to meet each other. The assumption that nonvirtual real-life conversation is needed for the consolidation of opinion, regardless of online virtual influences, is a topic worth exploring in the future.

Conclusions

In conclusion, the WhatsApp group had wildfire-like features and served to mobilize the members to act publicly in relation to the health issue. However, the very topic of controversy (effects of 5G mobile technologies on health) was not challenged in the form of debate, although the group members were predominantly health professionals and held academic positions.

Acknowledgments

This study was supported by the EU European Regional Development Fund, Operational Programme Competitiveness and Cohesion, grant agreement No.KK.01.1.1.01.0007, CoRE – Neuro. We are grateful to the administrator of the WhatsApp group for providing the anonymized data and group member analysis, but consequently the administrator name is not disclosed here. In addition, a sincere thank you to Anni Hasratian for her diligent proofreading of this paper.

Conflicts of Interest

None declared.

References

1. Ritzer G, Jurgenson N. Production, Consumption, Prosumption. Journal of Consumer Culture 2010 Mar 09;10(1):13-36. [doi: 10.1177/1469540509354673]
2. Svalastog AL, Allgaier J, Martinelli L, Gajovic S. Distortion, confusion, and impasses: could a public dialogue within Knowledge Landscapes contribute to better communication and understanding of innovative knowledge? Croat Med J 2014 Feb;55(1):54-60 [FREE Full text] [doi: 10.3325/cmj.2014.55.54] [Medline: 24577828]
3. Gajović S, Svalastog AL. When communicating health-related knowledge, beware of the black holes of the knowledge landscapes geography. Croat Med J 2016 Oct;57(5):504-509. [doi: 10.3325/cmj.2016.57.504] [Medline: 27815941]
4. Kopilaš V. Person-centered care and access to drugs in the digital environment: to cure or to sell. Croat Med J 2019 Apr;60(2):174-176. [doi: 10.3325/cmj.2019.60.174] [Medline: 31044591]
5. Keselman A, Arnott Smith C, Murcko AC, Kaufman DR. Evaluating the Quality of Health Information in a Changing Digital Ecosystem. J Med Internet Res 2019 Feb 08;21(2):e11129 [FREE Full text] [doi: 10.2196/11129] [Medline: 30735144]
6. Gil de Zúñiga H, Ardèvol-Abreu A, Casero-Ripollés A. WhatsApp political discussion, conventional participation and activism: exploring direct, indirect and generational effects. Information, Communication & Society 2019 Jul 20:1-18. [doi: 10.1080/1369118x.2019.1642933]
7. Barbosa S, Milan S. Do Not Harm in Private Chat Apps: Ethical Issues for Research on and with WhatsApp. Westminster Papers in Communication and Culture 2019;14(1):49-65. [doi: 10.16997/wpcc.313]
8. Coleman E, O'Connor E. The role of WhatsApp® in medical education; a scoping review and instructional design model. BMC Med Educ 2019 Jul 25;19(1):279. [doi: 10.1186/s12909-019-1706-8] [Medline: 31345202]
9. Carmona S, Alayed N, Al-Ibrahim A, D'Souza R. Realizing the potential of real-time clinical collaboration in maternal–fetal and obstetric medicine through WhatsApp. Obstet Med 2018 Mar 19;11(2):83-89. [doi: 10.1177/1753495x18754457] [Medline: 29997691]
10. Calleja-Castillo JM, Gonzalez-Calderon G. WhatsApp in Stroke Systems: Current Use and Regulatory Concerns. Front. Neurol 2018 May 31;9:388. [doi: 10.3389/fneur.2018.00388] [Medline: 29904369]
11. Durmaz S, Ergin I, Durusoy R, Hassoy H, Caliskan A, Okyay P. WhatsApp embedded in routine service delivery for smoking cessation: effects on abstinence rates in a randomized controlled study. BMC Public Health 2019 Apr 8;19(1):387. [doi: 10.1186/s12889-019-6727-z] [Medline: 30961557]
12. Fischer A, Chadyiwa M, Tshuma N, Nkosi V. Acceptability of Mobile Health Interventions to Increase Diabetic Risk Factor Awareness Among the Commuter Population in Johannesburg: Descriptive Cross-Sectional Study. JMIR Diabetes 2019 Sep 20;4(3):e12600. [doi: 10.2196/12600] [Medline: 31586363]
13. Eysenbach G. Infodemiology: the epidemiology of (mis)information. The American Journal of Medicine 2002 Dec;113(9):763-765. [doi: 10.1016/S0002-9343(02)01473-0] [Medline: 12517369]
14. Webb H, Burnap P, Procter R, Rana O, Stahl BC, Williams M, et al. Digital Wildfires: Propagation, Verification, Regulation, and Responsible Innovation. ACM Trans. Inf. Syst 2016 May 05;34(3):Article 15. [doi: 10.1145/2893478]
15. World Economic Forum. Digital Wildfires in a Hyperconnected World. 2013 URL: http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/ [accessed 2019-12-11]
16. Hemsley J. Studying the viral growth of a connective action network using information event signatures. First Monday 2016 Jul 16;21(8):6650 [FREE Full text] [doi: 10.5210/fm.v21i8.6650]
17. DiFonzo N. Rumour research can douse digital wildfires. Nature 2013 Jan 8;493(7431):135-135. [doi: 10.1038/493135a] [Medline: 23302823]
18. Hatfield E, Cacioppo JT, Rapson RL. Emotional contagion. Cambridge, UK: Cambridge University Press; 1994.
19. Ferrara E, Yang Z. Measuring Emotional Contagion in Social Media. PLoS ONE 2015 Nov 6;10(11):e0142390. [doi: 10.1371/journal.pone.0142390] [Medline: 26544688]
20. Bai Q, Dan Q, Mu Z, Yang M. A Systematic Review of Emoji: Current Research and Future Perspectives. Front. Psychol 2019 Oct 15;10:2221. [doi: 10.3389/fpsyg.2019.02221] [Medline: 31681068]

21. Zeng R, Zhu D. A model and simulation of the emotional contagion of netizens in the process of rumor refutation. Sci Rep 2019 Oct 2;9(1):14164. [doi: 10.1038/s41598-019-50770-4] [Medline: 31578405]
22. Norman ID, Aikins M, Binka F. Earthquake hoax in Ghana: exploration of the Cry Wolf hypothesis. J Public Health Africa 2012 Apr 26;3(1):e14. [doi: 10.4081/jphia.2012.e14] [Medline: 28299086]
23. Orr D, Baram-Tsabari A, Landsman K. Social media as a platform for health-related public debates and discussions: the Polio vaccine on Facebook. Isr J Health Policy Res 2016 Nov 10;10(5):34. [doi: 10.1186/s13584-016-0093-4] [Medline: 27843544]
24. Shirky C. Here Comes Everybody: The Power of Organizing Without Organizations. London, UK: Penguin Books Ltd; Feb 28, 2008.

Edited by G Eysenbach; submitted 13.11.19; peer-reviewed by N Buzas, H Webb; comments to author 16.12.19; revised version received 10.02.20; accepted 27.02.20; published 10.08.20.

Please cite as:
Kopilaš V, Gajović S
Wildfire-Like Effect of a WhatsApp Campaign to Mobilize a Group of Predominantly Health Professionals With a University Degree on a Health Issue: Infodemiology Study
J Med Internet Res 2020;22(8):e17051
URL: https://www.jmir.org/2020/8/e17051
doi:10.2196/17051
PMID:32442138

Original Paper

Online Japanese-Language Information on Lifestyle Factors Associated With Reduced Fertility: Content Analysis

Rie Yokota[1], MPH; Tsuyoshi Okuhara[2], PhD; Haruka Ueno[3], PhD; Hiroko Okada[2], PhD; Emi Furukawa[1], MD; Takahiro Kiuchi[2], MD, PhD

[1]Department of Health Communication, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

[2]Department of Health Communication, School of Public Health, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

[3]Department of Health and Dietetics, Faculty of Health and Medical Science, Teikyo Heisei Unievrsity, Toshima-ku, Tokyo, Japan

Corresponding Author:
Rie Yokota, MPH
Department of Health Communication
Graduate School of Medicine
The University of Tokyo
7-3-1 Hongo
Bunkyo-ku, Tokyo, 113-8655
Japan
Phone: 81 358006549
Email: yokotarie-tky@umin.ac.jp

Abstract

Background: Approximately one-third of Japanese couples currently worry or previously worried about infertility. To develop strategies for the primary prevention of infertility as a population approach, it is important for the general population to be knowledgeable about fertility and infertility. The internet may contribute to the dissemination of information regarding infertility and fertility. However, few studies have examined online information about fertility.

Objective: This study aimed to quantitatively examine online Japanese-language information about lifestyle factors associated with reduced fertility.

Methods: We conducted online searches, using the 10 search terms with the highest numbers of searches that people hoping to conceive are likely to input in two major search engines in Japan (Google Japan and Yahoo! Japan). From the 2200 retrieved websites, 1181 duplicates and 500 websites unrelated to our objective were excluded, resulting in a final dataset of 519 websites. Coding guidelines were developed for the following lifestyle factors associated with reduced fertility: sexually transmitted diseases, psychological stress, cigarette smoking, alcohol use, nutrition and diet, physical activity and exercise, underweight, overweight and obesity, and environmental pollutants.

Results: In terms of the website author's professional expertise, 69.6% of the coding instances for the selected lifestyle factors were mentioned by hospitals, clinics, or the media, whereas only 1.7% were mentioned by laypersons. Psychological stress (20.1%) and sexually transmitted diseases (18.8%) were the most frequently mentioned lifestyle factors associated with reduced fertility. In contrast, cigarette smoking, alcohol use, nutrition and diet, physical activity and exercise, underweight, overweight and obesity, and environmental pollutants were mentioned relatively infrequently. The association between reduced fertility and sexually transmitted diseases was mentioned significantly more frequently by hospitals and clinics than by the media (P<.001). The association between reduced fertility and nutrition and diet was mentioned significantly more frequently by the media than by hospitals and clinics (P=.008). With regard to the sex of the target audience for the information, female-specific references to psychological stress, sexually transmitted diseases, nutrition and diet, underweight, physical activity and exercise, and overweight and obesity were significantly more frequent than were male-specific references to these lifestyle factors (psychological stress: P=.002, sexually transmitted diseases: P<.001, nutrition and diet: P<.001, underweight: P<.001, physical activity and exercise: P<.001, overweight and obesity: P<.001).

Conclusions: Of the lifestyle factors known to be related to reduced fertility, cigarette smoking, alcohol use, and male-specific lifestyle factors are mentioned relatively infrequently in online information sources in Japan, and these factors should be discussed more in information published on websites.

(J Med Internet Res 2020;22(8):e19777) doi:10.2196/19777

KEYWORDS

content analysis; online information; lifestyle factor; fertility; infertility; reproductive health

Introduction

Background

At the 1994 United Nations International Conference on Population and Development, reproductive health was defined as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity, in all matters relating to the reproductive system and to its functions and processes" [1]. Infertility is increasingly acknowledged as a global public health issue by the World Health Organization [2], and reproductive health implies that "people are able to have the capability to reproduce and the freedom to decide if, when and how often to do so" [1]. To aid decision making concerning fertility and assist the reproductive-aged population in optimizing their fertility and reproductive health, knowledge of the lifestyle factors associated with reduced fertility is crucial [3,4].

Infertility is defined as the failure to achieve conception following at least 12 months of unprotected sexual intercourse [5]. Infertility affects as many as 186 million people worldwide [6], and about 10% to 15% of couples experience infertility [7]. In 2015, as many as one in three Japanese couples (approximately 35.0%) reported currently or previously worrying about infertility, and more than one-sixth (approximately 18.2%) reported currently or previously undergoing screening for infertility or trying to achieve pregnancy through assisted reproduction technologies [8].

Numerous factors may contribute to reduced fertility in men and women. In addition to genetic background and reproductive history, environmental factors and current lifestyle habits have been proposed as causes of male and female infertility [9]. According to large systematic reviews and meta-analyses, the lifestyle factors associated with reduced fertility are (1) sexually transmitted diseases, (2) psychological stress, (3) cigarette smoking, (4) alcohol use, (5) nutrition and diet, (6) physical activity and exercise, (7) underweight, (8) overweight and obesity, and (9) environmental pollutants [10-27]. Given the situation of one-third of Japanese couples currently or previously worrying about infertility, for those who are trying to conceive or hope to have a child in the future, knowledge about the lifestyle factors associated with reduced fertility may help to prevent infertility.

A recent study found that about 60% to 70% of the reproductive-aged population in Japan responded incorrectly to a question about the association between cigarette smoking and reduced fertility [28]. Similarly, when asked about the association between female overweight and reduced fertility, approximately 80% to 90% of the reproductive-aged population in Japan answered incorrectly [28]. These findings indicate a lack of knowledge about the associations between reduced fertility and lifestyle factors such as smoking cigarettes and overweight and obesity among the reproductive-aged population in Japan.

In terms of sex differences, several studies have shown that, compared with women, men are less knowledgeable regarding the associations between lifestyle factors and reduced fertility [3,29]. Although most men (88.5%) regarded themselves as knowledgeable on this topic, only half of the men (53.1%) participating in a population-based survey were able to identify the lifestyle factors associated with reduced fertility [4]. A recent survey in Japan found that only 46.4% of reproductive-aged men and 56.7% of reproductive-aged women were knowledgeable about male infertility factors, which constitute approximately 50% of infertility cases [30]. Moreover, according to the same survey, 38.0% of men reported that they did not intend to undergo a semen examination at a medical institution because they thought they had no infertility problems [30]. Because infertility is presumed to be a women's issue [31], the reproductive-aged population has relatively low knowledge about issues related to male fertility [32].

Currently, the internet is a preferred and common source of health information [33]. About 72% of internet users access health-related information via the web [32]. Health information is conveyed to targeted audiences to influence their attitudes or behaviors [34], and the presentation of information and framing used in media portrayals affect the general public's understanding of lifestyle factors associated with reduced fertility [35]. Thus, online information about fertility influences the general population's knowledge about conception.

Prior Work

Previous studies regarding information on infertility and fertility in the media have examined (1) newspaper reports about assisted reproductive technology [35]; (2) information on clinic websites [36]; (3) the readability, suitability, and quality of online information [32,37]; (4) online videos made by laypersons and informational infertility-related educational videos [38]; and (5) online emotional support and social media on infertility-seeking to reduce isolation [39,40]. As mentioned above, although people who are trying to conceive tend to have relatively low levels of knowledge about lifestyle factors associated with reduced fertility, especially male infertility, to our knowledge, few studies have investigated the content of online information on the lifestyle factors associated with reduced fertility.

People who are diagnosed with infertility and those receiving infertility treatment may receive accurate information on fertility from health care professionals. However, no studies have examined what kinds of online content are accessible to the general population, although online information is especially important for people who do not visit medical institutions but are trying to conceive or hoping to have a child in the future. Thus, it is crucial to examine online information on the lifestyle factors associated with reduced fertility.

Goal of the Study

The study aimed to (1) quantitatively examine the information on lifestyle factors associated with reduced fertility accessible to people hoping to have a child who seek information regarding fertility on the internet and (2) identify the characteristics of this information in terms of lifestyle factors, the webpage author's professional expertise, and the sex of the target audience (ie, information for men, women, or both).

Methods

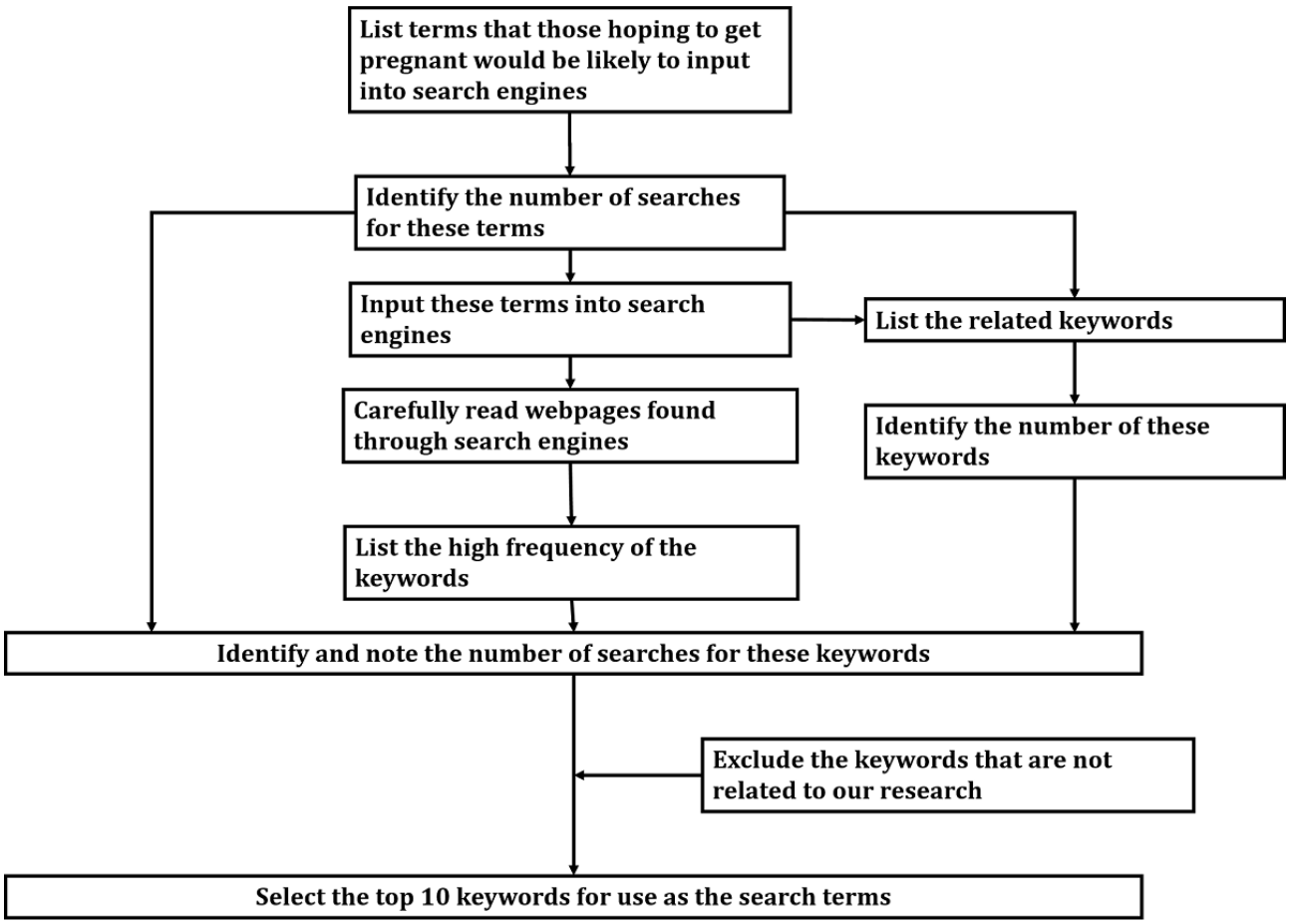
Search Engine

We used a Japanese-language search string input into the two most popular search engines in Japan, Google Japan [41] and Yahoo! Japan [42]. Google Japan and Yahoo! Japan accounted for roughly 75% and 19% of all internet searches, respectively, at the end of October 2019 [43].

Search Terms

A flow diagram depicting the search terms generation procedure is presented in Figure 1. The search terms were determined using the following procedure. Few studies have used content analysis to examine online information about the lifestyle factors associated with reduced fertility. Although English speakers frequently refer to "fertility," Japanese speakers do not commonly use the word "ninyousei" (fertility). Thus, deriving search terms from the previous literature was difficult. The main search terms thought to be used by people who are trying to conceive were derived in the following manner. First, we listed terms such as "ninshin" (pregnancy), "ninkatsu" (trying to conceive), and "funin" (infertility) to capture the most common and basic keywords that those hoping to get pregnant would be likely to input into search engines. We identified the number of searches for each of these keywords using a keyword search calculation tool [44]. Although both Google (Keyword Planner) and Yahoo (keyword advice tool) have tools to check the number of monthly searches, we were not able to use these services because we would need to register as a corporation to do so. Therefore, we used the keyword search calculation tool produced by Devo Inc [44]. This keyword search tool provides the total number of searches in one month for the search term, as well as the total number of searches for the top 50 webpages, per Google Japan and Yahoo! Japan [44]. After carefully reading websites found through Google Japan and Yahoo! Japan, we listed additional keywords such as "bebimachi" (waiting for a baby) that people who are trying to conceive would be likely to input. Then, we described the number of searches for each keyword using the abovementioned keyword search calculation tool. The top 6 keywords in terms of the number of searches were "funin" (infertility), "ninkatsu" (trying to conceive), "ninshindekinai" (I cannot get pregnant), "akachanhosii" (hoping to have a baby), "shizenninshin" (natural conception), and "bebimachi" (waiting for a baby).

Figure 1. Flow diagram of the search terms generation procedure.



As a next step, we entered these 6 keywords into Google Japan and Yahoo! Japan and listed the related keywords shown by the search engines (eg, “ninkatsu sapurimento” [trying to conceive AND supplement]). Related keywords are words that are frequently searched in combination with the main keywords [45]. Additionally, we described the number of searches for related keywords such as “funin genin” (infertility cause) by entering the top 6 keywords into the keyword search calculation tool [44]. The objective of this study was not to examine the information available to people who have already been diagnosed with infertility or those who are receiving infertility treatment. Rather, the objective was to examine the information accessible to those who are trying to conceive or hoping to have a child in the future. Therefore, we excluded the following from the listed search phases keywords concerning specific commercial products or services, financial support for infertility, and specific risk factors for infertility such as endometriosis or azoospermia. Of the remaining keywords, the top 10 were listed using the keyword search calculation tool. Using Google Trends, we then confirmed that the number of searches for the top 10 search terms did not represent a surge, compared with the history of searches over the last 10 years [46].

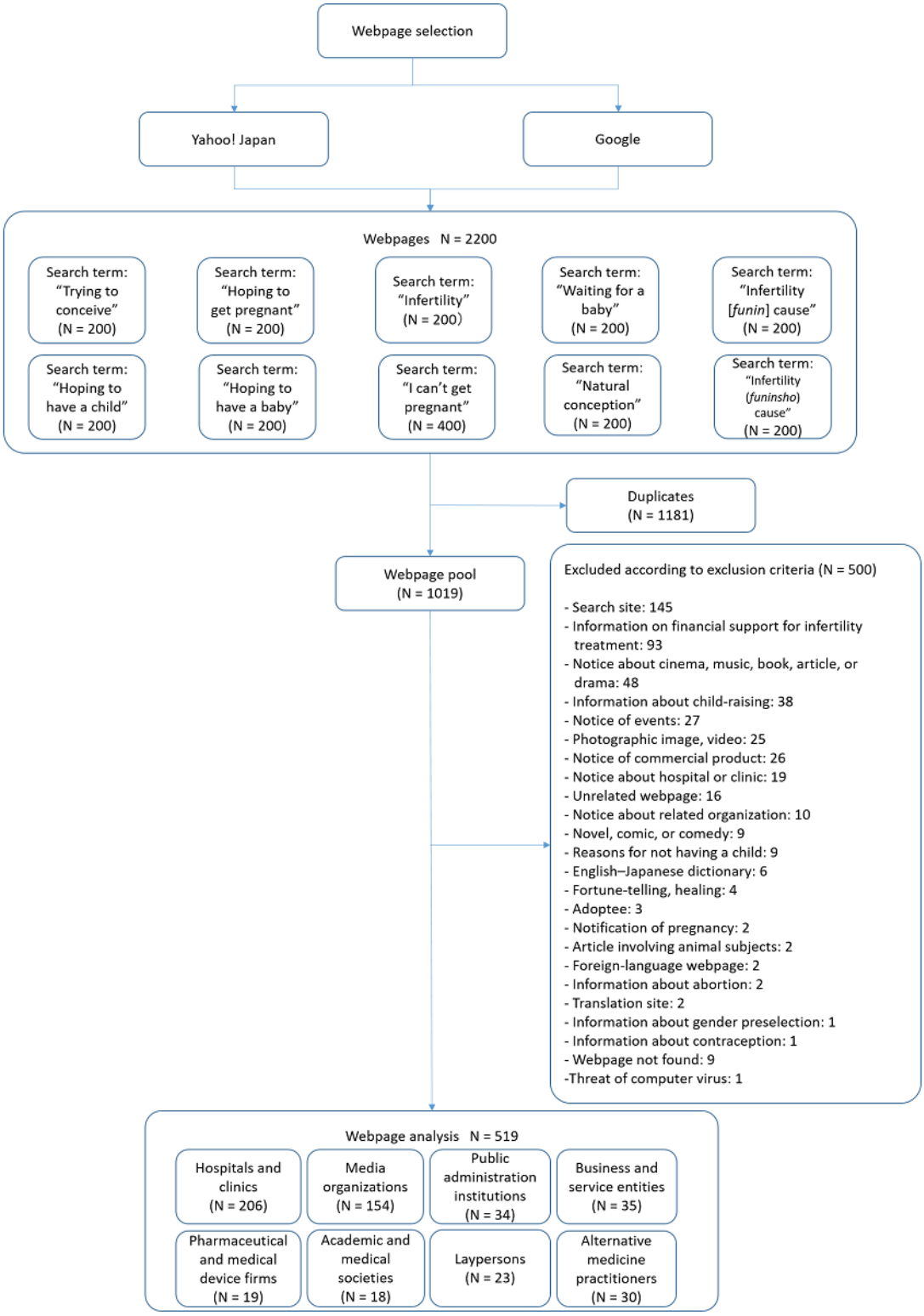
At the end of this process, the final search was performed using the following keywords: “ninkatsu” (trying to conceive), “ninshinshitai” (hoping to get pregnant), “funin” (infertility), “bebimachi” (waiting for a baby), “funin genin” (infertility cause), “kodomohosii” (hoping to have a child), “akachanhosii” (hoping to have a baby), “ninshindekinai” (I cannot get pregnant), “shizenninshin” (natural conception), and “huninsho genin” (infertility cause).

Material Collection

The unit of analysis in this study was the webpages suggested by the search engines. Because websites differ considerably in size, coding entire websites could introduce biases based on size [47,48]. Moreover, it has been found that visitors rarely look though all the pages of a website [47,48]. Additionally, a previous study indicated that those who used search engines stayed on each website for a mean of only 1 minute and 9 seconds (median 37 seconds) [49]. Therefore, for each webpage suggested by the search engines, this study analyzed the major part of the website that visitors were likely to read.

The first 100 results retrieved using each search engine were collected by the first author. Each time we input search keywords or changed search engines, the search history and cookies were cleared. We conducted online searches from October 21 to November 3, 2019. The flow diagram of the webpage selection is presented in Figure 2. Of the total 2200 webpages, 1181 duplicates were excluded. Additionally, 500 webpages not directly related to the objective of this study were excluded. Ultimately, 519 webpages were included in the analysis. The URL and ranking of each result for the 2200 webpages were saved in Excel (Microsoft Corporation). The 519 webpages to be analyzed were stored as PDFs and, if that was impossible, they were copied and pasted into Word (Microsoft Corporation).

Figure 2. Flow diagram of the webpage selection.



Coding Guidelines and Procedures

There are no current clinical practice guidelines in Japan regarding lifestyle factors associated with reduced fertility. We created a list of lifestyle factors associated with reduced fertility by reviewing information from the Japan Society of Obstetrics and Gynecology [50], Ministry of Health, Labour and Welfare [51], Cochrane Library [5], British Fertility Society [52], American Society for Reproductive Medicine [53], and previous research [10-27]. This list was then reviewed by two obstetrician-gynecologists. Through discussion, we reached consensus to include the following 9 lifestyle factors associated with reduced fertility: (1) sexually transmitted diseases, (2) psychological stress, (3) cigarette smoking, (4) alcohol use, (5) nutrition and diet, (6) physical activity and exercise, (7) underweight, (8) overweight and obesity, and (9) environmental pollutants. These lifestyle factors and the major references supporting their inclusion are summarized in Table 1.

Each webpage was categorized according to the author's professional expertise: hospitals and clinics, media organizations, public administration institutions, business and service entities, academic and medical societies, pharmaceutical and medical device firms, alternative medicine practitioners, or laypersons. Hospitals and clinics indicated that the content appeared on the website of a hospital or clinic, including blogs written by clinicians. Media organizations indicated that the content appeared on the website of a mass media organization such as a newspaper, magazine, or news site. Public administration institutions indicated that the content appeared on the website of a public organization such as the government, a municipality, public health care center, or specialized public consultation support center. Business and service entities indicated that the content appeared on the website of an enterprise such as a marriage support service company, bridal company, recruiting company, children's goods retail business, cooking school, counseling organization, architectural firm, stationery company, or hospital search service company. Academic and medical societies indicated that the content appeared on the website of a medical society such as an association of physicians or medical institutes. Pharmaceutical and medical device firms indicated that the content appeared on the website of a pharmaceutical company, medical device firm, or pharmacy. Alternative medicine practitioners indicated that the content appeared on the website of a practitioner of alternative medicine such as osteopathy, herbal medicine Kampo, acupuncture, moxibustion, or yoga or on the website of a health food company. Laypersons meant that the content was written by persons who were currently trying or had previously tried to conceive, including patients.

Table 1. Lifestyle factors and major references supporting their inclusion.

Lifestyle factor and reference	Relevant sex		
	Male	Female	Unknown
Sexually transmitted diseases			
Cochrane Library			x
Japan Society of Obstetrics and Gynecology		x	
American Society for Reproductive Medicine	x	x	
Ministry of Health, Labour, and Welfare	x	x	
Psychological stress			
Japan Society of Obstetrics and Gynecology	x		
American Society for Reproductive Medicine		x	
Ministry of Health, Labour, and Welfare	x		
Cigarette smoking			
Cochrane Library	x	x	
Japan Society of Obstetrics and Gynecology		x	
British Fertility Society			x
American Society for Reproductive Medicine	x	x	
Alcohol use			
Cochrane Library	x	x	
British Fertility Society			x
American Society for Reproductive Medicine			x
Nutrition and diet			
Cochrane Library	x	x	
Physical activity and exercise			
Cochrane Library	x	x	
Underweight			
Cochrane Library	x	x	
Japan Society of Obstetrics and Gynecology		x	
Ministry of Health, Labour, and Welfare		x	
Overweight and obesity			
Cochrane Library	x	x	
Japan Society of Obstetrics and Gynecology		x	
British Fertility Society			x
American Society for Reproductive Medicine		x	
Ministry of Health, Labour, and Welfare	x	x	
Environmental pollutants			
Cochrane Library	x	x	
American Society for Reproductive Medicine			x
Ministry of Health, Labour, and Welfare	x	x	

We created coding rules for the selected lifestyle factors. These coding guidelines are summarized in Table 2. Our coding included expressions directly related to the selected lifestyle factors associated with reduced fertility. However, expressions exclusively concerning the influence of these factors on fetuses or babies were excluded.

Table 2. Coding guidelines.

Lifestyle factor	Description
Sexually transmitted diseases	Content directly related to sexually transmitted diseases, including *Chlamydia trachomatis* and gonorrhea, is included. Additionally, content related to sexually transmitted disease screening or examination for causes of infertility is included.
Psychological stress	Content related to psychological stress in daily life or occupational life is included. Additionally, content related to psychological stress caused by infertility is included.
Cigarette smoking	Content directly related to cigarette smoking is included. Additionally, expressions concerning quitting smoking are included.
Alcohol use	Content directly related to alcohol is included. Additionally, expressions concerning alcohol drinking are included.
Nutrition and diet	Content directly related to nutrition and diet is included. Additionally, expressions concerning nutrient factors are included. Content only related to nutrient factors for the purpose of marketing (eg, information about supplement only) is excluded.
Physical activity and exercise	Content directly related to physical activity and exercise is included. Additionally, expressions concerning obesity prevention and exercise in regular life are included.
Underweight	Content directly related to underweight is included. Additionally, expressions concerning precipitous weight loss, dieting, and underweight as indicated by body mass index are included.
Overweight and obesity	Content directly related to overweight and obesity is included. Additionally, expressions concerning precipitous weight gain, and overweight and obesity as indicated by body mass index are included.
Environmental pollutants	Content directly related to environmental pollutants is included. Additionally, expressions concerning environmental hormones are included.

We analyzed the textual data on the websites retrieved using the search terms described above. First, we read the text carefully. Second, all data were coded on the selected lifestyle factors, author's professional expertise, and sex of the target audience (ie, information for men, women, or both). Finally, data from all webpages were assembled and pooled in Microsoft Excel. When information on lifestyle factors was provided, the code of 1 was assigned. When no information on lifestyle factors was provided, we assigned the code of 0. The URL and title of each webpage were saved as a reference during the data analysis. Because multiple types of lifestyle factors could be listed on a single website, instead of calculating the number of webpages mentioning a particular lifestyle factor, we calculated the number of mentions (codes) for each selected lifestyle factor associated with reduced fertility.

Interrater Reliability

Approximately 20% of the final dataset (100/519, 19.3%) was evaluated by two independent, blinded raters (RY and EF) to examine interrater reliability. Using the coding guidelines created by the first author of this study (RY), EF was instructed on applying the coding system in a training session that lasted about 1 hour. In a pilot test phase, the two raters applied the coding system to 10 webpages randomly selected from the full sample. No problems were identified during this phase. After a formal reliability assessment phase was completed, the first author (RY) calculated the interrater reliability index.

Statistical Analysis

To assess interrater reliability, the Gwet agreement coefficient (AC1) statistic, which is less affected by prevalence compared with the Cohen kappa [54], was used to assess interrater agreement of the coding. We also conducted a test to assess differences in frequency between the two most common types of professional expertise. In addition, we compared differences between male-specific and female-specific information on lifestyle factors associated with reduced fertility by categorizing the information into two groups: (1) information for men plus information for both men and women and (2) information for women plus information for both men and women. Lifestyle factors that could not be classified in this way were treated as missing in the above tests. Differences in the author's professional expertise, lifestyle factors, and sex of the target audience were assessed using count data analyzed by using the chi-square test and Fisher exact test. Statistical significance was set at $P<.05$ for all comparisons. Analyses were performed using R for Windows version 3.5.1 (R Foundation for Statistical Computing).

Ethical Considerations

This study was granted an exemption from the requirement of ethics approval by the ethical review committee at the Graduate School of Medicine, University of Tokyo, because we aimed to analyze online information, meaning that this study was not medical research involving human subjects, and because the authors had no conflicts of interest related to this study.

Results

Distributions of Author's Professional Expertise

The assignment of all codes ranged from 0 to 9 (mean 1.017) per page. The assignment of codes ranged from 0 to 8 (mean 1.058) for hospitals and clinics, from 0 to 8 (mean 0.974) for media organizations, from 0 to 9 (mean 1.147) for public administration institutions, from 0 to 6 (mean 1.029) for business and service entities, from 0 to 6 (mean 1.389) for academic and medical societies, from 0 to 7 (mean 2.105) for pharmaceutical and medical device firms, from 0 to 3 (mean 0.367) for alternative medicine practitioners, and from 0 to 3 (mean 0.391) for laypersons. Of the webpages retrieved, 0 codes were assigned

to 60.7% (315/519). The number of webpages for which 0 codes were assigned was 53.4% (110/206) for hospitals and clinics, 66.2% (102/154) for media organizations, 68% (23/34) for public administration institutions, 60% (21/35) for business and service entities, 50% (9/18) for academic and medical societies, 53% (10/19) for pharmaceutical and medical device firms, 77% (23/30) for alternative medicine practitioners, and 74% (17/23) for laypersons.

The distributions of the webpage author's professional expertise are summarized in Multimedia Appendix 1. Of the webpages retrieved, 39.7% (206/519) were produced by hospitals and clinics, 29.7% (154/519) by media organizations, 6.6% (34/519) by public administration institutions, 6.7% (35/519) by business and service entities, 5.8% (30/519) by alternative medicine practitioners, 4.4% (23/519) by laypersons, 3.7% (19/519) by pharmaceutical or medical device firms, and 3.5% (18/519) by academic and medical societies. Of the codes assigned for the examined lifestyle factors associated with reduced fertility, 41.2% (218/528) were published by hospitals and clinics, 28.4% (150/528) by media organizations, 7.6% (40/528) by pharmaceutical and medical device firms, 7.4% (39/528) by public administration institutions, 6.8% (36/528) by business and service entities, 4.7% (25/528) by academic and medical societies, 2.1% (11/528) by alternative medicine practitioners, and 1.7% (9/528) by laypersons.

Interrater Reliability

A review of the coding by the two independent, blinded raters on the randomly selected subsample of 100 webpages revealed that the interrater agreement was excellent. The two raters agreed on the assignment of codes in 98.00% (3528/3600) of coding instances. The Gwet AC1 statistic for the assignment of codes ranged from 0.969 to 1.000 (mean AC1 0.985) for the author's professional expertise, from 0.957 to 0.992 (mean AC1 0.979) for lifestyle factors, and from 0.950 to 0.986 (mean AC1 0.978) for the sex of the target audience. For the data analysis in this study, we used the coding of the first author.

Distribution of Lifestyle Factors

Of the coding instances concerning lifestyle factors associated with reduced fertility, 20.1% (106/528) related to psychological stress and 18.8% (99/528) related to sexually transmitted diseases. Together, codes on these two lifestyle factors accounted for 38.9% (205/528) of all lifestyle factor-coding instances. Websites referring to the associations between reduced fertility and nutrition and diet, cigarette smoking, underweight, physical activity and exercise, overweight and obesity, alcohol use, and environmental pollutants were relatively rare.

Distribution of Lifestyle Factors by Author's Professional Expertise

Figure 3 illustrates the distribution of codes for the examined lifestyle factors associated with reduced fertility by author's professional expertise. Of the coding instances referring to the association between reduced fertility and sexually transmitted diseases, 67% (66/99) were published by hospital or clinics and 14% (14/99) were published by media organizations. We assessed differences in the frequency of each lifestyle factor code between the two most common types of author's professional expertise (ie, hospitals/clinics and media organizations) using the chi-square test and Fisher exact test (Table 3). The results of these tests showed that the association between sexually transmitted diseases and reduced fertility was mentioned more frequently by hospitals and clinics than by media organizations (P<.001). In contrast, of the coding instances referring to the association between nutrition and diet and reduced fertility, 26% (18/68) were published by hospitals and clinics and 40% (27/68) were published by media organizations; the association between nutrition and diet and reduced fertility was significantly more frequently mentioned by the media than by hospitals and clinics (P=.008). The other lifestyle factors associated with reduced fertility did not show statistically significant differences between hospitals/clinics and media organizations.

Figure 3. Distribution of codes for lifestyle factors associated with reduced fertility by the author's professional expertise.

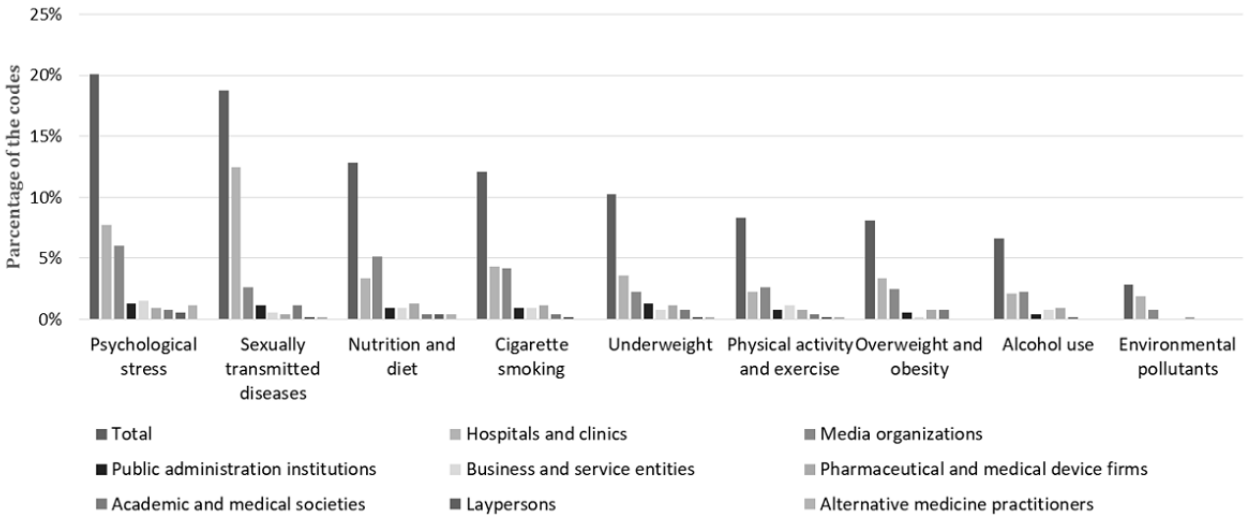


Table 3. Lifestyle factors and their associations with webpage author's professional expertise for hospitals or clinics and media organizations.

Lifestyle factor	Author's professional expertise, n (%)		Chi-square	*df*[a]	*P* value[b]
	Hospitals and clinics (n=218)	Media organizations (n=150)			
Sexually transmitted diseases			21.7	1	<.001[c]
Number of webpages with codes	66 (30.3)	14 (9.3)			
Number of webpages without codes	152 (69.7)	136 (90.7)			
Psychological stress			0.2	1	.64[c]
Number of webpages with codes	41 (18.8)	32 (21.3)			
Number of webpages without codes	177 (81.2)	118 (78.7)			
Cigarette smoking			1	1	.31[c]
Number of webpages with codes	23 (10.6)	22 (14.7)			
Number of webpages without codes	195 (89.4)	128 (85.3)			
Alcohol use			0.9	1	.35[c]
Number of webpages with codes	11 (5.0)	12 (8.0)			
Number of webpages without codes	207 (95.0)	138 (92.0)			
Nutrition and diet			7	1	.008[c]
Number of webpages with codes	18 (8.3)	27 (18.0)			
Number of webpages without codes	200 (91.7)	123 (82.0)			
Physical activity and exercise			1.4	1	.23[c]
Number of webpages with codes	12 (5.5)	14 (9.3)			
Number of webpages without codes	206 (94.5)	136 (90.7)			
Underweight			0	1	.96[c]
Number of webpages with codes	19 (8.7)	12 (8.0)			
Number of webpages without codes	199 (91.3)	138 (92.0)			
Overweight and obesity			0	1	>.99[c]
Number of webpages with codes	18 (8.3)	13 (8.7)			
Number of webpages without codes	200 (91.7)	137 (91.3)			
Environmental pollutants			—	—	.42[d]
Number of webpages with codes	10 (4.6)	4 (2.7)			
Number of webpages without codes	208 (95.4)	146 (97.3)			

[a]Degrees of freedom.

[b]*P* values compare media organizations with hospitals and clinics.

[c]Chi-square test.

[d]Fisher exact test.

Distribution of Lifestyle Factors by the Sex of the Target Audience

Figure 4 illustrates the distribution of codes for lifestyle factors associated with reduced fertility by the sex of the target audience. Of the total lifestyle factor-coding instances, 12.9% (68/528) were about men's lifestyle factors, 60.2% (318/528) were about women's lifestyle factors, and 17.4% (92/528) were about both women's and men's lifestyle factors (see Multimedia Appendix 2). Across all examined types of lifestyle factors, female-specific information was observed more frequently than was male-specific information. Information referring to psychological stress, sexually transmitted diseases, nutrition and diet, underweight, physical activity and exercise, and overweight and obesity was significantly more frequently directed toward women than toward men (Table 4; psychological stress: *P*=.002, sexually transmitted diseases: *P*<.001, nutrition

and diet: *P*<.001, underweight: *P*<.001, physical activity and exercise: *P*<.001, overweight and obesity: *P*<.001).

Figure 4. Distribution of codes for lifestyle factors associated with reduced fertility by the sex of the target audience.

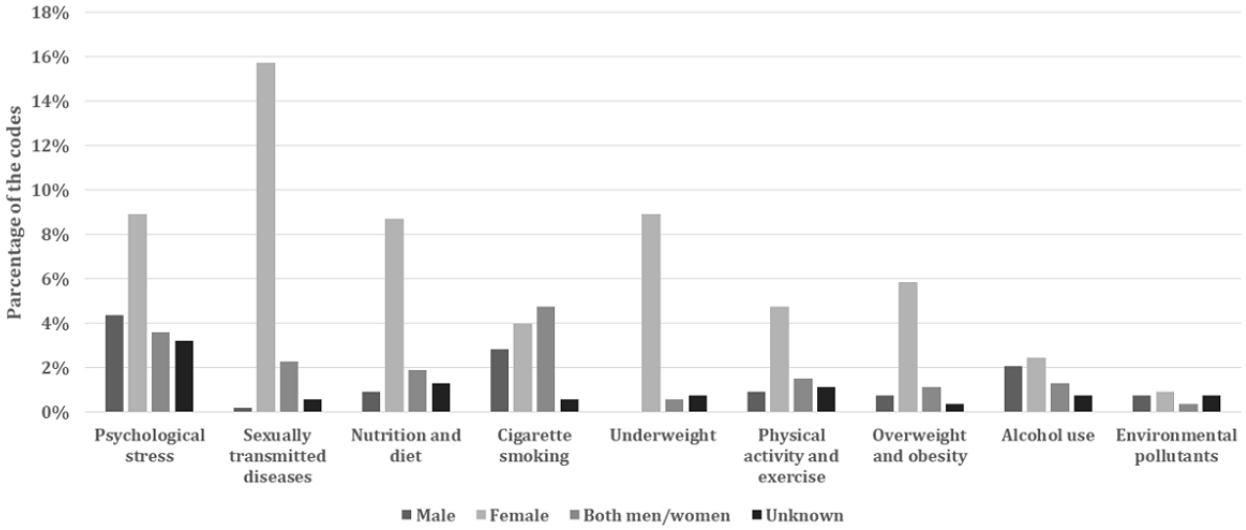


Table 4. Lifestyle factors and their associations with the sex of the target audience.

Lifestyle factor	Sex of target audience		Chi-square	*df*[b]	*P* value[c]
	Male[a], n (%)	Female[a], n (%)			
Sexually transmitted diseases (n=108)			121.5	1	<.001[d]
Number of webpages with codes	13 (12.0)	95 (88.0)			
Number of webpages without codes	95 (88.0)	13 (12.0)			
Psychological stress (n=108)			9.8	1	.002[d]
Number of webpages with codes	42 (38.9)	66 (61.1)			
Number of webpages without codes	66 (61.1)	42 (38.9)			
Cigarette smoking (n=86)			0.6	1	.45[d]
Number of webpages with codes	40 (46.5)	46 (53.5)			
Number of webpages without codes	46 (53.5)	40 (46.5)			
Alcohol use (n=38)			0.1	1	.82[d]
Number of webpages with codes	18 (47.4)	20 (52.6)			
Number of webpages without codes	20 (52.6)	18 (47.4)			
Nutrition and diet (n=71)			45.1	1	<.001[d]
Number of webpages with codes	15 (21.1)	56 (78.9)			
Number of webpages without codes	56 (78.9)	15 (21.1)			
Physical activity and exercise (n=46)			15.7	1	<.001[d]
Number of webpages with codes	13 (28.3)	33 (71.7)			
Number of webpages without codes	33 (71.7)	13 (28.3)			
Underweight (n=53)			—	—	<.001[e]
Number of webpages with codes	3 (5.7)	50 (94.3)			
Number of webpages without codes	50 (94.3)	3 (5.7)			
Overweight and obesity (n=47)			28.8	1	<.001[d]
Number of webpages with codes	10 (21.3)	37 (78.7)			
Number of webpages without codes	37 (78.7)	10 (21.3)			
Environmental pollutants (n=13)			—	—	>.99[e]
Number of webpages with codes	6 (46.2)	7 (53.8)			
Number of webpages without codes	7 (53.8)	6 (46.2)			

[a]We categorized the information into two groups: (1) information for men plus information for both men and women and (2) information for women plus information for both men and women. Lifestyle factors that could not be classified in this way were treated as missing.

[b]Degrees of freedom.

[c]*P* values compare the information for men with the information for women.

[d]Chi-square test.

[e]Fisher exact test.

Discussion

Principal Findings

This study quantitatively examined online information on the lifestyle factors associated with reduced fertility. The main findings of the study confirm that a large proportion of the information on lifestyle factors associated with reduced fertility was disseminated by hospitals, clinics, and the media. Furthermore, the findings show that the frequencies of descriptions provided by different entities varied by the particular lifestyle factor examined.

Distribution of Author's Professional Expertise

In terms of the professional expertise of the author publishing online information on lifestyle factors associated with reduced fertility, following the categorization of professional expertise described in the Methods section, our study demonstrated that, of the codes assigned for the examined lifestyle factors associated with reduced fertility, 41.2% (218/528) were published by hospitals and clinics, 28.4% (150/528) by media organizations, and 1.7% (9/528) by laypersons. This may be because of the strategy of search engine optimization used by hospitals and clinics to make their websites appear higher in user lists of search results generated by the search engines. This may also be because lifestyle factors associated with reduced fertility are discussed in the context of infertility examinations or treatments. The large number of references to lifestyle factors associated with reduced fertility by media organizations may be associated with the mass media coverage of the public debate [55] after the Japanese government's proposal for the need for fertility education in 2013 [28]. Conversely, the low number of references to lifestyle factors associated with reduced fertility by laypersons may be associated with perceived stigma around infertility [56]: Laypersons may avoid discussing these factors because of concerns about stigma related to infertility or not being able to get pregnant. In contexts where the desire for children is generally regarded as the social norm, higher levels of stigma consciousness may be associated with reductions in disclosure of one's own infertility [56]. Those who disclose their infertility may also have more negative social experiences associated with their infertility in such contexts [56] and consequently avoid further disclosing their infertility in the future. Therefore, the stigma around infertility may explain why laypersons were found to rarely mention lifestyle factors associated with reduced fertility in this study.

Distribution of Lifestyle Factors

Psychological stress and sexually transmitted diseases were the two most frequently mentioned lifestyle factors associated with reduced fertility. In contrast, the associations between reduced fertility and cigarette smoking, alcohol use, physical activity and exercise, underweight, and overweight and obesity were less frequently discussed. Our findings are consistent with those of a previous study showing that people lack a general understanding of fertility, including the associations between reduced fertility and cigarette smoking and overweight and obesity [28]. In 2017, the National Health and Nutrition Survey demonstrated that (1) 40% to 50% of the reproductive-aged population in Japan drank alcohol [57], (2) more than one-quarter of reproductive-aged men were obese [57], (3) 10% to 20% of reproductive-aged women were underweight [57], and (4) approximately 20% of the population smoked cigarettes [57]. Despite this situation, the associations between reduced fertility and alcohol consumption of both men and women, male overweight and obesity, female underweight, and cigarette smoking of both men and women were mentioned relatively infrequently on the websites analyzed in this study. These lifestyle factors should be discussed more frequently. We also believe that public initiatives for the reproductive-aged population and a fertility-related information strategy should be established in Japan. Several exemplary websites with evidence-based information exist that could serve as a model for these changes [58,59].

Distribution of Lifestyle Factors by Author's Professional Expertise

Our research has revealed that associations between reduced fertility and both psychological stress and sexually transmitted diseases were more frequently discussed by hospitals and clinics than by the other types of webpage author. This may be because lifestyle factors associated with reduced fertility are discussed in the context of infertility distress regarding infertility examinations or treatments [32,60]. We also found that the association between reduced fertility and nutrition and diet was more frequently discussed by the media than by hospitals and clinics. This may be because lifestyle factors associated with reduced fertility are presented less as health-oriented information, with media outlets orienting health messages toward entertainment.

Providing treatment is the main role of hospitals and clinics, and these entities may use their websites as a tool for patient acquisition. However, for individuals seeking fertility information on the internet because they are trying to conceive, the content of hospital and clinic websites may be a major source of information on fertility. This idea is supported by the finding that hospital and clinic websites rank higher in lists of search results when users use web search engines. Hence, various kinds of information provided by hospitals and clinics may be beneficial for people who are trying to conceive. In previous work, scholars have recommended that, despite the pressures of a competitive environment, hospitals and clinics offering fertility treatment should present educational information in an ethically balanced manner [61]. The same scholars have also recommended that medical experts lead the way for best practices among doctors by creating guidelines regarding the provision of online information [61]. We additionally recommend the ethically balanced presentation of information on fertility in the media. Media websites should incorporate benchmarks for collaboration between experts and the media.

Distribution of Lifestyle Factors by the Sex of the Target Audience

We found that there were fewer mentions of lifestyle factors associated with reduced fertility for men than for women. This may be because (1) men are less likely to seek information on infertility [32], (2) male and female infertility is highly stigmatized, and (3) men have a relatively low level of knowledge regarding male infertility [62]. Regarding stigma around male infertility, reproductive health is often regarded as a women's issue [31]. This may be because fertility treatment has mainly focused on women's bodies, although male factors also contribute to infertility [31]. Likewise, previous work indicates that infertility tends to be regarded as a women's issue in Japan [63]. Message senders may be influenced by this social norm, and they may therefore convey messages regarding infertility in a way that is consistent with this topic being a women's issue. Conversely, male infertility may be related to stigma stemming from ideas about masculinity that many men consider to be a social norm [60,64]. Attempts to hide stigmatized conditions often lead to delays in

information-seeking behavior [65]. Therefore, it is important that published discussions on the lifestyle factors associated with reduced fertility include information for both men and women.

Infertility and Stigma

As mentioned above, infertility may harm the self-esteem of people who are trying to conceive because of its latently stigmatizing nature [56]. People with a stigmatized condition tend to use the internet to seek health information more often compared with people with nonstigmatized conditions [65]. Those who experience infertility may feel a sense of isolation and be less likely to seek social support because of reduced self-esteem [56]. Advantages of the internet include user anonymity, optional disclosure, and the lack of geographical barriers [66]. These advantages could mitigate the threat of social stigma and distress around disclosure, providing much needed information on fertility. It is also important to address online discussion boards in terms of informational, emotional, and appraisal support [66]. Moreover, because the internet may be a good public education tool for people with stigmatized conditions [65], we recommend the use of online media for fertility-related education [67].

Limitations

This study has several limitations. First, we limited our examination to online information, which does not capture all circulating public messages. Television broadcasts and print newspapers and magazines may also be widely used sources of information about lifestyle factors associated with reduced fertility. Second, although a substantial number of websites (n=519) were retrieved for this analysis, availability, accessibility, and time limitations made it unfeasible to analyze all relevant sites. Third, despite the fact that we selected search terms based on related words indicated by the selected search engines, these search terms may have reflected our own biases. For example, it is possible that search terms were relatively easy for women to use. However, because men were less likely to seek information regarding infertility [32], the webpages for women should include information for men. Fourth, it is possible that the selected lifestyle factors associated with reduced fertility may also have reflected our biases; however, this study examined several publications from academic societies, Cochrane Library, and previous studies to determine which lifestyle factors to investigate. For example, age was not treated as a lifestyle factor associated with reduced fertility in this study because Cochrane Library did not directly consider age when defining lifestyle factors that may both influence fertility and affect the chances of a healthy, live birth [5]. Additionally, a recent study found that most of the Japanese reproductive-aged population (60% to 70%) were knowledgeable about the association between age and reduced fertility [28]. Therefore, this study did not include age in the analysis. However, age should be explored as a potential factor associated with reduced fertility in future research. Fifth, when identifying lifestyle factors associated with reduced fertility, we checked for critical threshold levels for the contribution of particular factors, but study designs, outcomes, and sample sizes varied across the examined studies. Because there are presently no guidelines concerning lifestyle factors associated with reduced fertility in any country, we could not identify unified critical thresholds for judging the impact of each lifestyle factor on reduced fertility. Sixth, we did not evaluate the accuracy of the information presented on the webpages. It is possible that some websites provided accurate information and others provided incorrect information. The accuracy of information provided on this topic should be explored in future research. Seventh, our study was conducted via online search at the end of October 2019. Although we confirmed that the number of searches for the top 10 most frequent main search terms did not represent a surge, we did not explore how the number and frequency of search results concerning the examined lifestyle factors associated with reduced fertility changed over the time. Finally, our analysis was restricted to Japanese-language online information, which may limit its generalizability to other contexts. However, considering context is crucial when examining media messages, which is an advantage of focusing the analysis on Japanese-language websites [35]. Although there are limitations to the study, to our knowledge, this study is the first to examine online information on the lifestyle factors associated with reduced fertility using quantitative content analysis.

Conclusions

In terms of the webpage author's professional expertise, authors from hospitals, clinics, and the media relatively frequently mentioned lifestyle factors associated with reduced fertility, whereas laypersons mentioned them relatively rarely. Regarding the specific lifestyle factors associated with reduced fertility, psychological stress and sexually transmitted diseases were more frequently discussed compared with the other factors. The association between reduced fertility and sexually transmitted diseases was more frequently discussed by hospitals and clinics than by the media. Conversely, the association between reduced fertility and nutrition and diet was significantly more frequently mentioned by the media than by hospitals and clinics. With regard to the sex of the target audience, male lifestyle factors were less frequently discussed than were female lifestyle factors. The authors of fertility-related websites should more frequently mention information on lifestyle factors associated with reduced fertility overall, moving beyond only psychological stress and sexually transmitted diseases to also discuss nutrition and diet, cigarette smoking, underweight, physical activity and exercise, overweight and obesity, alcohol use, and environmental pollutants. These authors should also make an effort to provide specific information on men's lifestyle factors.

Acknowledgments

RY contributed to the study design, data collection, data analysis, interpretation, and manuscript drafting. RY and TO conceived and designed the study. TO contributed to the critical revision and editing of this manuscript. HO and HU provided feedback on

the manuscript. EF performed the analysis and interpretation of the data. TK supervised the study. All authors read and approved the final manuscript for publication. The authors would like to thank Ritsuko Shirabe and Saeko Higuchi for their assistance in identifying lifestyle factors associated with reduced fertility. We also thank Jennifer Barrett, PhD, from Edanz Group for editing a draft of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution of lifestyle factor codes by webpage author's professional expertise.
[DOCX File , 15 KB - jmir_v22i8e19777_app1.docx]

Multimedia Appendix 2
Distribution of lifestyle factor codes by the author's professional expertise and the sex of the target audience.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v22i8e19777_app2.xlsx]

References

1. Health topics: reproductive health. World Health Organization. URL: https://www.who.int/westernpacific/health-topics/reproductive-health [accessed 2019-11-17]
2. World HO. Infertility: a tabulation of available data on prevalence of primary and secondary infertility. Programme on Maternal and Child Health and Family Planning, Division of Family Health. Geneva: World Health Organization; 1991. URL: https://apps.who.int/iris/bitstream/handle/10665/59769/WHO_MCH_91.9.pdf?sequence=1&isAllowed=y [accessed 2019-11-17]
3. Bunting L, Tsibulsky I, Boivin J. Fertility knowledge and beliefs about fertility treatment: findings from the International Fertility Decision-making Study. Hum Reprod 2013 Feb;28(2):385-397 [FREE Full text] [doi: 10.1093/humrep/des402] [Medline: 23184181]
4. Daumler D, Chan P, Lo KC, Takefman J, Zelkowitz P. Men's knowledge of their own fertility: a population-based survey examining the awareness of factors that are associated with male infertility. Hum Reprod 2016 Dec;31(12):2781-2790 [FREE Full text] [doi: 10.1093/humrep/dew265] [Medline: 27816924]
5. Anderson K, Norman RJ, Middleton P. Preconception lifestyle advice for people with subfertility. Cochrane Database Syst Rev 2010 Apr 14(4):CD008189. [doi: 10.1002/14651858.CD008189.pub2] [Medline: 20393968]
6. Inhorn MC, Patrizio P. Infertility around the globe: new thinking on gender, reproductive technologies and global movements in the 21st century. Hum Reprod Update 2015;21(4):411-426 [FREE Full text] [doi: 10.1093/humupd/dmv016] [Medline: 25801630]
7. McLaren JF. Infertility evaluation. Obstet Gynecol Clin North Am 2012 Dec;39(4):453-463. [doi: 10.1016/j.ogc.2012.09.001] [Medline: 23182553]
8. Basic survey on social security and population problems. National Institute of Population and Social Security Research. 2015. URL: http://www.ipss.go.jp/ps-doukou/j/doukou15/NFS15_gaiyou.pdf [accessed 2019-11-17]
9. Bunting L, Boivin J. Knowledge about infertility risk factors, fertility myths and illusory benefits of healthy habits in young people. Hum Reprod 2008 Aug;23(8):1858-1864. [doi: 10.1093/humrep/den168] [Medline: 18469334]
10. Tang W, Mao J, Li KT, Walker JS, Chou R, Fu R, et al. Pregnancy and fertility-related adverse outcomes associated with Chlamydia trachomatis infection: a global systematic review and meta-analysis. Sex Transm Infect 2020 Aug;96(5):322-329. [doi: 10.1136/sextrans-2019-053999] [Medline: 31836678]
11. Xiong Y, Chen Y, Cheng M, He W, Chen Q. The risk of human papillomavirus infection for male fertility abnormality: a meta-analysis. Asian J Androl 2018 Oct;20(5):493-497 [FREE Full text] [doi: 10.4103/aja.aja_77_17] [Medline: 29623908]
12. Li Y, Lin H, Li Y, Cao J. Association between socio-psycho-behavioral factors and male semen quality: systematic review and meta-analyses. Fertil Steril 2011 Jan;95(1):116-123. [doi: 10.1016/j.fertnstert.2010.06.031] [Medline: 20674912]
13. Veltman-Verhulst SM, Boivin J, Eijkemans MJC, Fauser BJCM. Emotional distress is a common risk in women with polycystic ovary syndrome: a systematic review and meta-analysis of 28 studies. Hum Reprod Update 2012;18(6):638-651. [doi: 10.1093/humupd/dms029] [Medline: 22824735]
14. Augood C, Duckitt K, Templeton AA. Smoking and female infertility: a systematic review and meta-analysis. Hum Reprod 1998 Jun;13(6):1532-1539. [doi: 10.1093/humrep/13.6.1532] [Medline: 9688387]
15. Bundhun PK, Janoo G, Bhurtu A, Teeluck AR, Soogund MZS, Pursun M, et al. Tobacco smoking and semen quality in infertile males: a systematic review and meta-analysis. BMC Public Health 2019 Jan 08;19(1):36 [FREE Full text] [doi: 10.1186/s12889-018-6319-3] [Medline: 30621647]
16. Ricci E, Al Beitawi S, Cipriani S, Candiani M, Chiaffarino F, Viganò P, et al. Semen quality and alcohol intake: a systematic review and meta-analysis. Reprod Biomed Online 2017 Jan;34(1):38-47. [doi: 10.1016/j.rbmo.2016.09.012] [Medline: 28029592]

17. Fan D, Liu L, Xia Q, Wang W, Wu S, Tian G, et al. Female alcohol consumption and fecundability: a systematic review and dose-response meta-analysis. Sci Rep 2017 Oct 23;7(1):13815 [FREE Full text] [doi: 10.1038/s41598-017-14261-8] [Medline: 29062133]
18. Chavarro JE, Rich-Edwards JW, Rosner BA, Willett WC. Protein intake and ovulatory infertility. Am J Obstet Gynecol 2008 Feb;198(2):210-217 [FREE Full text] [doi: 10.1016/j.ajog.2007.06.057] [Medline: 18226626]
19. Salas-Huetos A, Bulló M, Salas-Salvadó J. Dietary patterns, foods and nutrients in male fertility parameters and fecundability: a systematic review of observational studies. Hum Reprod Update 2017 Jul 01;23(4):371-389. [doi: 10.1093/humupd/dmx006] [Medline: 28333357]
20. Arab A, Rafie N, Mansourian M, Miraghajani M, Hajianfar H. Dietary patterns and semen quality: a systematic review and meta-analysis of observational studies. Andrology 2018 Jan;6(1):20-28 [FREE Full text] [doi: 10.1111/andr.12430] [Medline: 29024507]
21. Liu Y, Zhang W. Association between body mass index and endometriosis risk: a meta-analysis. Oncotarget 2017 Jul 18;8(29):46928-46936 [FREE Full text] [doi: 10.18632/oncotarget.14916] [Medline: 28159926]
22. Guo D, Xu M, Zhou Q, Wu C, Ju R, Dai J. Is low body mass index a risk factor for semen quality? A PRISMA-compliant meta-analysis. Medicine (Baltimore) 2019 Aug;98(32):e16677 [FREE Full text] [doi: 10.1097/MD.0000000000016677] [Medline: 31393367]
23. Sermondade N, Faure C, Fezeu L, Shayeb AG, Bonde JP, Jensen TK, et al. BMI in relation to sperm count: an updated systematic review and collaborative meta-analysis. Hum Reprod Update 2013;19(3):221-231 [FREE Full text] [doi: 10.1093/humupd/dms050] [Medline: 23242914]
24. Lan L, Harrison CL, Misso M, Hill B, Teede HJ, Mol BW, et al. Systematic review and meta-analysis of the impact of preconception lifestyle interventions on fertility, obstetric, fetal, anthropometric and metabolic outcomes in men and women. Hum Reprod 2017 Sep 01;32(9):1925-1940. [doi: 10.1093/humrep/dex241] [Medline: 28854715]
25. Best D, Avenell A, Bhattacharya S. How effective are weight-loss interventions for improving fertility in women and men who are overweight or obese? A systematic review and meta-analysis of the evidence. Hum Reprod Update 2017 Nov 01;23(6):681-705. [doi: 10.1093/humupd/dmx027] [Medline: 28961722]
26. Fathi Najafi T, Latifnejad Roudsari R, Namvar F, Ghavami Ghanbarabadi V, Hadizadeh Talasaz Z, Esmaeli M. Air pollution and quality of sperm: a meta-analysis. Iran Red Crescent Med J 2015 Apr;17(4):e26930 [FREE Full text] [doi: 10.5812/ircmj.17(4)2015.26930] [Medline: 26023349]
27. Sharma R, Biedenharn KR, Fedor JM, Agarwal A. Lifestyle factors and reproductive health: taking control of your fertility. Reprod Biol Endocrinol 2013 Jul 16;11:66 [FREE Full text] [doi: 10.1186/1477-7827-11-66] [Medline: 23870423]
28. Maeda E, Sugimori H, Nakamura F, Kobayashi Y, Green J, Suka M, et al. A cross sectional study on fertility knowledge in Japan, measured with the Japanese version of Cardiff Fertility Knowledge Scale (CFKS-J). Reprod Health 2015 Jan 31;12:10 [FREE Full text] [doi: 10.1186/1742-4755-12-10] [Medline: 25638172]
29. Pedro J, Brandão T, Schmidt L, Costa ME, Martins MV. What do people know about fertility? A systematic review on fertility awareness and its associated factors. Ups J Med Sci 2018 Jun;123(2):71-81 [FREE Full text] [doi: 10.1080/03009734.2018.1480186] [Medline: 29957086]
30. [Survey on infertility awareness]. Recruit Lifestyle Co. URL: https://www.recruit-lifestyle.co.jp/news/pressrelease/others/nw26561_20180730 [accessed 2019-11-17]
31. Hinton L, Miller T. Mapping men's anticipations and experiences in the reproductive realm: (in)fertility journeys. Reprod Biomed Online 2013 Sep;27(3):244-252. [doi: 10.1016/j.rbmo.2013.06.008] [Medline: 23871363]
32. Robins S, Barr HJ, Idelson R, Lambert S, Zelkowitz P. Online health information regarding male infertility: an evaluation of readability, suitability, and quality. Interact J Med Res 2016 Oct 21;5(4):e25 [FREE Full text] [doi: 10.2196/ijmr.6440] [Medline: 27769954]
33. Survey report on spending time of communication media and information behavior. Ministry of Internal Affairs and Communications. 2019. URL: http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000082.html [accessed 2019-11-17]
34. Morris DS, Rooney MP, Wray RJ, Kreuter MW. Measuring exposure to health messages in community-based intervention studies: a systematic review of current practices. Health Educ Behav 2009 Dec;36(6):979-998. [doi: 10.1177/1090198108330001] [Medline: 19605622]
35. King L, Tulandi T, Whitley R, Constantinescu T, Ells C, Zelkowitz P. What's the message? A content analysis of newspaper articles about assisted reproductive technology from 2005 to 2011. Hum Fertil (Camb) 2014 Jun;17(2):124-132. [doi: 10.3109/14647273.2014.895427] [Medline: 24669833]
36. Mohammadi L, Aranda D, Martínez-Martínez S. The narratives of fertility clinic's websites in Spain. El profesional de la información 2019;28(2):e280219 [FREE Full text] [doi: 10.3145/epi.2019.mar.19]
37. Okamura K, Bernstein J, Fidler AT. Assessing the quality of infertility resources on the World Wide Web: tools to guide clients through the maze of fact and fiction. J Midwifery Womens Health 2002;47(4):264-268. [doi: 10.1016/s1526-9523(02)00260-x] [Medline: 12138934]
38. Kelly-Hedrick M, Grunberg PH, Brochu F, Zelkowitz P. "It's totally okay to be sad, but never lose hope": content analysis of infertility-related videos on YouTube in relation to viewer preferences. J Med Internet Res 2018 May 23;20(5):e10199 [FREE Full text] [doi: 10.2196/10199] [Medline: 29792296]

39. Malik SH, Coulson NS. Coping with infertility online: an examination of self-help mechanisms in an online infertility support group. Patient Educ Couns 2010 Nov;81(2):315-318. [doi: 10.1016/j.pec.2010.01.007] [Medline: 20144521]
40. Osadchiy V, Mills JN, Eleswarapu SV. Understanding patient anxieties in the social media era: qualitative analysis and natural language processing of an online male infertility community. J Med Internet Res 2020 Mar 10;22(3):e16728 [FREE Full text] [doi: 10.2196/16728] [Medline: 32154785]
41. Google Japan. URL: https://www.google.co.jp/ [accessed 2019-11-17]
42. Yahoo! Japan. URL: https://www.yahoo.co.jp/ [accessed 2019-11-17]
43. Search engine market share Japan. StatCounter Global Stars. URL: https://gs.statcounter.com/search-engine-market-share/all/japan/#desktop+mobile+tablet+console_search_engine--bar [accessed 2019-11-17]
44. Calculated tool for number of website searches. Devo Inc. URL: http://aramakijake.jp/ [accessed 2019-10-20]
45. What are related searches? Yahoo Japan Corporation. URL: https://support.yahoo-net.jp/PccSearch/s/article/H000011628 [accessed 2019-10-20]
46. Let's check trends of website searches. Google Trends. URL: https://trends.google.co.jp/trends/?geo=JP [accessed 2019-11-17]
47. Ha L, James EL. Interactivity reexamined: a baseline analysis of early business web sites. J Broadc Electr Media 1998 Sep;42(4):457-474. [doi: 10.1080/08838159809364462]
48. Zhou X. E-government in China: a content analysis of national and provincial Web Sites. J Comput-Mediated Comm 2004;9(4):948. [doi: 10.1111/j.1083-6101.2004.tb00297.x]
49. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 09;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
50. Infertility. Japan Society of Obstetrics and Gynecology. URL: http://www.jsog.or.jp/modules/diseases/index.php?content_id=15 [accessed 2020-08-08]
51. What do you know about male body and female body? Basic knowledge for spending fulfilling life. Ministry of Health, Labour, and Welfare. URL: https://www.mhlw.go.jp/seisakunitsuite/bunya/kodomo/kodomo_kosodate/boshi-hoken/dl/gyousei-01-01.pdf#search='%E5%A6%8A%E5%A8%A0%E5%87%BA%E6%9D%A5%E3%81%AA%E3%81%84 [accessed 2019-11-17]
52. What are the main preventable cases of infertility? British Fertility Society. URL: https://www.britishfertilitysociety.org.uk/fei/what-are-the-main-preventable-causes-of-infertility/ [accessed 2019-11-17]
53. Protecting your fertility. American Society for Reproductive Medicine. URL: https://www.asrm.org/topics/topics-index/?filterbycategoryid=109 [accessed 2019-11-17]
54. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol 2013;13:61 [FREE Full text] [doi: 10.1186/1471-2288-13-61] [Medline: 23627889]
55. The proposal for breaking though the fertility crisis. The Task Force for Tackling the Crisis of Falling Birth Rate. 2013. URL: https://www8.cao.go.jp/shoushi/shoushika/meeting/taskforce/k_4/pdf/teian.pdf [accessed 2020-06-07]
56. Slade P, O'Neill C, Simpson AJ, Lashen H. The relationship between perceived stigma, disclosure patterns, support and distress in new attendees at an infertility clinic. Hum Reprod 2007 Aug;22(8):2309-2317. [doi: 10.1093/humrep/dem115] [Medline: 17580298]
57. National Health and Nutrition Survey 2017. Ministry of Health, Labour, and Welfare. URL: https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450171&kikan=00450&stat_infid=000031041748&cycle=7&tclass1=000001123588&survey=%E5%81%A5%E5%BA%B7&result_page=1&cycle_facet=cycle [accessed 2019-11-17]
58. Delbaere I, Mokangi P, Roelens K, De Sutter A, Gellynck X, Beeckman D, et al. Systematic development of an evidence-based website on preconception care. Ups J Med Sci 2016 Nov;121(4):264-270 [FREE Full text] [doi: 10.1080/03009734.2016.1216481] [Medline: 27586661]
59. Before pregnancy. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/preconception/index.html [accessed 2020-06-05]
60. Dooley M, Nolan A, Sarma K. The psychological impact of male factor infertility and fertility treatment on men: a qualitative study. Irish J Psychol 2011 Mar;32(1-2):14-24. [doi: 10.1080/03033910.2011.611253]
61. Wong L, Yan H, Margel D, Fleshner NE. Urologists in cyberspace: a review of the quality of health information from American urologists' websites using three validated tools. Can Urol Assoc J 2013;7(3-4):100-107 [FREE Full text] [doi: 10.5489/cuaj.501] [Medline: 23671523]
62. Daniluk JC, Koert E. The other side of the fertility coin: a comparison of childless men's and women's knowledge of fertility and assisted reproductive technology. Fertil Steril 2013 Mar 01;99(3):839-846. [doi: 10.1016/j.fertnstert.2012.10.033] [Medline: 23148926]
63. Yui H. Visualization of male infertility and maternity protection concept. Ann Rep Fam Res 2015;40:7-23 [FREE Full text] [doi: 10.14965/afs.40.7]
64. Dooley M, Dineen T, Sarma K, Nolan A. The psychological impact of infertility and fertility treatment on the male partner. Hum Fertil (Camb) 2014 Sep;17(3):203-209. [doi: 10.3109/14647273.2014.942390] [Medline: 25116275]

65. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
66. Richard J, Badillo-Amberg I, Zelkowitz P. "So much of this story could be me": men's use of support in online infertility discussion boards. Am J Mens Health 2017 Dec;11(3):663-673 [FREE Full text] [doi: 10.1177/1557988316671460] [Medline: 27702886]
67. Maeda E, Nakamura F, Kobayashi Y, Boivin J, Sugimori H, Murata K, et al. Effects of fertility education on knowledge, desires and anxiety among the reproductive-aged population: findings from a randomized controlled trial. Hum Reprod 2016 Sep;31(9):2051-2060 [FREE Full text] [doi: 10.1093/humrep/dew133] [Medline: 27301362]

Abbreviations

AC1: Gwet agreement coefficient 1

Edited by G Eysenbach; submitted 02.05.20; peer-reviewed by E Maeda, M Kelly-Hedrick; comments to author 01.06.20; revised version received 23.06.20; accepted 07.07.20; published 25.08.20.

Please cite as:
Yokota R, Okuhara T, Ueno H, Okada H, Furukawa E, Kiuchi T
Online Japanese-Language Information on Lifestyle Factors Associated With Reduced Fertility: Content Analysis
J Med Internet Res 2020;22(8):e19777
URL: https://www.jmir.org/2020/8/e19777/
doi:10.2196/19777
PMID:

Original Paper

Technology Evaluation and Assessment Criteria for Health Apps (TEACH-Apps): Pilot Study

Erica Camacho[1], MS; Liza Hoffman[1,2], LICSW; Sarah Lagan[1], BS; Elena Rodriguez-Villa[1], BS; Natali Rauseo-Ricupero[1], LCSW; Hannah Wisniewski[1], BS; Philip Henson[1], MS; John Torous[1], MD

[1]Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

[2]Bicycle Health, Boston, MA, United States

Corresponding Author:
John Torous, MD
Division of Digital Psychiatry
Beth Israel Deaconess Medical Center
Harvard Medical School
330 Brookline Avenue
Boston, MA, 02215
United States
Phone: 1 7143359858
Email: jtorous@bidmc.harvard.edu

Abstract

Background: Despite the emergence of app evaluation tools, there remains no well-defined process receptive to diverse local needs, rigorous standards, and current content. The need for such a process to assist in the implementation of app evaluation across all medical fields is evident. Such a process has the potential to increase stakeholder engagement and catalyze interest and engagement with present-day app evaluation models.

Objective: This study aimed to develop and pilot test the Technology Evaluation and Assessment Criteria for Health apps (TEACH-apps).

Methods: Tailoring a well-known implementation framework, Replicating Effective Programs, we present a new process to approach the challenges faced in implementing app evaluation tools today. As a culmination of our experience implementing this process and feedback from stakeholders, we present the four-part process to aid the implementation of mobile health technology. This paper outlines the theory, evidence, and initial versions of the process.

Results: The TEACH-apps process is designed to be broadly usable and widely applicable across all fields of health. The process comprises four parts: (1) preconditions (eg, gathering apps and considering local needs), (2) preimplementation (eg, customizing criteria and offering digital skills training), (3) implementation (eg, evaluating apps and creating educational handouts), and (4) maintenance and evolution (eg, repeating the process every 90 days and updating content). TEACH-apps has been tested internally at our hospital, and there is growing interest in partnering health care facilities to test the system at their sites.

Conclusions: This implementation framework introduces a process that equips stakeholders, clinicians, and users with the foundational tools to make informed decisions around app use and increase app evaluation engagement. The application of this process may lead to the selection of more culturally appropriate and clinically relevant tools in health care.

(J Med Internet Res 2020;22(8):e18346) doi:10.2196/18346

KEYWORDS

app; mobile phones; smartphones; app evaluation; technology

Introduction

Excitement over the myriad digital health apps available on commercial marketplaces has been tempered by emerging privacy, efficacy, usability, and implementation concerns [1]. According to industry reports, there may be over 300,000 health-related apps [2], which compared to the 20,000 prescription drug products approved for marketing by the Food and Drug Administration [3], represents the scope of the challenge in helping both consumers and providers find, evaluate, and use the right apps. As a solution, we introduce an implementation framework for app evaluation that offers an

evidence-based, practical, and impactful process to utilize across all medical fields in evaluating apps.

Challenges with current app evaluation tools and websites can be best understood from an implementation science perspective. The Replicating Effective Programs (REP) framework consists of four phases: preconditions, preimplementation, implementation, and maintenance and evolution. In the preconditions phase, needs and implementation barriers are identified. The preimplementation phase includes gathering community opinions and pilot testing the intervention. Implementation requires intervention education and technical assistance, while the maintenance and evolution phase requires the establishment of sustainable practices. Utilizing the REP framework for health care interventions [4], the need to consider stages of implementation is clear. For example, finding the right app for a consumer or clinic begins with the identification of local needs and fit; this coincides with the preconditions stage in REP. Likewise, orienting staff and consumers, customizing how apps will be used in care settings, and ensuring technical assistance is available parallels the preimplementation stage of REP. The actual evaluation of apps may be considered equivalent to the implementation stage. However, process evaluation, feedback, and refinement are often lacking in efforts today. Current app evaluation schemes are not regularly updated [5], which renders them unresponsive to dynamic changes in the app marketplace [6]. Finally, the maintenance and evolution phase of REP is missing in nearly all app evaluation efforts with results of out-of-date and incorrect information being promulgated [5]. In this paper, we present the Technology Evaluation and Assessment Criteria for Health apps (TEACH-apps) process that aims to address the current challenges surrounding app assessment and assessment maintenance.

Preconditions

The vast number of health apps available offers an opportunity to select unique tools to meet the local needs of consumers and clinics. Considerations of foreign languages that apps need to support, price points, and literacy level are all important factors that impact the adoption [7] and use of apps yet are rarely considered in current app evaluation efforts. Just as there is no single best medication or therapy, an understanding of clinical needs, fit, and resources can guide the selection of a good match—so consideration of preconditions can ensure the selection of apps that will be a good match for the end users.

Preimplementation

Customizing app evaluation to be responsive to preconditions is necessary not only to ensure that local needs are met but also that buy-in and support from clinicians and consumers are obtained. Lower levels of support around the use of apps in care today are due to many factors, but lack of active engagement from end users and clinicians plays a notable role [8,9]. Offering technical support and assistance for end users to ensure their comfort with basic app competencies and skills, such as downloading and installing apps, is another critical aspect often lacking in today's efforts.

Implementation

The focus of many existing efforts involves the actual rating of apps, as seen in the myriad of scales, scoring systems, and websites that exist for this purpose. A core challenge of these efforts is the lack of validity or reliability in their resulting metrics [10]. This deficiency of validity and reliability is understandable, given that preconditions of the actual needs of diverse consumers, regions, and cultures are not considered in most evaluation efforts. Instead, the single score generated by these efforts is meant to reflect suitability to anyone in the world—a seemingly impossible task.

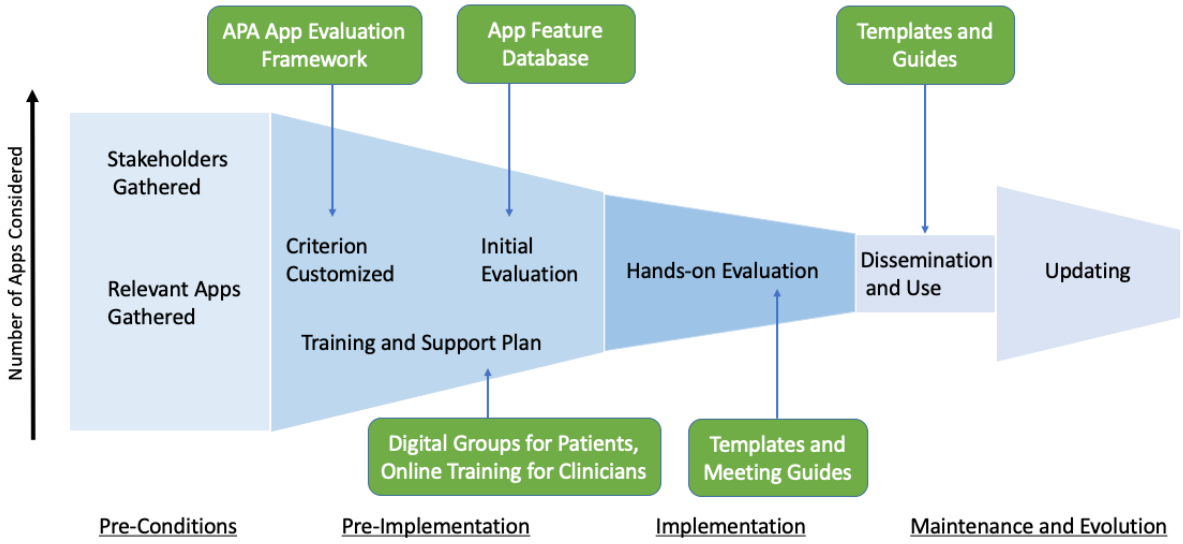
Maintenance and Evolution

Disregarding maintenance and evolution in any app evaluation process remains a critical flaw. The first version of the United Kingdom's National Health Service app evaluation library was suddenly shuttered amid research showing that the app ratings displayed did not account for recent updates to apps. These apps were no longer protecting data nor offering evidence-based care as they did when they were first evaluated [11]. A recent research report on mental health app evaluation websites noted that the average time since the last review of an app was over one year (473 days)—reflecting the inability of these services to keep pace with the dynamic world of apps which may be updated as often as weekly [5]. A tangible example is an app rating website that rated an app called Mood Triggers. At the time of this writing, clicking on the link to find this app on the marketplaces led to an error message because the app was no longer available [12]. The potential for harm as the result of offering consumers and providers out-of-date and incorrect information is evident when maintenance and evolution are ignored.

Solution

Current app evaluation efforts prioritize and concentrate on the hands-on review of apps. In the model shown in Figure 1, we present an app evaluation solution designed to ensure app evaluation is responsive to local needs, rigorous in standards, and current in content. Unlike existing processes, our method does not score or rank apps; this allows recommendations to be made based on the specific aspects matter most to a population or user. Resources available to facilitate the Technology Evaluation and Assessment Criteria for Health apps (TEACH-apps) process are shown in green. In this paper, we offer a series of customizable resources in the form of handouts and worksheets that can be found in Appendix 1. To the best of our knowledge, this is the first implementation-focused app evaluation effort that provides handouts, resources, and tools to support the app evaluation process further.

Figure 1. Schematic of the TEACH-apps model for implementing app evaluation. This model presents the app evaluation process step-by-step from left to right. The height of each TEACH-apps stage represents the number of apps considered. The preconditions stage has the largest height because it involves the initial gathering of all of the apps to be considered. After apps have been evaluated within the implementation stage, apps that meet the criteria are disseminated in the maintenance and evolution stage. The 'updating' portion of the final stage, which occurs 90 days from the dissemination period, has an increased height to account for new apps that may have surfaced or apps that have improved since the dissemination period. APA: American Psychiatric Association.



Methods

The American Psychiatric Association (APA) App Evaluation framework offers a useful tool to guide informed decision making around apps. Supported by evidence [13,14], international stakeholders [15], and frequently cited in research [16,17] on app evaluation, the framework offers a simple and ethically grounded approach that first considers access, then in sequence privacy/safety, evidence, usability, and finally clinical integration. The hierarchical nature of the framework ensures consideration of factors frequently overlooked by many other app evaluation tools, such as privacy and safety [14]. While the framework is notable and derives its name from the professional organization that has adopted it, no aspect of the framework is specific to mental health, a reflection of its broad approach and generalizability. As outlined below, we use the APA App Evaluation framework as the tool to be customized in the app evaluation process and acknowledge that other frameworks may be substituted. The focus should not be on any single tool but rather on how people use it in the four steps discussed next.

Participants

Ten committee members were recruited from our team's hospital at Beth Israel Lahey Health. Our committee consisted of a range of stakeholders from undergraduate students to research assistants to clinicians. Input from this diverse pool of stakeholders holds unique potential to yield a comprehensive understanding of the benefits and concerns regarding each app.

Preconditions

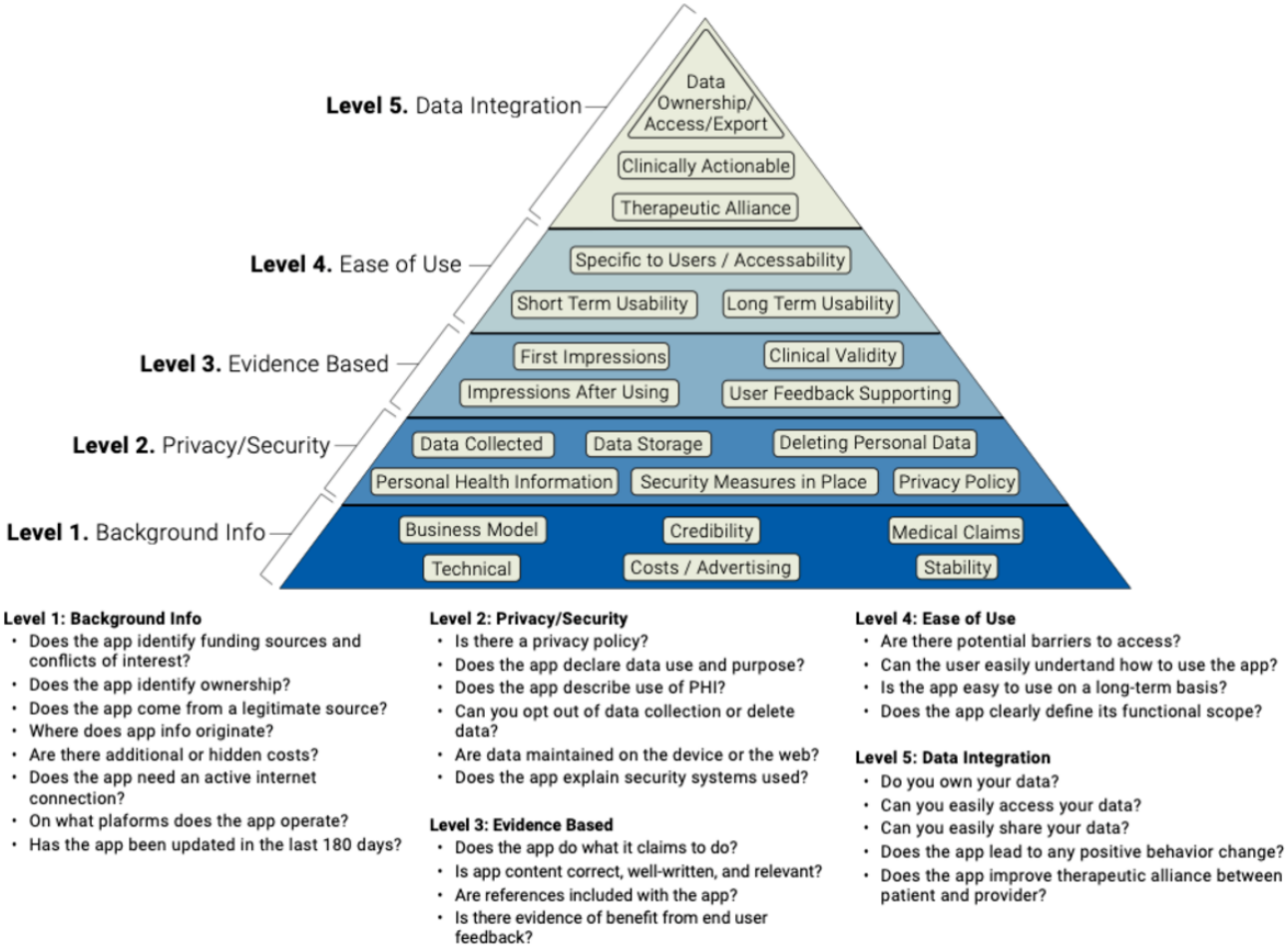
Asking all stakeholders to submit the names of any apps they may have heard of, used, or have an interest in will create a pool of apps for consideration. To supplement the app list produced by stakeholders, conduct a narrow search on the app store by choosing a specific category of apps such as physical health or mindfulness. Local app needs can be identified by asking participants to name areas in which apps may be of help, even if they do not know of an app that suits, will highlight local needs around apps.

In parallel, at this stage, recruitment of a committee interested in app evaluation begins. The committee should be diverse and represent a variety of stakeholders with unique perspectives. Technology expertise is not a requirement for the committee, but interest in the topic and willingness to learn more is encouraged. Committee members are responsible for evaluating a certain number of apps broadly, then making decisions on whether to push the app forward into the next stage of the evaluation process.

Preimplementation

The preimplementation stage involves an in-person meeting with the app evaluation committee. Although an online, self-directed survey assessment is feasible, it has not yet been tested by our team. The initial task of the committee is to reflect on feedback from the preconditions stage and utilize it to establish what app evaluation criteria should be added or removed from the APA App Evaluation framework. For example, if people are not willing or able to pay for apps, a free criterion feature in the "access" level is necessary. Likewise, if the need for strong evidence is not considered crucial as people will be using these apps under close supervision and as an adjunct to care, related questions may be removed. There is no right or wrong customization of the APA App Evaluation framework, and the goal is to ensure it reflects the priorities and needs of the local users. It is easy to add or remove questions as deemed appropriate and necessary during the preceding stages of the process. As Henson et al demonstrated, the APA framework organizes the primary evaluation considerations into five levels—background information gathering, privacy, evidence, ease of use, and interoperability (Figure 2) [14]. Sample questions for each level are also provided in Figure 2.

Figure 2. American Psychiatric Association App Evaluation framework and sample questions corresponding to each level.



The committee can also add any apps to the list of those submitted they believe to be potentially valuable and may remove any for which they have serious concerns. By personalizing the framework and eliminating apps that fail to meet the consumer's needs, stakeholders further narrow their list of apps. At this stage, the committee can also reflect on the need for technical assistance and user support in working with apps.

The final component of preimplementation may be the most time consuming but can be done outside of the in-person committee meeting. The goal is to evaluate the available apps selected for consideration in the context of the final criteria and note which apps best align with the defined criteria. No scores or points are assigned, as the goal at this stage is to allow a closer inspection of the apps to separate those that have serious flaws from those that may be acceptable.

Implementation

The list of apps deemed acceptable in the preimplementation stage is now brought back to the committee, and members are invited to interact with and test drive the apps. While some apps can likely be tested in a matter of minutes (eg, informational apps), others such as medication reminders or habit-forming apps may require a more extensive test drive. The goal at this stage is to collect feedback on which apps may be a better fit in terms of usability and offer the most clinical value. See Multimedia Appendix 1 for a sample questionnaire. These are inherently more subjective decisions that rely on the support from the prior three stages and are informed by local needs, fit, and resources. The final apps that are deemed appropriate can then be transformed into an educational handout that explains why the app was selected and briefly outlines the pros and cons.

Maintenance and Evolution

While there is no absolute number of days within which app recommendations must be updated, there is evidence that 180 days may be a good target [18]. Thus, we recommend the process be repeated at least biannually, noting that updating apps will often be faster and simpler. Regular repetition of the process, in addition to ensuring that recommended apps are up to date and of high quality, also enables clinicians and stakeholders to account for changes in consumer preferences over time. New apps should be given priority for evaluation, but the volume will likely be less given that with further rounds, the apps that meet user needs should rise to the top. Of course, some top apps from180 days ago may no longer be available, have changed dramatically, or now irrelevant; the committee can check for this in the implementation stage of the process. The maintenance and evolution phase of such a system has been outlined elsewhere, with future directions including self-certification by app developers and verification by committees [19]. Developers answer questions related to the safety and efficacy of an app. Responses from the self-certification checklist are made publicly available so users can confirm or reject the validity of answers. This system helps

hold developers accountable and encourages them to build apps that are safe and effective [19].

Results

Preconditions

The TEACH-apps process was pilot tested at our hospital by a team of ten committee members with diverse levels of expertise. During the initial committee meeting, the team discussed and determined the focus of their app evaluation efforts: health-related apps. Although this focus reduced the list of apps significantly from about two million to over three hundred thousand, the team relied on clinician recommendations, personal experiences or familiarity with apps, and popularity on the Apple App Store and the Google Play Store to further narrow their list. This process resulted in the creation of an initial list of 180 apps.

We understand that not all committees following the TEACH-apps process will have the opportunity to host an initial meeting to gather relevant apps. Thus, this stage can be conducted through email (Multimedia Appendix 2), or discussion at staff meetings. Another avenue to collect this information is to provide handouts in waiting rooms (Multimedia Appendix 3), which can be collected by front desk staff. The goal is to cast a wide net to see what apps people are using or interested in and what roles they hope apps could play.

Preimplementation

Stage two of TEACH-apps involved an in-person committee meeting where the team was asked to determine what app evaluation criteria should be added or removed from the APA App Evaluation framework. Analysis of the APA App Evaluation framework allowed the committee to develop their top eight general categories of interest in no specific order: (1) privacy, (2) medical evidence, (3) price, (4) ratings, (5) attributes, (6) features, (7) onboarding, and (8) performance.

This initial evaluation allowed the committee to narrow the list from 180 to 56 apps. We acknowledge there is no 'perfect' app, and those evaluating apps during this stage will have to use some level of discretion informed by the precondition and preimplementation stages. Our team recognized that not everyone interested in conducting the TEACH-apps process would be able to dedicate this much time to the evaluation. Since then, our team has created a database of common and popular apps where we have coded their features, thus creating a resource to help expedite the preimplementation process. The database allows participants to narrow their search by applying filters such as language or disorder. This database is publicly available [20].

Implementation

The implementation stage of TEACH-apps consisted of a hands-on app evaluation by the committee. The in-depth analysis of each app involved reading the privacy policy of all 56 apps and documenting key aspects from the customized criteria developed in the preimplementation stage (Table 1). Each committee member was responsible for reviewing 5-6 apps independently, which took an average of 30 minutes per app. As members became increasingly familiar and comfortable with the evaluation process, the time required to review an app decreased. Since apps are not scored, apps with an overall positive evaluation from stakeholders and little to no privacy concerns were then offered to residents at Beth Israel Lahey Health for further clinical evaluation (Multimedia Appendix 1). The hands-on evaluation conducted by the committee resulted in the reduction of the app list from 56 apps to 27. Apps that clinicians rated a 3 ("There are several people I would recommend this app to") or higher were turned into a simple handout that outlined the pros, cons, cost, and download file size (Figure 3). Privacy was not included on this handout because all apps that reached this stage were previously checked for privacy and safety concerns. This handout can be offered to clinicians, consumers, and others who may find it of value. Resources related to helping use apps and technology, in general, can also be offered as a section on the handout. A disclaimer should be included to ensure users understand the handout provides educational resources but is not offering clinical recommendations.

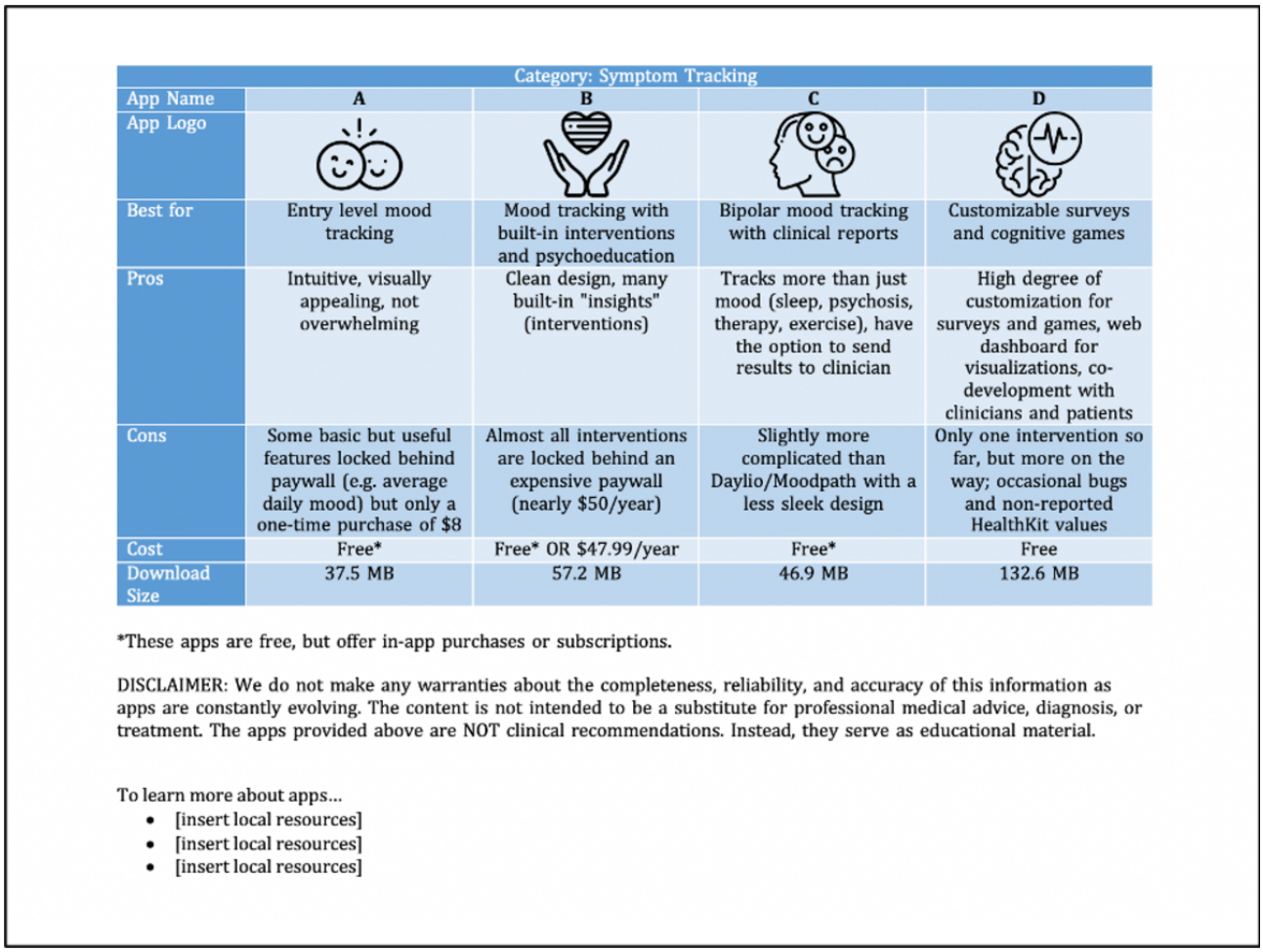
Table 1. App evaluation categories and questions. Privacy is evaluated by assessing each app's privacy policy. Medical evidence is assessed through a quick internet search. The Price, Ratings, and Attributes categories can be answered with information collected in the App Store. Lastly, the Features, Onboarding, and Performance categories are evaluated by downloading and interacting with the app.

Category and subcategory	Evaluation instructions
Privacy	
Privacy policy	0 – Does not have a privacy policy 1 – Does have a privacy policy
Data type	0 – Personalized data or can't find 1 – Deidentified data 2 – Anonymized data or none
Disclaimer	0 – App claims to provide a medical intervention 0 – Can't find statement / unsure 1 – App does not claim to be a substitute for medical care
Delete data	0 – No 0 – Can't find statement on this/unsure 1 – Yes
Data protection	0 – No statement 0 – Can't find/unsure 1 – Statement on data protection from start to finish
Data location	0 – Data is stored elsewhere 0 – Can't find a statement on this 1 – Data remains on the device or data is intended to leave
Data sharing I	0 – Data is sold or shared 0 – Can't find statement on this 1 – Data is NOT sold or shared
Data sharing II	0 – Can't find statement on this 1 – Data is sold or shared if/when the company is sold 2 – Data is NOT sold or shared if/when the company is sold
Trusted developer	0 – No trust or unsure 1 – Trust
Total	0-10
Medical	
Evidence (publications of app efficacy)	Yes or no
Price	
Cost	Input value
Business model	Check all that apply: • Free • Free with in-app purchase • One-time purchase • Subscription
Ratings	
Stars	Check 1-5
Number of reviews	<100 <1000 <10,000 <100,000 100,000+
Attributes	

Category and subcategory	Evaluation instructions
Download size	0-50 MB 50-100 MB 100-200 MB 200+ MB
Availability	Check all that apply: • Android • iOS • Web
Internet required	Yes For some features No
Accessibility	Check all that apply: • Multiple languages • Text/button size • Literacy level • Microphone option for data input
Features	
Advertisements	Yes or no
Features	Check all that apply: • Mood tracking • Step count • Medication Tracker • Sleep Tracking • Psychoeducation/resources • Journal • Picture gallery/hope board • Connection to coach/therapist • Mindfulness meditation • Relaxation exercises • Deep breathing • ACT/Cognitive diffusion • CBT thought exercises • Worry time • Peer support • Goal setting/habits
Onboarding	
Login requirement	Check all that apply: • Requires sign up (eg, username and password) • Requests personal information (eg, birthday, name, email) • No signup or personal information required
Data collected	Check all that apply: • Surveys • Health Kit/Google Fit • Birthday • Photos • Location (eg, GPS) • Medication
Performance	
Effectiveness	1 – App is broken; no/insufficient/inaccurate response (eg, crashes/bugs/broken features, etc) 2 – Some functions work, but lagging or contains major technical problems 3 – App works overall. Some technical problems need fixing or slow at times 4 – Mostly functional with minor/negligible problems 5 – Perfect/timely response; no technical bugs found, or contains a “loading time left” indicator (if relevant)

Category and subcategory	Evaluation instructions
Endorsement	1 – I would not recommend this app to anyone 2 – There are very few people I would recommend this app to 3 – There are several people I would recommend this app to 4 – There are many people I would recommend this app to 5 – I would recommend this app to everyone
Key takeaways: Any usability issues (small text, difficult to navigate)? Who might this be most useful for, and for what purpose? Who might struggle using this app?	Open-ended box
Share with clinicians?	Yes or no

Figure 3. Sample educational handout with major app characteristics documented.



Category: Symptom Tracking				
App Name	A	B	C	D
App Logo				
Best for	Entry level mood tracking	Mood tracking with built-in interventions and psychoeducation	Bipolar mood tracking with clinical reports	Customizable surveys and cognitive games
Pros	Intuitive, visually appealing, not overwhelming	Clean design, many built-in "insights" (interventions)	Tracks more than just mood (sleep, psychosis, therapy, exercise), have the option to send results to clinician	High degree of customization for surveys and games, web dashboard for visualizations, co-development with clinicians and patients
Cons	Some basic but useful features locked behind paywall (e.g. average daily mood) but only a one-time purchase of $8	Almost all interventions are locked behind an expensive paywall (nearly $50/year)	Slightly more complicated than Daylio/Moodpath with a less sleek design	Only one intervention so far, but more on the way; occasional bugs and non-reported HealthKit values
Cost	Free*	Free* OR $47.99/year	Free*	Free
Download Size	37.5 MB	57.2 MB	46.9 MB	132.6 MB

*These apps are free, but offer in-app purchases or subscriptions.

DISCLAIMER: We do not make any warranties about the completeness, reliability, and accuracy of this information as apps are constantly evolving. The content is not intended to be a substitute for professional medical advice, diagnosis, or treatment. The apps provided above are NOT clinical recommendations. Instead, they serve as educational material.

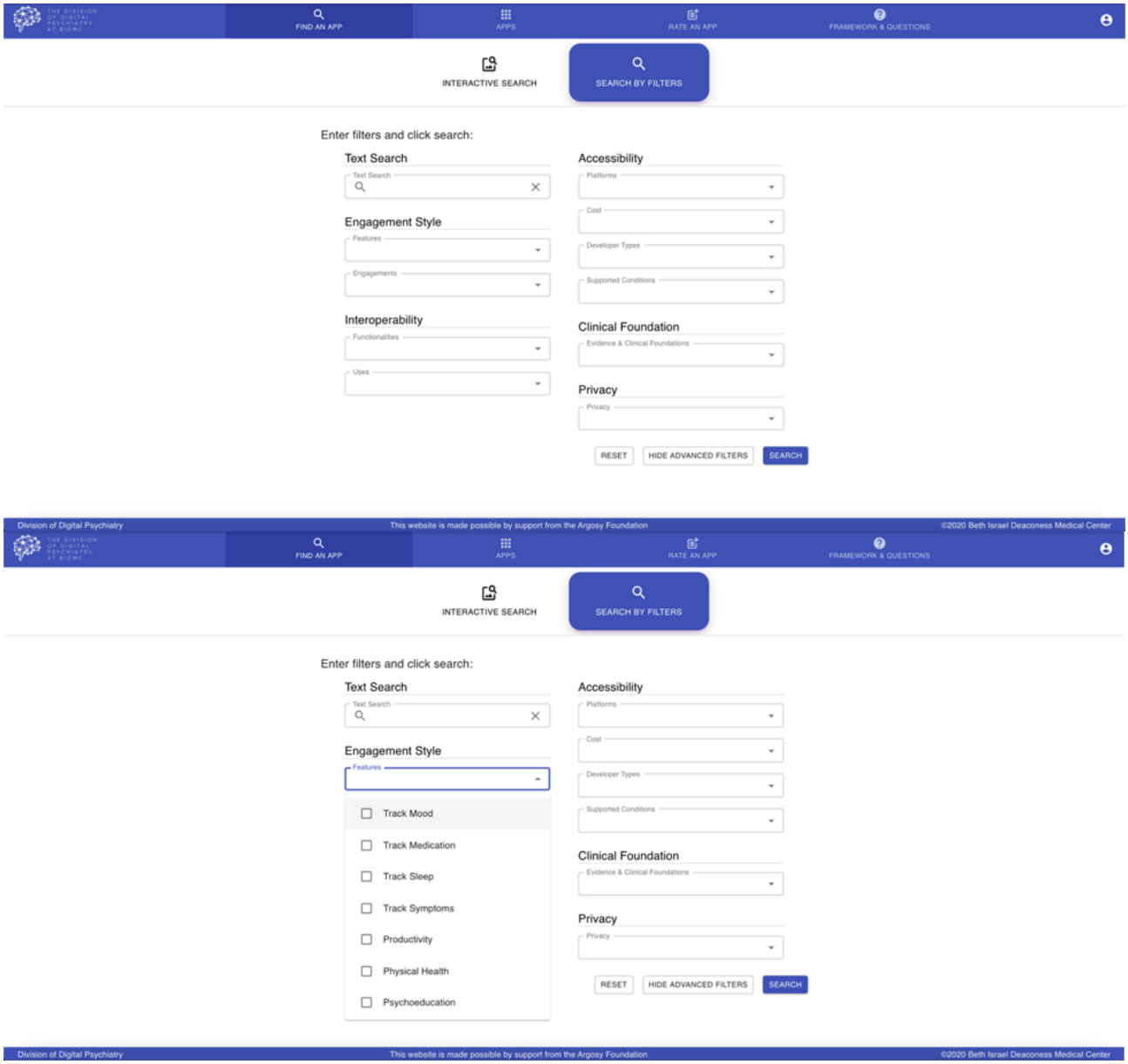
To learn more about apps...

- [insert local resources]
- [insert local resources]
- [insert local resources]

Maintenance and Evolution

While maintenance and evolution will vary by site and need, our solution includes a publicly accessible online database [20]. Unlike a paper handout, which is impossible to update except with redistribution to each person, a website can be updated in real time to ensure that information is accurate and up to date. A website also offers an easy means to track history and changes in the evaluation of individual apps, creating a transparent record of maintenance and evolution. We created a search feature so that individuals accessing the database could enter their preferences and learn which apps may be a close match. A screenshot of the website is shown in Figure 4.

Figure 4. Screenshots of the database for smartphone app filtering. The top panel showcases the various filters that users can use to find an app. The lower panel displays some of the features users can choose under 'Engagement Style.' The rest of the features in this category can be seen by scrolling on the website.



Discussion

Outcomes of Success

Like any process, we will keep updating the app evaluation process based on feedback. To that end, we offer a pre- and postsurvey to both committee members and end users to identify areas for continual improvement (see Multimedia Appendix 4 for committee members, and Multimedia Appendices 5 and 6 for end users). We propose that administering the presurvey at the first in-person committee meeting during the preimplementation stage. The committee members complete the post-survey after implementation. An analysis of success must consider both the degree to which the proposed process met local needs in app evaluation and the feasibility of process maintenance. The surveys thus pose questions about comfort level and training with technology, the potential for integration of technology into care, and the potential for adaptation with technology over time. The same set of questions is answered before and after implementation to assess the impact of the proposed app evaluation process.

Additionally, users of the app recommendations derived from the process will complete a survey before and after receiving the educational material (Multimedia Appendices 5 and 6). These surveys focus on app use before and after perusal of the handout as well as familiarity with different components of the app evaluation process, such as privacy.

The templates included here serve only as guidelines. Thorough consideration of the effectiveness of the app evaluation process, local needs and goals should be considered. There is thus a great deal of latitude to amend survey questions as committee members and stakeholders see fit.

Limitations

Our study has several weaknesses that must be considered. With hundreds of thousands of apps available on the dashboard, our initial evaluation of 180 apps only scratches the surface of app evaluation. This preliminary list, generated during the preconditions phase, was based on recommendations, personal experiences, and app popularity. Thus, the preconditions phase, in particular, may result in the exclusion of less popular or less well-known apps. Another limitation is that while there is interest in implementing this process at other health care facilities, this process has only been tested once due to disruptions caused by coronavirus disease (COVID-19), and our newly developed database has not been tested. Although we have provided comprehensive resources to conduct both the preconditions and preimplementation phases digitally, an additional limitation stems from the fact that we were able to hold in-person meetings, and our pilot process was thus informed by in-person discussion and debate that may be lost in online surveys.

Conclusions

As digital health apps continue to grow in number and prominence, there is an increasing need for solutions that help consumers and providers find and evaluate relevant apps while taking into account concerns about privacy, efficacy, usability, and implementation.

We have proposed an implementation framework that equips clinicians and users with the tools to make informed decisions around app use. The framework may be applied to a wide variety of medical contexts and can be customized to address specific segments of the estimated 300,000 health apps currently on the market. For example, although the preconditions for an assessment of meditation apps versus apps for diabetes management differ, the outline of the process remains intact.

The process is comprehensive, with a preconditions stage focused on gathering relevant apps, a preimplementation stage to customize the criteria of evaluation to local needs, and the ultimate implementation comprising a hands-on test run of the apps that met the tailored set of criteria. With this process completed, stakeholders, clinicians, and users can more easily navigate the dynamic digital health space and utilize health apps for the advancement of care and well-being.

Authors' Contributions

Conception/model: JT, PH, LH
Data collection: EC, JT, PH, LH
Analysis/interpretation: EC, SL, HW, ERV, NRR, PH, JT, LH
Writing: EC, SL, HW, ERV, NRR, PH, JT, LH*
Editing: EC, JT
*Resource 3, sections of Table 1, sections of Figure 3.

Conflicts of Interest

None declared.

Multimedia Appendix 1
App evaluation questionnaire to be given to clinicians.
[PNG File , 78 KB - jmir_v22i8e18346_app1.png]

Multimedia Appendix 2
An example of a Google Form that can be easily circulated via email. Although this illustration is centered around health in general, the form can be tailored to fit the local needs.
[PNG File , 77 KB - jmir_v22i8e18346_app2.png]

Multimedia Appendix 3
Sample handout for distribution in waiting rooms.
[PNG File , 158 KB - jmir_v22i8e18346_app3.png]

Multimedia Appendix 4
Survey templates to be completed before and after app evaluation by committee members.
[PNG File , 49 KB - jmir_v22i8e18346_app4.png]

Multimedia Appendix 5
Survey templates to be completed by end users before viewing educational material.
[PNG File , 41 KB - jmir_v22i8e18346_app5.png]

Multimedia Appendix 6
Survey templates to be completed by end users after viewing educational material.
[PNG File , 57 KB - jmir_v22i8e18346_app6.png]

References

1. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019 Feb;18(1):97-98 [FREE Full text] [doi: 10.1002/wps.20592] [Medline: 30600619]
2. Pohl M. 325,000 mobile health apps available in 2017. URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017/ [accessed 2018-01-01]
3. US Food & Drug Administration. Fact Sheet: FDA at a Glance. URL: https://www.fda.gov/about-fda/fda-basics/fact-sheet-fda-glance [accessed 2019-10-18]
4. Kilbourne AM, Neumann MS, Pincus HA, Bauer MS, Stall R. Implementing evidence-based interventions in health care: application of the replicating effective programs framework. Implement Sci 2007 Dec 09;2:42 [FREE Full text] [doi: 10.1186/1748-5908-2-42] [Medline: 18067681]
5. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019 Jun 17;2(1):54 [FREE Full text] [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]
6. Larsen ME, Nicholas J, Christensen H. Quantifying App Store Dynamics: Longitudinal Tracking of Mental Health Apps. JMIR Mhealth Uhealth 2016 Aug 09;4(3):e96 [FREE Full text] [doi: 10.2196/mhealth.6020] [Medline: 27507641]
7. Young AS, Cohen AN, Niv N, Nowlin-Finch N, Oberman RS, Olmos-Ochoa TT, et al. Mobile Phone and Smartphone Use by People With Serious Mental Illness. Psychiatr Serv 2020 Mar 01;71(3):280-283. [doi: 10.1176/appi.ps.201900203] [Medline: 31744429]
8. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
9. Feijt MA, de Kort YA, Bongers IM, IJsselsteijn WA. Perceived Drivers and Barriers to the Adoption of eMental Health by Psychologists: The Construction of the Levels of Adoption of eMental Health Model. J Med Internet Res 2018 Apr 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]
10. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater Reliability of mHealth App Rating Measures: Analysis of Top Depression and Smoking Cessation Apps. JMIR Mhealth Uhealth 2016 Feb 10;4(1):e15 [FREE Full text] [doi: 10.2196/mhealth.5176] [Medline: 26863986]
11. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 25;13(1). [doi: 10.1186/s12916-015-0444-y]
12. Mood Triggers: Anxiety Sadness Review. One Mind PsyberGuide. URL: https://psyberguide.org/apps/mood-triggers-anxiety-sadness/ [accessed 2020-03-15]
13. Torous JB, Chan SR, Gipson SYT, Kim JW, Nguyen T, Luo J, et al. A Hierarchical Framework for Evaluation and Informed Decision Making Regarding Smartphone Apps for Clinical Care. Psychiatr Serv 2018 May 01;69(5):498-500. [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]
14. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. The Lancet Digital Health 2019 Jun;1(2):e52-e54. [doi: 10.1016/s2589-7500(19)30013-5]
15. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019 Feb;18(1):97-98 [FREE Full text] [doi: 10.1002/wps.20592] [Medline: 30600619]
16. Hategan A, Giroux C, Bourgeois JA. Digital Technology Adoption in Psychiatric Care: an Overview of the Contemporary Shift from Technology to Opportunity. J. technol. behav. sci 2019 Mar 7;4(3):171-177. [doi: 10.1007/s41347-019-00086-x]
17. Kowatsch T, Otto L, Harperink S, Cotti A, Schlieter H. A design and evaluation framework for digital health interventions. it-Information Technology 2019 Nov 20;61(5-6):253-263. [doi: 10.1515/itit-2019-0019]
18. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela J, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evid Based Ment Health 2019 Feb;22(1):4-9 [FREE Full text] [doi: 10.1136/ebmental-2018-300069] [Medline: 30635262]
19. Rodriguez-Villa E, Torous J. Regulating digital health technologies with transparency: the case for dynamic and multi-stakeholder evaluation. BMC Med 2019 Dec 03;17(1):226 [FREE Full text] [doi: 10.1186/s12916-019-1447-x] [Medline: 31801532]
20. Digital Psychiatry App Database.: Division of Digital Psychiatry at Beth Israel Deaconess Medical Center URL: https://apps.digitalpsych.org/ [accessed 2020-07-06]

Abbreviations

APA: American Psychiatric Association
COVID-19: coronavirus disease
REP: Replicating Effective Programs
TEACH-apps: Technology Evaluation and Assessment Criteria for Health apps

Edited by G Eysenbach; submitted 20.02.20; peer-reviewed by L Bonet, S Acquilano, B Saalfeld; comments to author 08.04.20; revised version received 17.05.20; accepted 13.06.20; published 27.08.20.

Please cite as:
Camacho E, Hoffman L, Lagan S, Rodriguez-Villa E, Rauseo-Ricupero N, Wisniewski H, Henson P, Torous J
Technology Evaluation and Assessment Criteria for Health Apps (TEACH-Apps): Pilot Study
J Med Internet Res 2020;22(8):e18346
URL: https://www.jmir.org/2020/8/e18346
doi:10.2196/18346
PMID:32535548

Original Paper

Personal Health Information Inference Using Machine Learning on RNA Expression Data from Patients With Cancer: Algorithm Validation Study

Solbi Kweon[1,2*], BS; Jeong Hoon Lee[1*], MS; Younghee Lee[3], PhD; Yu Rang Park[1], PhD

[1]Department of Biomedical System Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Department of Medical Engineering, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[3]Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT, United States

[*]these authors contributed equally

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical System Informatics
Yonsei University College of Medicine
50-1 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 ext 2493
Fax: 82 2 2227 8308
Email: yurangpark@yuhs.ac

Abstract

Background: As the need for sharing genomic data grows, privacy issues and concerns, such as the ethics surrounding data sharing and disclosure of personal information, are raised.

Objective: The main purpose of this study was to verify whether genomic data is sufficient to predict a patient's personal information.

Methods: RNA expression data and matched patient personal information were collected from 9538 patients in The Cancer Genome Atlas program. Five personal information variables (age, gender, race, cancer type, and cancer stage) were recorded for each patient. Four different machine learning algorithms (support vector machine, decision tree, random forest, and artificial neural network) were used to determine whether a patient's personal information could be accurately predicted from RNA expression data. Performance measurement of the prediction models was based on the accuracy and area under the receiver operating characteristic curve. We selected five cancer types (breast carcinoma, kidney renal clear cell carcinoma, head and neck squamous cell carcinoma, low-grade glioma, and lung adenocarcinoma) with large samples sizes to verify whether predictive accuracy would differ between them. We also validated the efficacy of our four machine learning models in analyzing normal samples from 593 cancer patients.

Results: In most samples, personal information with high genetic relevance, such as gender and cancer type, could be predicted from RNA expression data alone. The prediction accuracies for gender and cancer type, which were the best models, were 0.93-0.99 and 0.78-0.94, respectively. Other aspects of personal information, such as age, race, and cancer stage, were difficult to predict from RNA expression data, with accuracies ranging from 0.0026-0.29, 0.76-0.96, and 0.45-0.79, respectively. Among the tested machine learning methods, the highest predictive accuracy was obtained using the support vector machine algorithm (mean accuracy 0.77), while the lowest accuracy was obtained using the random forest method (mean accuracy 0.65). Gender and race were predicted more accurately than other variables in the samples. On average, the accuracy of cancer stage prediction ranged between 0.71-0.67, while the age prediction accuracy ranged between 0.18-0.23 for the five cancer types.

Conclusions: We attempted to predict patient information using RNA expression data. We found that some identifiers could be predicted, but most others could not. This study showed that personal information available from RNA expression data is limited and this information cannot be used to identify specific patients.

(J Med Internet Res 2020;22(8):e18387) doi:10.2196/18387

KEYWORDS

cancer; privacy issue; personal information; prediction; RNA sequencing; machine learning

Introduction

High-throughput sequencing and array technologies, such as next-generation sequencing and microarrays, can be applied to personalized genomics and for medical purposes. These technologies will enable comprehensive multiomics analysis at various levels, including genomics, transcriptomics, and proteomics. In the last decade, the ability to collect and store personal data has increased significantly. A growing number of studies around the world have used multidimensional cancer genome data sets to obtain biological insights and develop clinical applications [1]. The ability to collect and store personal data has exploded, making genomic analysis a viable method for improving diagnostic accuracy and personalized medicine.

These advances require both the collection and sharing of high-resolution genetic profiles among researchers and institutions. However, this large-scale use of detailed individual-level data raises legitimate privacy concerns. It has been proposed that genetic profiles should not be collected and shared due to the potential for privacy breaches and risk of participant identification. There are standards outlined by modern data protection laws, such as the General Data Protection Regulation in the European Union, for the anonymization of data before sharing. The new General Data Protection Regulation explores the major provisions of this new regulation with regard to processing genetic data and includes it as a special category of sensitive data [2]. The Privacy Rule of the Health Insurance Portability and Accountability Act (HIPAA) sets standards for the privacy and security of health records in the United States [3]. Public databases such as The Cancer Genome Atlas (TCGA) obtain patient consent to share their genetic data. Consent is obtained due to the possible risk of exposing patient information obtained from multiomics data [4].

We have yet to discover everything there is to learn from genomes [5]. The study of personal genome interpretation using genomic data has continued to evolve and has now reached the point of being able to explain individual characteristics [6,7]. RNA expression data, a next-generation sequencing-based method for analysis of transcription, provides valuable information on the expression of specific genes [8], and it is also considered to be sensitive data [9,10]. RNA expression analysis is performed on bulk tissue samples or cell populations. Differences in cellular RNA expression profiles are caused by various factors such as cell cycle status, differentiation, and morphologic position. Despite increasing research using genomic data, there is a lack of research to determine the appropriate level of data sharing based on the predictability of personal information.

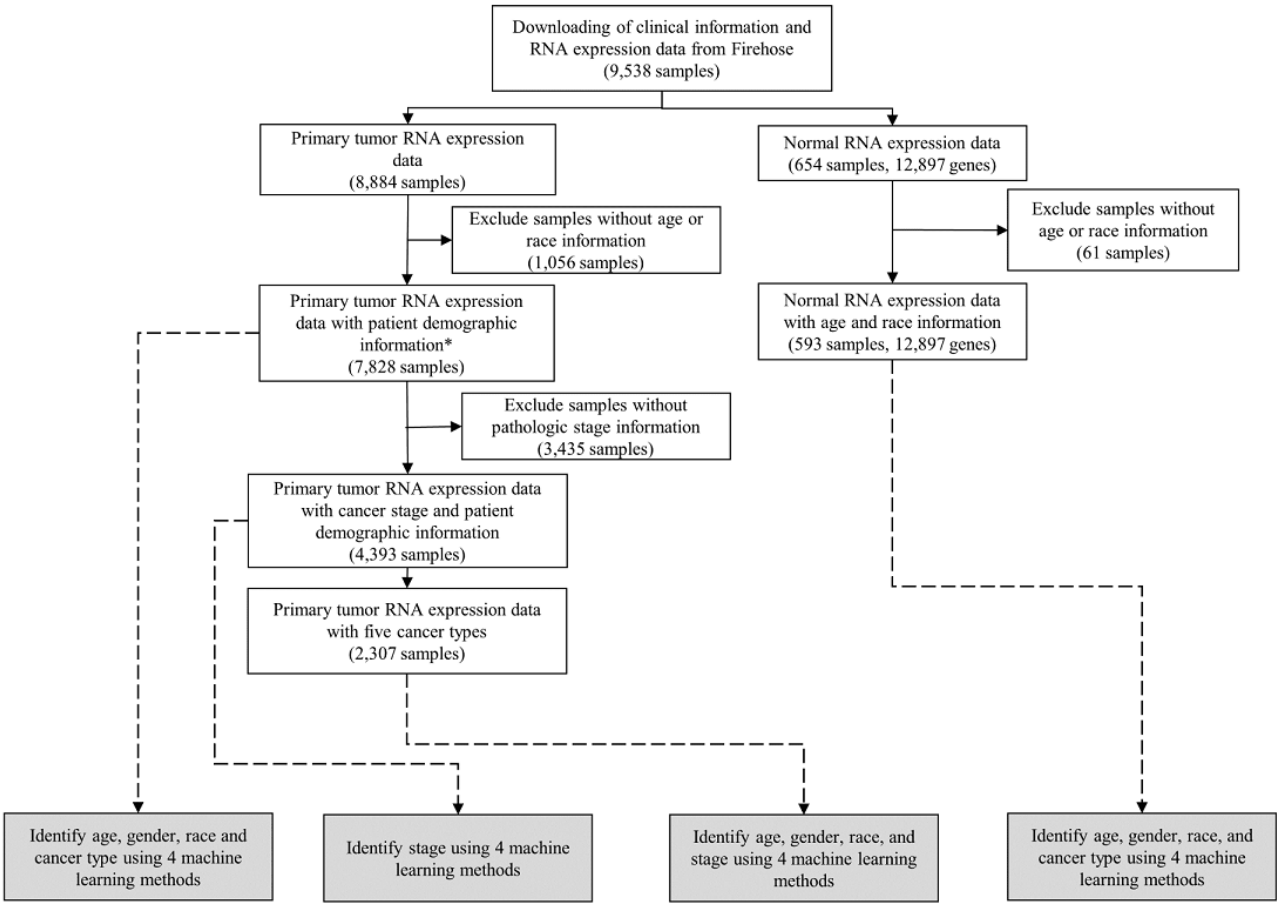
In order to protect personal information while sharing genomic data, it is necessary to evaluate all patient information that could be revealed by genomic data only. Since 2000, numerous papers that use machine learning algorithms in genome-wide analyses have been published [11]. The aim of this study is to assess whether machine learning algorithms can use RNA expression data from a public cancer genome database (TCGA) to identify patients' personal information.

Methods

Study Design

We selected five personal data features from 9538 RNA expression samples for identification: gender, age, cancer type, race, and pathologic (cancer) stage. Information on these five features was extracted from the clinical data and separately predicted for each sample. Our data processing workflow is summarized in Figure 1. Gender consisted of two groups (male and female), age was classified into nine groups (each ranging 10 years of age), cancer type consisted of 32 groups, race consisted of five groups, and cancer stage consisted of four groups (stages I to IV). Since gender, age, cancer type, and race are typical variables, patient data with all four variables were included. However, cancer stage data were missing from many samples, and its definition differs according to cancer type, thus patient data missing this information were still included.

Figure 1. Patient inclusion and exclusion criteria (white boxes) and flowchart of the study design. The dashed lines indicate RNA expression data used to develop machine learning models. The gray boxes show machine learning models. *Demographic information: age, gender, and race.



Availability of Data and Materials

All RNA expression data were generated as part of The Cancer Genome Atlas [12]. These data were obtained from the FireBrowse website [13].

Database

Gene expression information (level 3) from TCGA was downloaded from the Firehose analysis infrastructure (Broad Institute Genome Data Analysis Center). The RNA expression level 3 data contained reads per kilobase million mapped reads [14], RNA expression by expectation-maximization [15], read count, and clinical data. The TCGA level 3 RNA expression data set contained quantifications of transcript levels by normalized counts calculated using the expectation-maximization method.

We used 7828 primary tumor samples from 32 cancer types and 593 normal samples from 22 cancer types. Samples where clinical data did not exist were excluded. We used HiSeq (HiSeq 2000; Illumina Inc) RNA expression values for 12,897 genes (12,883 genes excluding the expression values of the genes on the Y chromosome) in our machine learning platforms to predict gender. In addition to the data set with samples from multiple cancer types, we created five separate data sets, each consisting of samples from a single cancer, in order to compare predictability between cancer types. These five data sets represented breast carcinoma (957/7828, 12.24%), kidney renal clear cell carcinoma (519/7828, 6.64%), head and neck squamous cell carcinoma (496/7828, 6.35%), low-grade glioma (486/7828, 6.23%), and lung adenocarcinoma (416/7828, 5.34%). These data sets were used to determine whether personal information was able to be predicted in each cancer type. Specifically, the data sets were used to determine whether the stage was predictable and how the prediction accuracy of other identifiers would change.

Processing of RNA-Sequence Data

Level 3 RNA expression data processing and quality control were performed by the Broad Institute TCGA workgroup. Data were processed in R (version 3.6.0; packages: edgeR, limma). Expectation maximization–normalized data were preprocessed using the DGEList function (edgeR), and only genes expressed with counts per million above zero in at least 20% of samples were retained using the CPM function (edgeR) [16]. Level 3 RNA expression data were normalized within each cancer type. We performed preprocessing of level 3 RNA expression data of primary tumor samples and normal samples using the voom function (limma), which is also an alternative variance-stabilizing multiple-testing framework for RNA expression data [17]. Comparison of gene expression between cancer types was performed using linear regression models and the transformed data were then used to derive the final differential gene expression list (voom; limma). Our integrated data set contained expression values of 12,897 genes from 9538 samples.

Personal Health Identifiers

Under the HIPAA privacy rules, personal information typically includes information that can be used to identify or track an individual, such as their name, social security number, or biometric records, either alone or in combination with other information linkable to a specific individual, such as a date or place of birth [18]. Among the available patient data, we selected five informative features that could be connected to a specific individual. HIPAA provides 18 identifiers; however, the personal information identifiers provided by TCGA are restricted. There were 29 variables associated with TCGA patient information data, from which we excluded the sample barcode, version, and survival variables. Additionally, variables missing more than 60% of data were excluded. TCGA-provided identifiers existed only for age, and we added the demographic information (ie, gender, race, stage, and cancer type) to the personal information.

Selection of Significant Genes for Predicting Personal Information

The genes in the primary tumor data set relating to gender were analyzed using two-tailed *t* tests [19,20] with Bonferroni correction for any two-group comparisons. Other genes in the primary tumor data set relating to the remaining variables were analyzed using a one-way analysis of variance (ANOVA) [21,22] and Bonferroni posthoc tests for multiple comparisons. In addition to the 12,897 genes in the primary tumor data set, which included RNA gene expression levels, we created two data sets for each gene (P value≤.01) based on the P values of the ANOVA and *t* tests (Multimedia Appendix 1). The purpose of creating these data sets was to provide alternative data sets for evaluating whether personal information could be identified by selecting significant genes.

Supervised Machine Learning Algorithms

We used four different supervised machine learning algorithms to generate classification models. Support vector machines are a group of related supervised learning methods used for classification and regression [23,24]. Decision tree structures use leaves to represent classifications, while branches represent conjunctions of features that lead to those classifications [23,25]. Random forest is a classifier consisting of many decision trees and determines the class, which is the mode of classes generated, by individual trees [25,26]. Artificial neural networks are an interconnected group of nodes that use a computational model for information processing. Its structure changes based on external or internal information that flows through the network. Artificial neural networks can be used to model complex relationships between inputs and outputs and find patterns in data [25,27].

The four supervised machine learning algorithms were trained on the five features subsets and cross-validated. Random forest models were generated using 100 trees. The support vector machines used linear kernels while artificial neural networks used four-layer networks. To compare the models of the four supervised machine learning algorithms, the study population was randomly stratified and split into 70% training and 30% independent testing data sets.

Performance Evaluation

To evaluate the generated prediction models, we employed various metrics recommended for evaluating classifier performance such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUROC). The multiclass AUROC was the mean of several AUROC classes. Quantitative measures of accuracy and AUROC were used to assess the overall performance of each classifier. AUROC is a measure of model performance, which is based on the receiver operating characteristic curve that plots the tradeoff between sensitivity and specificity of these values using commonly accepted criteria [27].

Results

Description of the RNA Expression Data Set

A total of 7828 primary tumor samples consisting of 32 cancer types were collected from TCGA to construct and test five personal information prediction models (Table 1). To compare tumor samples with normal samples, we extracted data from 593 normal samples from TCGA (Multimedia Appendix 2). The primary tumor data set and normal data set are structurally analogous.

Table 1. Personal information of the study population.

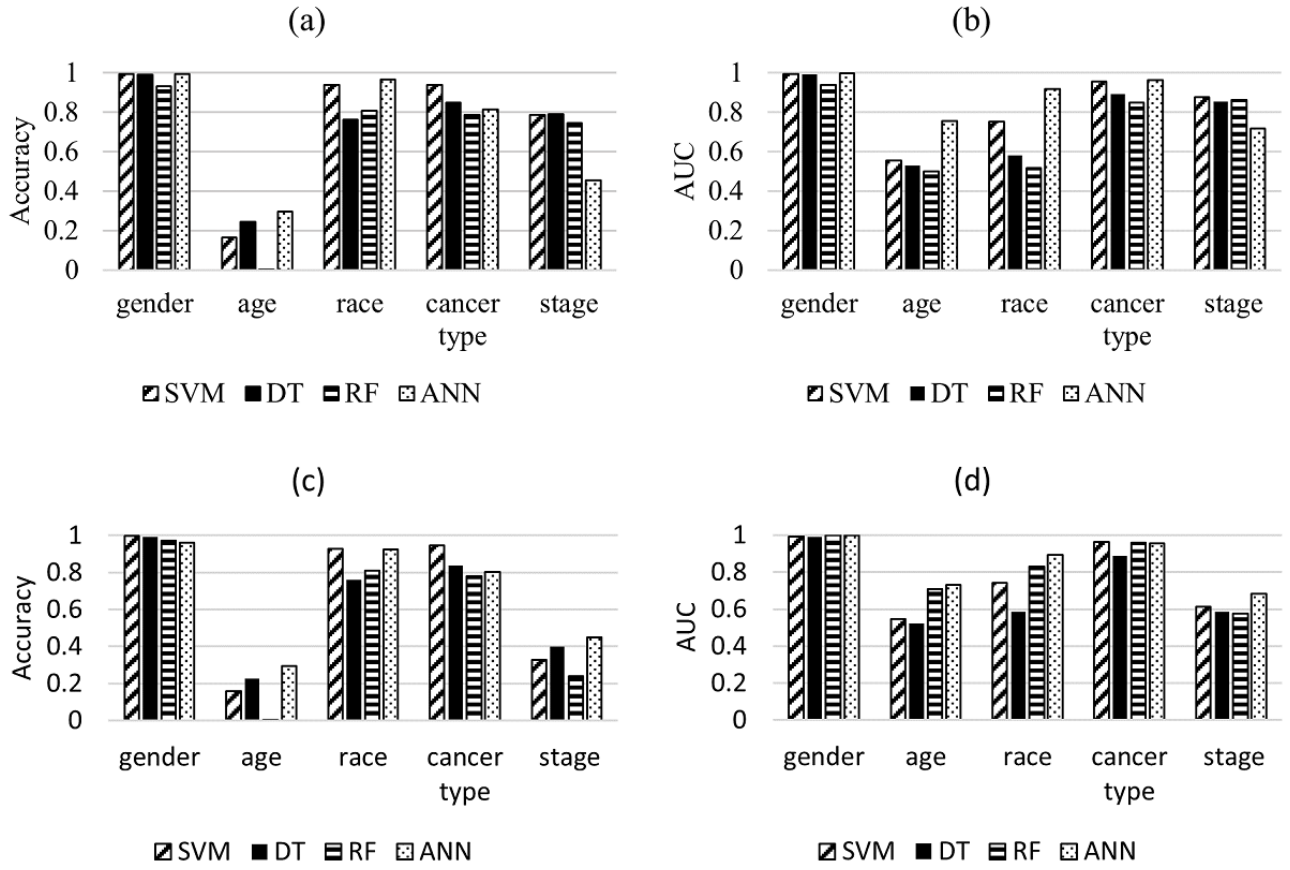
Data set information	Female, n (%)	Male, n (%)	Total, n (%)
Participants	4046 (51.69)	3782 (48.31)	7828 (100)
Age (years)			
10-19	14 (0.18)	16 (0.2)	30 (0.38)
20-29	127 (1.62)	134 (1.71)	261 (3.33)
30-39	330 (4.22)	256 (3.27)	586 (7.49)
40-49	629 (8.04)	451 (5.76)	1080 (13.8)
50-59	946 (12.08)	882 (11.27)	1828 (23.35)
60-69	995 (12.71)	1100 (14.05)	2095 (26.76)
70-79	772 (9.86)	737 (9.41)	1509 (19.27)
80-89	225 (2.87)	200 (2.55)	425 (5.42)
90+	8 (0.1)	6 (0.08)	14 (0.18)
Race			
American or Alaska Native	13 (0.17)	7 (0.09)	20 (0.26)
Asian	253 (3.23)	346 (4.42)	599 (7.65)
Black or African American	519 (6.63)	261 (3.33)	780 (9.96)
Hawaiian or Pacific Islander	6 (0.08)	1 (0.01)	7 (0.09)
Caucasian	3255 (41.58)	3167 (40.46)	6422 (82.04)
Cancer stage[a]			
Stage I	738 (16.8)	638 (14.52)	1376 (31.32)
Stage II	866 (19.71)	552 (12.57)	1418 (32.28)
Stage III	569 (12.95)	491 (11.18)	1060 (24.13)
Stage IV	184 (4.19)	355 (8.08)	539 (12.27)

[a]The percentages may not add up to 100% because of missing values.

Prediction of the Five Personal Variables Using Four Machine Learning Algorithms in Multiple Cancers

Figure 2 presents the performance of the four machine learning algorithms using 12,897 genes to predict target outcomes. The accuracies of the five personal information variables in the independent data set were 0.93-0.99 (gender), 0.0026-0.29 (age), 0.76-0.96 (race), 0.78-0.94 (cancer type), and 0.45-0.79 (stage). The AUROC of the five personal information variables in the independent data set were 0.94-0.99 (gender), 0.50-0.75 (age), 0.51-0.91 (race), 0.84-0.96 (cancer type), and 0.71-0.87 (stage). The accuracy of all models in predicting the five personal information variables ranged variedly, the lowest being in random forest and the highest being in support vector machine (Multimedia Appendix 3).

Figure 2. Prediction performance of personal identifiers according to independent gene sets: (a) accuracy and (b) AUROC of the personal information classifiers from a training dataset consisting of 12,897 genes; (c) accuracy and (d) AUROC of the personal information classifiers as analyzed by prediction models made from a training dataset consisting of significant genes selected through statistical analysis. ANN: artificial neural network; DT: decision tree; RF: random forest; SVM: support vector machine.



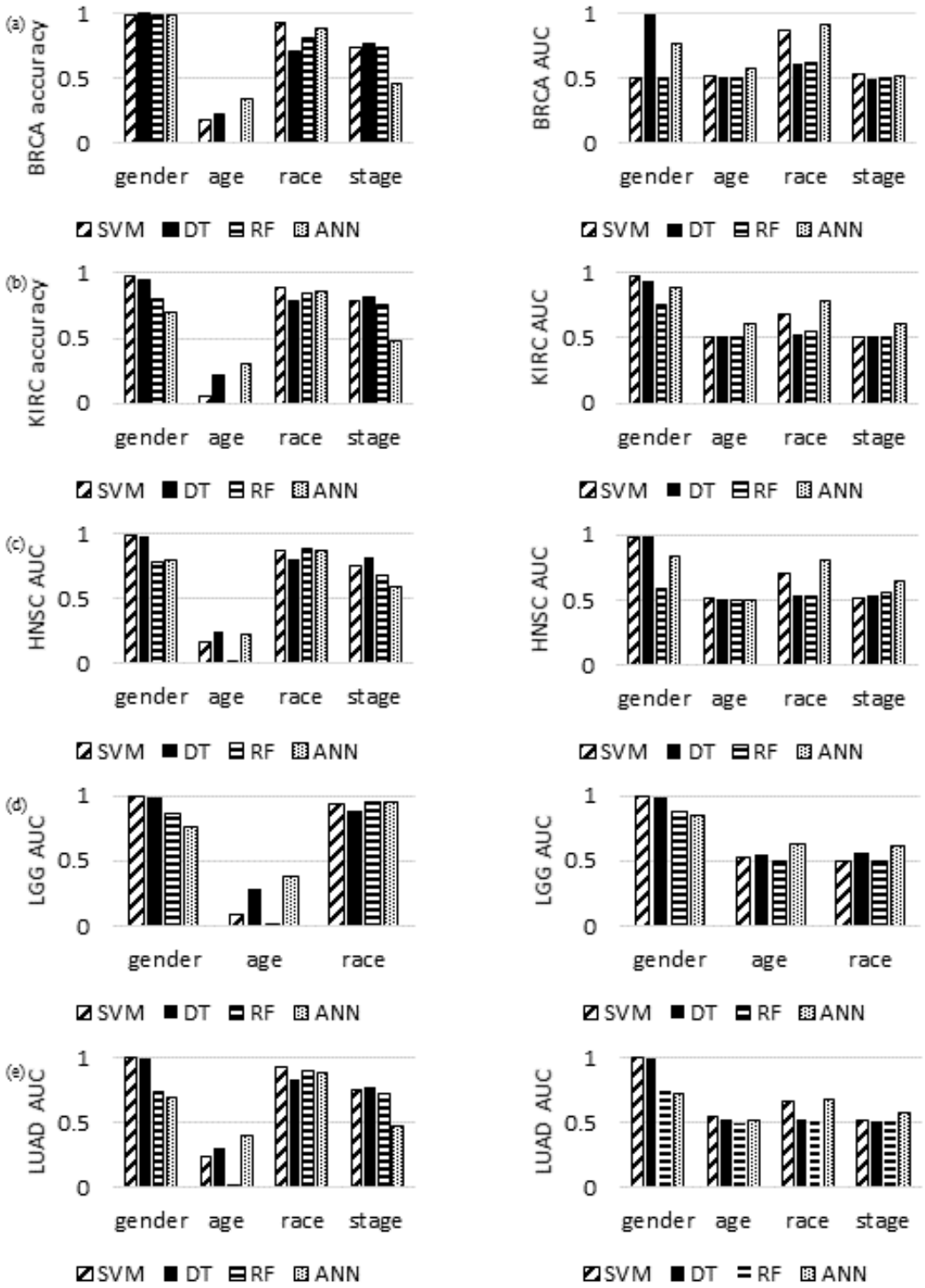
For gender, accuracy and AUROC were the highest compared to other variables: precision and recall ranges were 0.98-1.00 and 0.91-1.00, respectively. We also applied the RNA expression data of the remaining genes (excluding genes on the Y chromosome) using the machine learning algorithms, and the ranges of prediction accuracy and AUROC for gender were 0.91-0.98 and 0.96-0.99, respectively (Multimedia Appendix 4). For age, accuracy and AUROC were low, and precision and recall were less than 0.20. The accuracy and AUROC were lower for race than for gender. In the Caucasian group alone, the F1 scores were high, in the range of 0.86-0.98. However, other groups representing smaller percentages of the population were difficult to predict. Despite the multilayered structure of the 32 cancer types, the accuracy of cancer type and AUROC were lower than those of the gender variable. The accuracy and AUROC of cancer stage were lower than those of the cancer type variable, due to the multilayered structure of the 19 cancer types (Multimedia Appendix 5).

In addition to the gene data sets, results for data sets compiled with significant genes based on the *P* values were obtained. Most results were similar for all gene data sets, with accuracy and AUROC decreasing for all personal information variables, except for the cancer stage variable (Figure 2, Multimedia Appendix 5, and Multimedia Appendix 6). Thus, personal information variables such as age, cancer stage, race, cancer type, and gender were difficult to predict using the prediction models generated by machine learning.

Prediction of Personal Information Using Four Machine Learning Algorithms in the Top Five Cancer Types

We selected five types of cancers with large samples sizes from the primary tumor data set and developed predictive models for the four personal information variables from each cancer data set (Multimedia Appendix 7). To achieve this, we generated data sets for breast carcinoma, kidney renal clear cell carcinoma, head and neck squamous cell carcinoma, low-grade glioma, and lung adenocarcinoma, and predicted the personal information through machine learning algorithms (Figure 3). This analysis compared whether personal information was predicted differently depending on the type of cancer.

Figure 3. Results of model evaluation based on accuracy and AUROC for one type of cancer: (a) breast carcinoma, (b) kidney renal clear cell carcinoma, (c) head and neck squamous cell carcinoma, (d) low-grade glioma, and (e) lung adenocarcinoma. ANN: artificial neural network; DT: decision tree; RF: random forest; SVM: support vector machine.

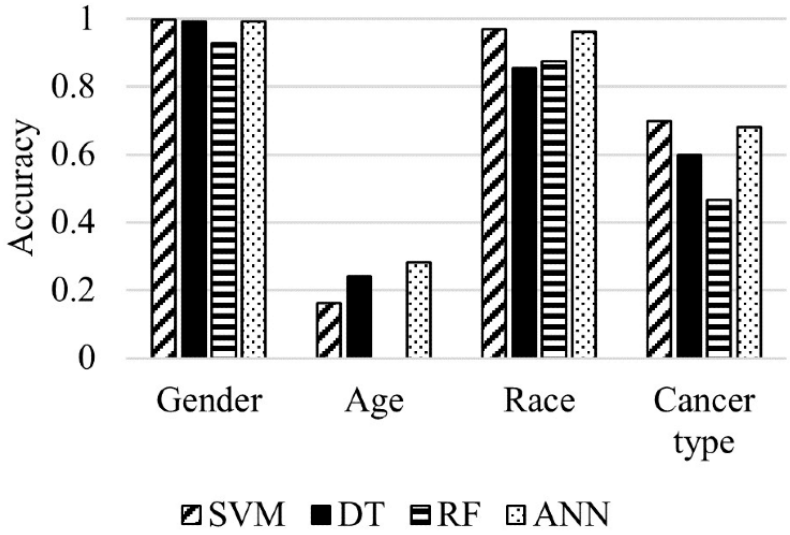


Gender accuracy was low and cancer stage accuracy was high when comparing the average accuracy of all cancers and the five specific cancers; when comparing the accuracy averages of the four machine learning algorithms, the lowest for gender was 0.85, the lowest for age was 0.14, the lowest for race was 0.83, and the lowest for cancer stage was 0.67 (Figure 3, Multimedia Appendix 8).

Validation Through Normal Tissue Samples

We also conducted a study to compare the predicted results for cancer samples with those for normal samples. We evaluated the optimized model using 593 normal tissue samples belonging to cancer patients from TCGA. Prediction of gender, age, and race had similar accuracies in normal samples as in tumor samples, but prediction of cancer type was less accurate in normal samples than prediction of cancer type in tumor samples (Figure 4).

Figure 4. Results of model evaluation based on accuracy and AUROC when applying normal samples as a test to model personal information. ANN: artificial neural network; DT: decision tree; RF: random forest; SVM: support vector machine.



Expression of Specific Genes Depending on Personal Information

Using the RNA expression data of primary tumors, prediction models of the five personal information variables were created using four machine learning algorithms. The genes for predicting personal information were compared using important features provided by random forest and decision tree.

To facilitate the comparison of genes associated with personal information, we compared genes that play an important role when creating models for personal information retrieval. Each of the five personal information variables had a separate list of corresponding important genes; we selected the top 100 genes in these lists, based on the *P* values generated by random forest and decision tree. The results showed a low level of association between genes and personal information: only gender showed gene-relatedness in random forest. Of the top 10 genes in the list of important genes to predict the gender provided by random forest, 9 genes were located on the Y chromosome. However, of the top 10 genes in the list of important genes to predict the gender provided by decision tree, only 2 genes were located on the Y chromosome.

Discussion

Principal Results

The prediction of personal information using RNA expression–based approaches is a rapidly developing subfield of cancer epigenetics that has great potential to provide accurate predictive outcomes. However, there is a risk that patient information might be disclosed, thereby limiting data collection and sharing. For accurate regulation of the use of genomic data, it is necessary to examine the possibility of private information leaking. In this study, we verified personal information predictability using RNA expression profiling and presented a new direction for studies on genomic data sharing.

When predicting personal information from RNA expression data using the four machine learning algorithms, we found that most personal information could not be predicted using RNA expression data, with the exceptions being gender and cancer type. Gender could be easily predicted by analyzing the expression of sex chromosome genes, as expected [28]. Furthermore, we confirmed whether gender could be predicted by RNA expression data of the remaining genes (excluding the genes on the Y chromosome) through machine learning. Since there is a regulatory network between genes located on chromosome X or Y and chromosome 1-22, gender could still be predicted by genes located on chromosomes 1-22 [29]. In addition, several studies have shown that there are marked differences in RNA expression between different cancer types [30-32]. Personal information can be more accurately predicted when combined with other information [33], which slightly increased the accuracy of the prediction when information about cancer type was provided (increased by 0.004-0.14). However, it is still difficult to say that personal information, such as age, race, and cancer stage, could be predicted.

Aging is a very complex process that is influenced by various genetic, lifestyle, and environmental factors [34]. It causes a variety of molecular modifications and adjustments in tissues and organs that accumulate over an individual's lifetime, including chemical modifications and alterations to gene expression [35]. Thus, it is difficult to predict age from RNA expression data alone, because of the influence of these molecular modifications and adjustments throughout the individual's lifetime. Although HIPAA describes birth date as an identifiable variable, it is an identifier virtually impossible to predict using RNA expression data. To predict a patient's age, epigenomic data such as DNA methylation profiles, which reflects the patient's age, should be used [34,35].

The race data set was not suitable for the machine learning algorithms because the sample ratios between group populations were unbalanced. Caucasians accounted for 82% of patient samples in the RNA expression data set. For this reason, the accuracy and recall results were higher for the Caucasian group and lower for the other groups. Machine learning algorithms can suffer a performance bias in relation to classification when data sets are unbalanced [36]. On the other hand, with regard to the cancer stage data set, predictions had low accuracy across

all machine learning platforms, findings that coincide with those of other studies [37,38].

Genomic data are sensitive data that can be used to identify individuals or for other purposes [39,40]. The individuality of some of the genomic data has been verified [34,35,41], but little has been studied for RNA data. Our study is thus the first to verify individual identification in RNA expression data testing the identification of personal information using RNA expression data provided by the public database TCGA. The prediction was rarely successful.

This study is the first paper to suggest determining the level of data sharing should be based on predictability of personal information. Clearly, data sharing will have a pivotal role in precision oncology. By promoting cancer genomic data sharing, researchers and clinicians will gather the information needed to improve our understanding of cancer genome and improve patient care and outcomes. Currently, projects that use genomic data use complex procedures to collect the genomic data or use previously published genomic data. Genomic data collected for a given project is often difficult to reuse in other projects. We believe that studying whether the personal information of the patient can be predicted when genomic data is shared can help determine the appropriate level of genome data sharing.

Limitations

Our study has several limitations. First, the study was limited in terms of scope of personal information studied; we used RNA expression data and machine learning algorithms to perform predictions regarding only five personal information variables. Limited information to predict personal characteristics can be retrieved from the available TCGA data. Thus, large-scale studies employing personal information are needed to gain further insights in this field. Second, the race data were unbalanced in the TCGA primary tumor data set, with predominantly Caucasian data. Therefore, the predictive ability of the machine learning algorithms in relation to race might be reduced or biased. To clarify whether RNA expression data can predict race, further research using genomic data from diverse and balanced races is needed. Third, this study used only RNA expression data to assess the predictability of personal information using machine learning algorithms. Future research should explore the possibility of predicting personal information using other genomic information, such as DNA methylation data. Fourth, this study was done without considering clinical data, because there was little clinical information in the TCGA public database. Further research will be needed considering clinical data is sensitive information.

Conclusions

In this study, we analyzed the ability of RNA expression data to predict patients' personal information using machine learning algorithms. We verified that RNA expression alone is not sufficient to identify personal information using the analysis techniques employed herein. These tentative conclusions await further validation by future similar studies.

Acknowledgments

We gratefully acknowledge the TCGA Consortium and all its members for the TCGA Project initiative, for providing sample, tissues, and data processing and making data and results available. This work was supported by the Technology Innovation Program (20002289), funded by the Ministry of Trade, Industry & Energy (Republic of Korea) and by a faculty research grant from Yonsei University College of Medicine (6-20190115).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Three data sets with *t* test and ANOVA.
[PDF File (Adobe PDF File), 29 KB - jmir_v22i8e18387_app1.pdf]

Multimedia Appendix 2
Normal samples of the four personal information in TCGA.
[PDF File (Adobe PDF File), 376 KB - jmir_v22i8e18387_app2.pdf]

Multimedia Appendix 3
Receiver operating characteristics curve of support vector machine, decision tree, random forest, artificial neural network for predicting Personal Healthcare Identifiers. (a) gender, (b) age, (c) race, (d) cancer type, and (e) stage on the primary tumor test-set.
[PNG File , 157 KB - jmir_v22i8e18387_app3.png]

Multimedia Appendix 4
Prediction performance of gender according to an independent gene set consisting of genes located on chromosomes 1-22 and X.
[PDF File (Adobe PDF File), 111 KB - jmir_v22i8e18387_app4.pdf]

Multimedia Appendix 5
Prediction performance of personal information according to independent gene sets consisting of 12,897 genes.
[PDF File (Adobe PDF File), 123 KB - jmir_v22i8e18387_app5.pdf]

Multimedia Appendix 6
Prediction performances of personal information according to independent gene sets consisting of significant genes selected through statistical analysis.
[PDF File (Adobe PDF File), 172 KB - jmir_v22i8e18387_app6.pdf]

Multimedia Appendix 7
Five types of cancers with large sample sizes from the primary tumor dataset.
[PDF File (Adobe PDF File), 143 KB - jmir_v22i8e18387_app7.pdf]

Multimedia Appendix 8
Statistical analysis of predictive performance of personal information for five cancer types.
[PDF File (Adobe PDF File), 356 KB - jmir_v22i8e18387_app8.pdf]

References

1. Bielinski SJ, Olson JE, Pathak J, Weinshilboum RM, Wang L, Lyke KJ, et al. Preemptive Genotyping for Personalized Medicine: Design of the Right Drug, Right Dose, Right Time—Using Genomic Data to Individualize Treatment Protocol. Mayo Clinic Proceedings 2014 Jan;89(1):25-33. [doi: 10.1016/j.mayocp.2013.10.021]
2. Shabani M, Borry P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. Eur J Hum Genet 2017 Nov 29;26(2):149-156. [doi: 10.1038/s41431-017-0045-7]
3. Office FCRH. Standards for privacy of individually identifiable health information. Final rule. Federal register 2002;67(157):53181. [doi: 10.1201/9781420000177.ch38]
4. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn) 2015;19(1A):A68-A77 [FREE Full text] [doi: 10.5114/wo.2014.47136] [Medline: 25691825]
5. Naveed M, Ayday E, Clayton EW, Fellay J, Gunter CA, Hubaux J, et al. Privacy in the Genomic Era. ACM Comput. Surv 2015 Sep 29;48(1):1-44. [doi: 10.1145/2767007]
6. Cariaso M, Lennon G. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. Nucleic Acids Res 2012;40(D1):a. [doi: 10.1093/nar/gkr798]
7. Badalato L, Kalokairinou L, Borry P. Third party interpretation of raw genetic data: an ethical exploration. Eur J Hum Genet 2017 Aug 23;25(11):1189-1194. [doi: 10.1038/ejhg.2017.126]
8. Marguerat S, Bähler J. RNA-seq: from technology to biology. Cell. Mol. Life Sci 2009 Oct 27;67(4):569-579. [doi: 10.1007/s00018-009-0180-6]
9. Lunter G, Goodson M. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Research 2010 Oct 27;21(6):936-939. [doi: 10.1101/gr.111120.110]
10. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009 Jan;10(1):57-63. [doi: 10.1038/nrg2484]
11. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nat Rev Genet 2015 May 7;16(6):321-332. [doi: 10.1038/nrg3920]
12. TCGA Research Network. The Cancer Genome Atlas. URL: http://cancergenome.nih.gov [accessed 2020-07-28]
13. FireBrowse. URL: https://gdac.broadinstitute.org [accessed 2020-07-28]
14. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008 May 30;5(7):621-628. [doi: 10.1038/nmeth.1226]
15. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011 Aug 4;12(1):323. [doi: 10.1186/1471-2105-12-323]
16. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010 Jan 01;26(1):139-140 [FREE Full text] [doi: 10.1093/bioinformatics/btp616] [Medline: 19910308]
17. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 2014 Feb 03;15(2):R29 [FREE Full text] [doi: 10.1186/gb-2014-15-2-r29] [Medline: 24485249]
18. Fisher Wilson J. Health Insurance Portability and Accountability Act Privacy Rule Causes Ongoing Concerns among Clinicians and Researchers. Ann Intern Med 2006 Aug 15;145(4):313. [doi: 10.7326/0003-4819-145-4-200608150-00019]
19. Long AD, Mangalam HJ, Chan BYP, Tolleri L, Hatfield GW, Baldi P. Improved Statistical Inference from DNA Microarray Data Using Analysis of Variance and A Bayesian Statistical Framework. J. Biol. Chem 2001 Mar 20;276(23):19937-19944. [doi: 10.1074/jbc.m010192200]

20. Baldi P, Long AD. A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes. Bioinformatics 2001 Jun 01;17(6):509-519. [doi: 10.1093/bioinformatics/17.6.509]
21. Kerr MK, Martin M, Churchill GA. Analysis of Variance for Gene Expression Microarray Data. Journal of Computational Biology 2000 Dec;7(6):819-837. [doi: 10.1089/10665270050514954]
22. Bharathi A, Natarajan A. Cancer Classification of Bioinformatics datausing ANOVA. IJCTE 2010:369-373. [doi: 10.7763/ijcte.2010.v2.169]
23. Mao Y, Zhou X, Pi D, Sun Y, Wong STC. Multiclass Cancer Classification by Using Fuzzy Support Vector Machine and Binary Decision Tree With Gene Selection. Journal of Biomedicine and Biotechnology 2005;2005(2):160-171. [doi: 10.1155/jbb.2005.160]
24. Brown MPS, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proceedings of the National Academy of Sciences 2000 Jan 04;97(1):262-267. [doi: 10.1073/pnas.97.1.262]
25. Pirooznia M, Yang JY, Yang MQ, Deng Y. A comparative study of different machine learning methods on microarray gene expression data. BMC Genomics 2008;9(Suppl 1):S13. [doi: 10.1186/1471-2164-9-s1-s13]
26. Díaz-Uriarte R, Alvarez de Andrés S. Gene selection and classification of microarray data using random forest. BMC Bioinformatics 2006 Jan 06;7:3. [doi: 10.1186/1471-2105-7-3] [Medline: 16398926]
27. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8-17 [FREE Full text] [doi: 10.1016/j.csbj.2014.11.005] [Medline: 25750696]
28. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, et al. Individuality and variation in gene expression patterns in human blood. Proceedings of the National Academy of Sciences 2003 Feb 10;100(4):1896-1901. [doi: 10.1073/pnas.252784499]
29. Srivastava A, Wang J, Zhou H, Melvin JE, Wong DT. Age and gender related differences in human parotid gland gene expression. Archives of Oral Biology 2008 Nov;53(11):1058-1070. [doi: 10.1016/j.archoralbio.2008.05.006]
30. Özdağ H, Teschendorff AE, Ahmed AA, Hyland SJ, Blenkiron C, Bobrow L, et al. Differential expression of selected histone modifier genes in human solid cancers. BMC Genomics 2006 Apr 25;7(1):90. [doi: 10.1186/1471-2164-7-90]
31. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature 2005 Jun;435(7043):834-838. [doi: 10.1038/nature03702]
32. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang C, Angelo M, et al. Multiclass cancer diagnosis using tumor gene expression signatures. Proceedings of the National Academy of Sciences 2001 Dec 11;98(26):15149-15154. [doi: 10.1073/pnas.211566398]
33. Jung J, Park P, Lee J, Lee H, Lee GK, Cha HS. A Determination Scheme for Quasi-Identifiers Using Uniqueness and Influence for De-Identification of Clinical Data. j med imaging hlth inform 2020 Feb 01;10(2):295-303. [doi: 10.1166/jmihi.2020.2966]
34. Vidaki A, Ballard D, Aliferi A, Miller TH, Barron LP, Syndercombe Court D. DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. Forensic Science International: Genetics 2017 May;28:225-236. [doi: 10.1016/j.fsigen.2017.02.009]
35. Horvath S. DNA methylation age of human tissues and cell types. Genome Biology 2013;14(10):R115. [doi: 10.1186/gb-2013-14-10-r115]
36. Chawla NV, Japkowicz N, Kotcz A. Editorial. SIGKDD Explor. Newsl 2004 Jun;6(1):1-6. [doi: 10.1145/1007730.1007733]
37. Bhalla S, Chaudhary K, Kumar R, Sehgal M, Kaur H, Sharma S, et al. Gene expression-based biomarkers for discriminating early and late stage of clear cell renal cancer. Sci Rep 2017 Mar 28;7(1):1-13. [doi: 10.1038/srep44997]
38. Rahimi A, Gönen M. Discriminating early- and late-stage cancers using multiple kernel learning on gene sets. Bioinformatics 2018 Jul 01;34(13):i412-i421 [FREE Full text] [doi: 10.1093/bioinformatics/bty239] [Medline: 29949993]
39. Anderlik MR, Rothstein MA. P C G I : What Rules for the New Science? Annu. Rev. Genom. Hum. Genet 2001 Sep;2(1):401-433. [doi: 10.1146/annurev.genom.2.1.401]
40. Greenbaum D, Sboner A, Mu XJ, Gerstein M. Genomics and Privacy: Implications of the New Reality of Closed Data for the Field. PLoS Comput Biol 2011 Dec 1;7(12):e1002278. [doi: 10.1371/journal.pcbi.1002278]
41. Kang S, Li Q, Chen Q, Zhou Y, Park S, Lee G, et al. CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. Genome Biol 2017 Mar 24;18(1):1-12. [doi: 10.1186/s13059-017-1191-5]

Abbreviations

AUROC: area under the receiver operating characteristic curve
HIPAA: Health Insurance Portability And Accountability Act
TCGA: The Cancer Genome Atlas.

Edited by G Eysenbach; submitted 28.02.20; peer-reviewed by SY Shin, H Soung; comments to author 10.03.20; revised version received 25.03.20; accepted 06.07.20; published 10.08.20.

Please cite as:
Kweon S, Lee JH, Lee Y, Park YR
Personal Health Information Inference Using Machine Learning on RNA Expression Data from Patients With Cancer: Algorithm Validation Study
J Med Internet Res 2020;22(8):e18387
URL: https://www.jmir.org/2020/8/e18387
doi:10.2196/18387
PMID:32773372

Original Paper

Characteristics and Outcomes of a Sample of Patients With COVID-19 Identified Through Social Media in Wuhan, China: Observational Study

Dong Liu[1], PhD; Yuyan Wang[2*], MD; Juan Wang[3*], MD; Jue Liu[4*], MD; Yongjie Yue[1*], BA; Wenjun Liu[1], PhD; Fuhai Zhang[5*], PhD; Ziping Wang[2], MD

[1]Renmin University of China, Beijing, China

[2]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Thoracic Medical Oncology, Peking University Cancer Hospital & Institute, Beijing, China

[3]Department of Radiology, Peking University Shougang Hospital, Beijing, China

[4]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China

[5]School of Education, Hebei Normal University, Shijiazhuang, China

[*]these authors contributed equally

Corresponding Author:
Ziping Wang, MD
Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing)
Department of Thoracic Medical Oncology
Peking University Cancer Hospital & Institute
No 52 Fucheng Road
Haidian District
Beijing
China
Phone: 86 13301212676
Fax: 86 10 87787471
Email: wangzp2007@126.com

Abstract

Background: The number of deaths worldwide caused by coronavirus disease (COVID-19) is increasing rapidly. Information about the clinical characteristics of patients with COVID-19 who were not admitted to hospital is limited. Some risk factors of mortality associated with COVID-19 are controversial (eg, smoking). Moreover, the impact of city closure on mortality and admission rates is unknown.

Objective: The aim of this study was to explore the risk factors of mortality associated with COVID-19 infection among a sample of patients in Wuhan whose conditions were reported on social media.

Methods: We enrolled 599 patients with COVID-19 from 67 hospitals in Wuhan in the study; 117 of the participants (19.5%) were not admitted to hospital. The demographic, epidemiological, clinical, and radiological features of the patients were extracted from their social media posts and coded. Telephone follow-up was conducted 1 month later (between March 15 and 23, 2020) to check the clinical outcomes of the patients and acquire other relevant information.

Results: The median age of patients with COVID-19 who died (72 years, IQR 66.5-82.0) was significantly higher than that of patients who recovered (61 years, IQR 53-69, $P<.001$). We found that lack of admission to hospital (odds ratio [OR] 5.82, 95% CI 3.36-10.1; $P<.001$), older age (OR 1.08, 95% CI 1.06-1.1; $P<.001$), diffuse distribution (OR 11.09, 95% CI 0.93-132.9; $P=.058$), and hypoxemia (odds ratio 2.94, 95% CI 1.32-6.6; $P=.009$) were associated with increasing odds of death. Smoking was not significantly associated with mortality risk (OR 0.9, 95% CI 0.44-1.85; $P=.78$).

Conclusions: Older age, diffuse distribution, and hypoxemia are factors that can help clinicians identify patients with COVID-19 who have poor prognosis. Our study suggests that aggregated data from social media can also be comprehensive, immediate, and informative in disease prognosis.

(J Med Internet Res 2020;22(8):e20108) doi:10.2196/20108

KEYWORDS

COVID-19; risk factors; web-based data; outcome; infectious disease; clinical characteristic; mortality; social media; prognosis, China; coronavirus

Introduction

In December 2019, a novel coronavirus disease (COVID-19) emerged in China and began to spread globally. As of April 08, 2020, the outbreak has resulted in 82,992 deaths worldwide [1]. Identification of early-stage clinical predictors of poor patient outcomes is essential to effectively prioritize resources for patients with the highest risks and lower death rates. Several case series from China and Italy have suggested that male sex, older age, hypertension, kidney disease, and myocardial injury are risk factors for severe COVID-19 [2-6]. Despite these studies, knowledge of the early stage risk factors associated with poor prognosis is still limited. In this paper, we present a series of cases reported on the internet (from 67 hospitals in Wuhan) with definite clinical outcomes (discharge or death as of March 30, 2020) and their early-stage characteristics (before hospital admission) to explore the early stage risk factors and clinical features of COVID-19 mortality.

Methods

Study Design

The study protocol was approved by the research ethics committee of Renmin University of China on February 5, 2020. Data were obtained from two sources: Weibo posts and a telephone survey. The Weibo data were posted on the internet by families impacted by COVID-19 between January 20 and February 15, 2020, and were collected between February 3 and February 15, 2020. Then, volunteers phoned each participant's family to describe the study and obtain their oral consent to participate. Over 60% of the patients (599/911, 65.8%) agreed to participate and completed most of the questions. One month later, a follow-up telephone call was conducted to collect the outcomes of each patient (between March 15 and 23, 2020).

Data Sources

In this study, we used patient reports from Weibo to conduct the analysis. Weibo is a Chinese microblogging website that resembles a hybrid of Twitter and Facebook; it uses a format similar to that of Twitter. This microblog provides a platform for patients infected with COVID-19 in Wuhan to seek help on the internet. Many patients reported their onset of symptoms, listed their symptoms, and uploaded their medical records and computed tomography (CT) images to seek medical care on the internet. We carefully monitored reports of patients infected with COVID-19 on Weibo between January 20 and February 15, 2020, and downloaded the data. We extracted and coded this information and deidentified the patient information by removing their names, home addresses, and contact information. We obtained 911 original COVID-19 patient reports. We only included patients who were diagnosed by positive COVID-19 tests according to the guidance provided by the Chinese National Health Commission. The positive tests consisted of either CT or real-time reverse transcriptase–polymerase chain reaction (RT-PCR) reports on the internet and were further confirmed by the follow-up telephone call.

The Weibo messages were posted from February 3 to 15, 2020. The median time from onset of symptoms to the posting of messages was 7 days (IQR 3-11). We conducted telephone follow-up calls 1 month later (between March 15 and 23) to check the clinical outcomes of the patients and acquire information about their smoking behavior, admission time, time to discharge or death, duration of time in the hospital, name of the hospital, and medication used for hypertension. Only patients who had a definite outcome (died or recovered) were included in our study.

Data Coding

All Weibo messages were collected on the internet from February 3 to February 15, 2020. The research team extracted vital information from individual patients' reports on the internet. A series of individual-level patient data, including demographic information, underlying comorbidities, symptoms, and signs, were coded. We double-checked and reviewed the data. The data were entered into a computerized database and cross-checked.

The symptoms coded included hypoxemia, inability to eat, cough, acute respiratory distress (ARD), dizziness, headache, confusion, unconsciousness, hemoptysis, chest pain, muscle pain, fatigue, vomiting, chest distress, loss of appetite, diarrhea, shortness of breath, and fever. The underlying comorbidities coded included chronic respiratory disease, chronic liver disease, hypertension, diabetes, chronic vascular disease, chronic lung disease, chronic heart disease, cancer, and kidney disease. Other information coded from Weibo messages included symptom onset date, sex, and age. In the telephone survey, we specifically asked which hypertension medications the patients were taking. We coded the type of medication as angiotensin-converting enzyme inhibitors (ACEIs), angiotensin II receptor blockers (ARBs), and others.

All CT images in the original posts were extracted and recorded. The CT images and CT reports were evaluated by an experienced radiologist (WJ) who was blinded to the patient survival results when she interpreted the images. The radiologist examined and coded the following features: lesion distribution, lesion characteristics, and pleural effusion.

Our data have been made public so that readers can replicate our analysis. The data can be found in the supplemental materials of this paper (Multimedia Appendix 1).

Statistical Analysis

Descriptive statistics were obtained for all study variables. Categorical variables were described as frequency rates and percentages and were compared for the outcomes of the study using the Fisher exact test. The continuous variables were described using the median, range, and SD values and were compared using *t* tests.

Cumulative rates of death were determined using the Kaplan-Meier method. The associations between age groups, hospital admission status, lesion distribution, and pleural effusion and death outcomes were examined using the Cox proportional hazard regression model. We used the Kaplan-Meier method to plot the survival curves and used multivariate Cox regression to determine the independent risk factors for mortality.

All statistical tests were 2-tailed. *P* values <.05 were considered significant. All analyses were performed in R (R Foundation).

Results

Key Time-Course Distribution

We enrolled 599 patients in our study; of these patients, 516 (86.1%) recovered, and 83 (14.9%) died.

The median time from symptom onset to discharge was much longer than the time to death, namely 36 days (IQR 29.0–44.0) and 14 days (IQR 9.0-20.0, *P*<.001), respectively. The median time from the onset of symptoms to hospital admission was 11 days (IQR 7.0-15.0) for the discharged patients, compared with 10 days (IQR 6.0-13.0) for the patients who died (*P*=.70). The median time of hospital stay was 25 days (IQR 18.0-32.0) for the discharged patients and was significantly shorter for the patients who died (6 days, IQR 3.0-12.0; *P*<.001). The median time from onset of symptoms to death for patients without hospital admission was 13 days (IQR 7.0-16.75), which was not significantly different from that of patients who died in hospital (*P*=.34).

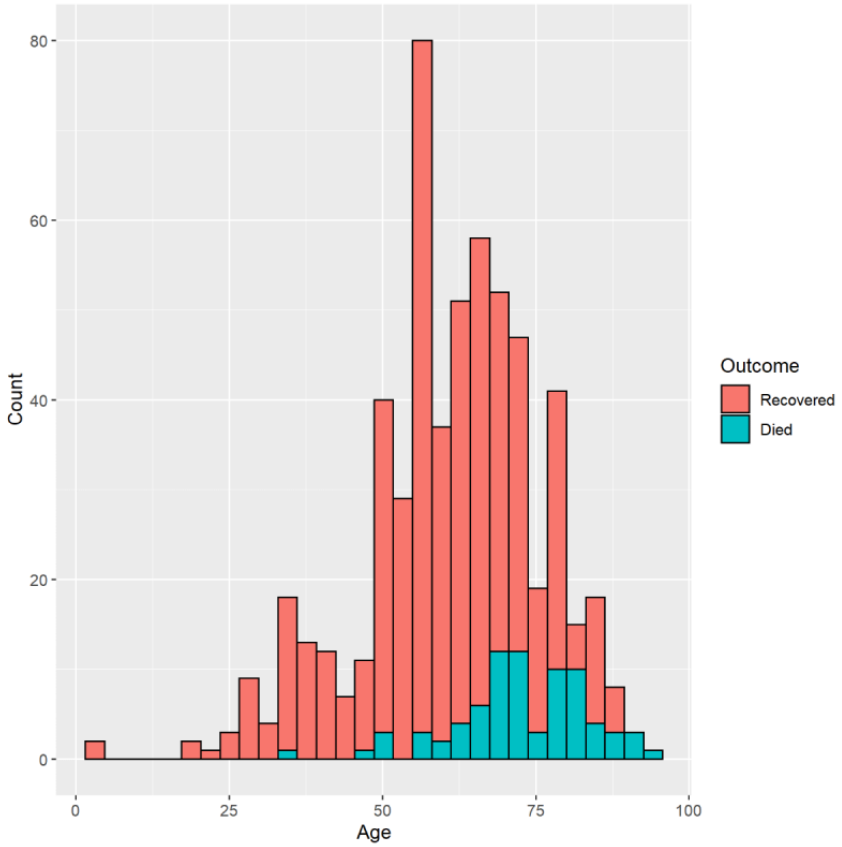
Age and Gender

The median age of the deceased patients (72 years, IQR 66.5-82.0) was significantly higher than that of the recovered patients (61 years, IQR 53-69, *P*<.001); see Table 1 and Figure 1. Female sex was more prevalent among patients who recovered (243/516, 47.1%) than among those who died (31/83, 37.3%, *P*=.05); see Table 1.

Table 1. Baseline characteristics of patients with coronavirus disease (N=599).

Characteristic	Patient outcomes			*P* value
	All (N=599)	Recovered (n=511)	Died (n=83)	
Age (years), median (range)	63 (2-93)	61 (2-89)	72 (33-93)	<.001
Female sex, n (%)	274 (46.1)	243 (47.6)	31 (37.3)	.10
Smoker, n (%)	123 (22.0)	106 (21.6)	17 (24.6)	.54
Signs and symptoms before admission, n (%)				
Fever	471 (79.8)	408 (80.3)	63 (76.8)	.46
Cough	241 (40.8)	216 (42.5)	25 (30.5)	.04
Hemoptysis	11 (1.9)	10 (2.0)	1 (1.2)	>.99
Dyspnea	291 (49.3)	244 (48.0)	47 (57.3)	.12
Shortness of breath	74 (12.5)	68 (13.4)	6 (7.3)	.15
Fatigue	199 (33.7)	175 (34.4)	24 (29.3)	.38
Muscle ache	19 (3.2)	19 (3.7)	0 (0)	.09
Diarrhea	80 (13.6)	73 (14.4)	7 (8.5)	.17
Chest pain	13 (2.2)	12 (2.4)	1 (1.2)	>.99
Vomiting	72 (12.2)	67 (13.2)	5 (6.1)	.07
Chest distress	64 (10.8)	61 (12.0)	3 (3.7)	.02
Loss of appetite	156 (26.4)	130 (25.6)	26 (31.7)	.28
Inability to eat	24 (4.0)	18 (3.5)	6 (7.2)	.13
Hypoxemia	33 (5.6)	23 (4.5)	10 (12.2)	.02
Confusion	17 (2.9)	12 (2.4)	5 (6.1)	.07
Unconsciousness	15 (2.5)	12 (2.4)	3 (3.7)	.45
Dizziness	18 (3.1)	16 (2.4)	2 (3.1)	>.99
Headache	22 (3.7)	20 (3.9)	2 (2.4)	.76
Underlying illness, n (%)				
Hypertension	87 (14.7)	79 (15.5)	8 (9.8)	.24
Diabetes	57 (9.6)	49 (9.6)	8 (9.8)	>.99
Chronic heart disease	70 (11.8)	61 (12.0)	9 (11.0)	.49
Chronic lung disease	25 (4.1)	22 (4.3)	3 (3.7)	>.99
Cerebrovascular disease	15 (2.5)	13 (2.5)	2 (2.4)	.95
Chronic kidney disease	28 (4.7)	26 (5.1)	2 (2.4)	.41
Chronic liver disease	14 (2.7)	0 (0)	14 (2.4)	.13
Chronic respiratory disease	12 (2.0)	11 (2.2)	1 (1.2)	>.99
Cancer	17 (2.9)	16 (3.1)	1 (1.2)	.49

Figure 1. Age distributions of patients with coronavirus disease in our study who recovered and who died.



Symptoms and Clinical Characteristics

At baseline, the most common symptoms were fever (471/599, 79.8%) and cough (241/599, 40.8%); see Table 1. According to the Fisher exact test, cough and chest distress were less frequent among patients who died (25/83, 30.5%, and 3/83, 3.7%, respectively) than among recovered patients (216/511, 42.5%, and 61/511, 12.0%, respectively). The incidence rates of muscle ache and vomiting were similar between patients who died and who recovered; these differences were only marginally significant (*P*=.09 and *P*=.07, respectively). On the other hand, hypoxemia and confusion were more frequent among patients who died (10/83, 12.2%, and 5/83, 6.1%) than among patients who recovered (23/511, 4.5%, and 12/511, 2.4%; *P*=.02 and .07, respectively).

Univariate Cox regression analysis showed that age older than 70 years (*P*=.02 for patients aged 70 to 79 years; *P*=.002 for patients aged ≥80 years) and hypoxemia (*P*=.03) were positively associated with death. Kaplan-Meier analysis in different groups using a log-rank test also revealed a significantly higher mortality rate for patients aged 70 to 79 years and ≥80 years and patients with hypoxemia (*P*<.001 and *P*=.02, respectively). Additionally, both univariate Cox regression and Kaplan-Meier analysis showed a lower risk of death in patients admitted to hospital (*P*<.001 for both analyses); see Table 2, Figure 2, and Figure 3.

The multivariable-adjusted Cox proportional hazard regression model showed a significantly lower risk of death in patients with hypertension when controlling for age and sex (*P*=.008); see Figure 4. The number of days between symptom onset and hospital admission was significantly shorter in the deceased patients (*P*=.02); see Table 1 and Figure 5. Here, we adjusted the model using only age and sex because all other covariates, including symptoms and underlying diseases, were not significant and were excluded from the final model.

Table 2. Multivariate Cox regression analysis of the risk factors associated with mortality in patients with coronavirus disease (n=83).

Factor	Death (%)	Univariate model		Multivariate model	
		Crude hazard ratio (95% CI)	*P* value	Adjusted hazard ratio (95% CI)	*P* value
Age (years)					
40-59	9 (4.6)	1.28 (0.15-10.93)	.82	1.82 (0.21-15.78)	.59
60-69	22 (12.5)	5.19 (0.69-39.04)	.11	8.27 (1.07-63.63)	.04
70-79	25 (23.4)	11.7 (1.58-86.87)	.02	14.01 (1.85-105.83)	.01
≥80	21 (46.7)	22.92 (3.01-174.35)	.002	36.14 (4.68-279.27)	.001
Female sex	31 (11.3)	0.74 (0.44-1.25)	.26	0.71 (0.40-1.26)	.24
Hospital admission	33 (7.0)	0.20 (0.12-0.34)	<.001	0.16 (0.093-0.28)	<.001
Hypertension	8 (9.2)	0.80 (0.38-1.69)	.57	0.24 (0.08-0.67)	.006
Hypoxemia	10 (30.3)	2.45 (1.11-5.38)	.03	3.39 (1.51-7.62)	.003

Figure 2. Cumulative incidence of death of patients with coronavirus disease grouped by (a) hospital admission, (b) time length between symptom onset and hospital admission, (c) hypoxemia, and (d) hypertension.

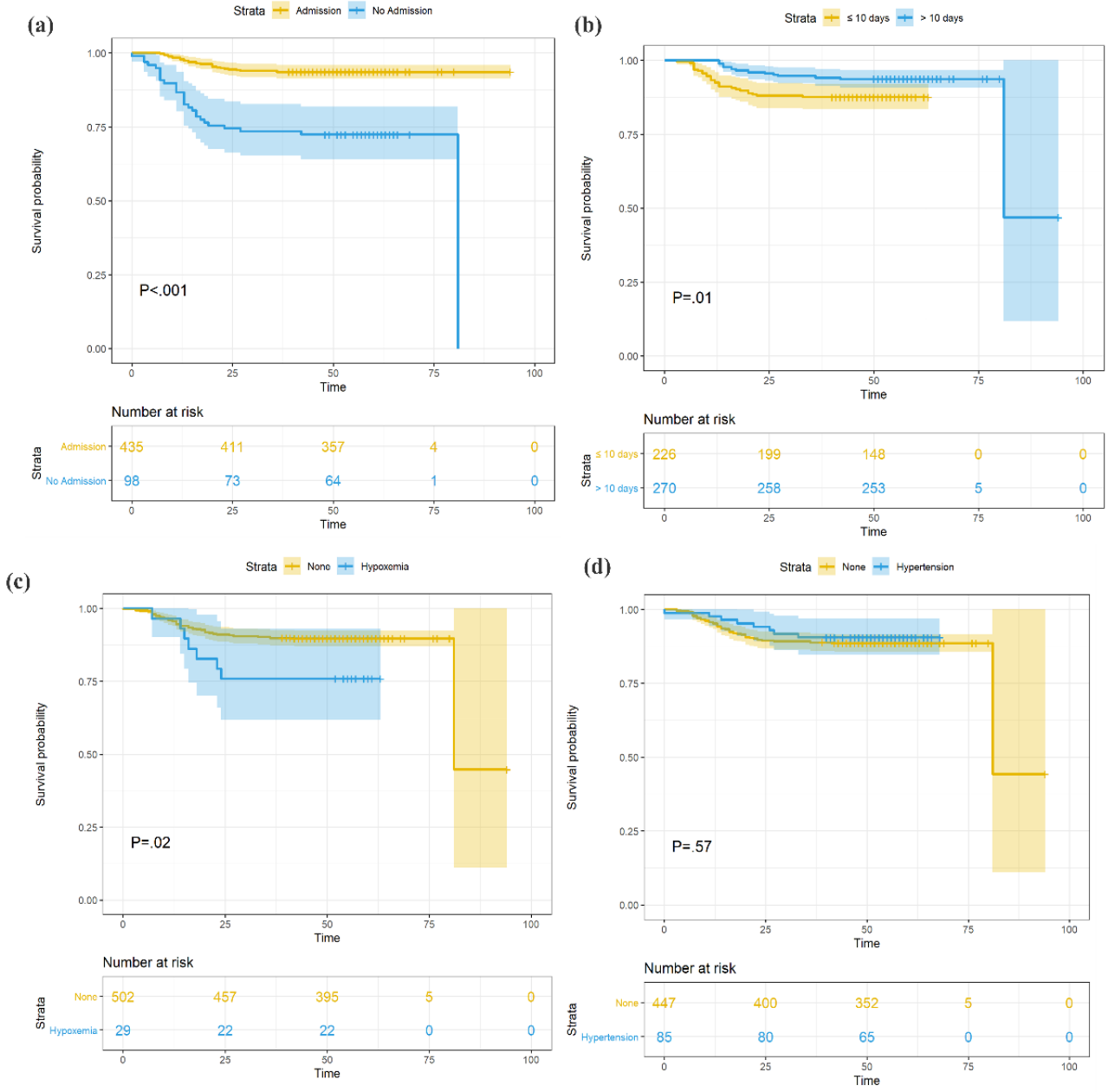


Figure 3. Cumulative incidence of death of patients with coronavirus disease grouped by (a) age group and (b) hospital admission and severity of illness. AC: admission, critically ill; AN: admission, not critically ill; NC: no admission, critically ill; NN: no admission, not critically ill.

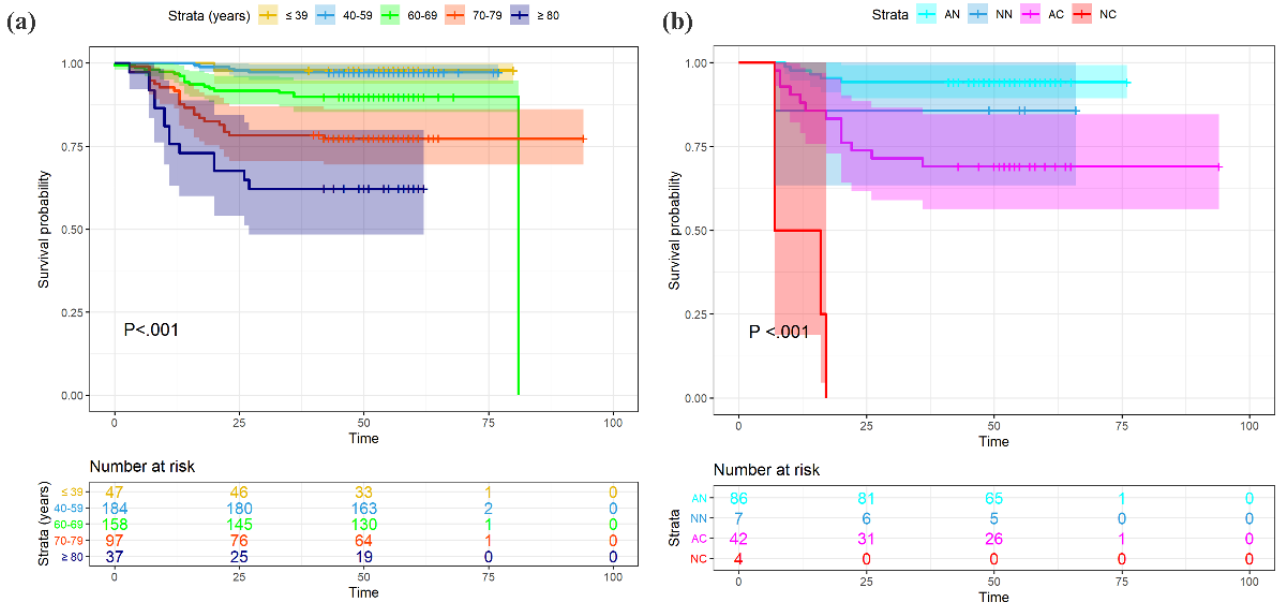


Figure 4. R software output showing the association of computed tomography characteristics with outcomes of patients with coronavirus disease. The hazard ratios of each variable were obtained using proportional hazard Cox models after adjustment for age and sex. ***P*<.01.

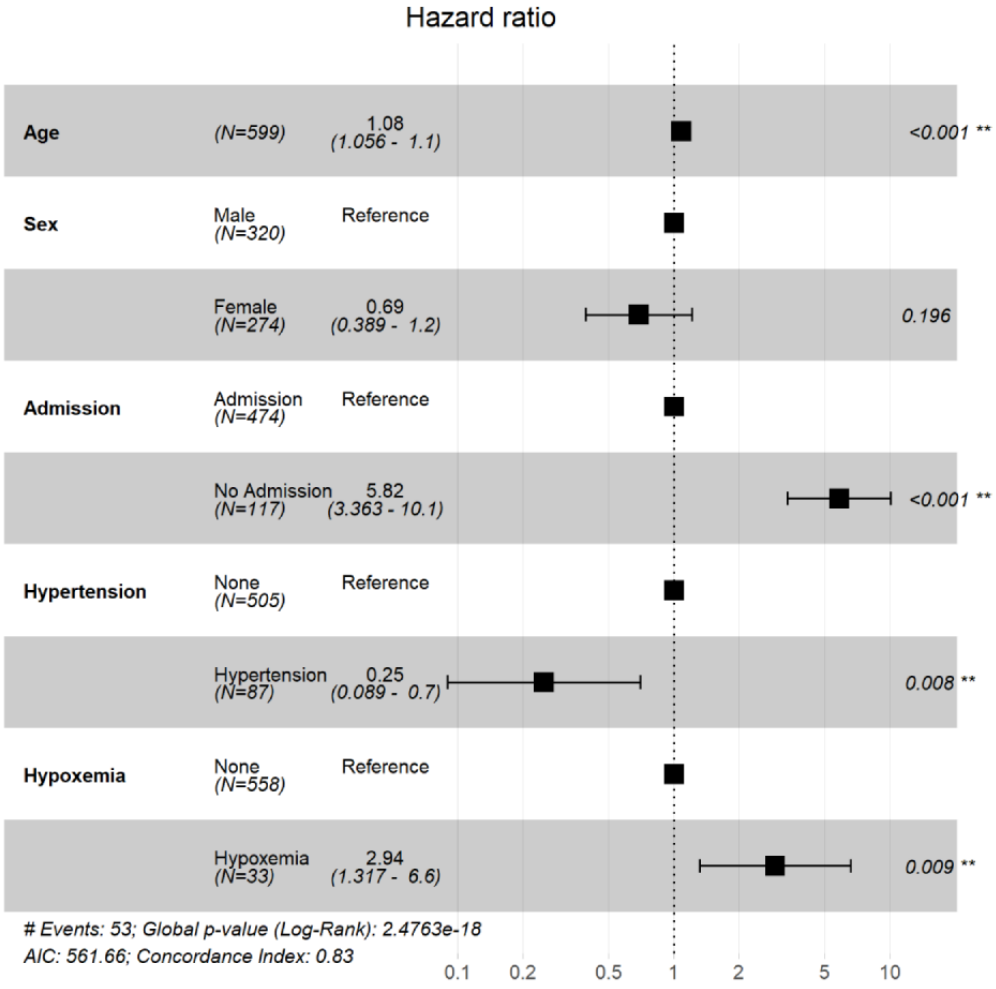
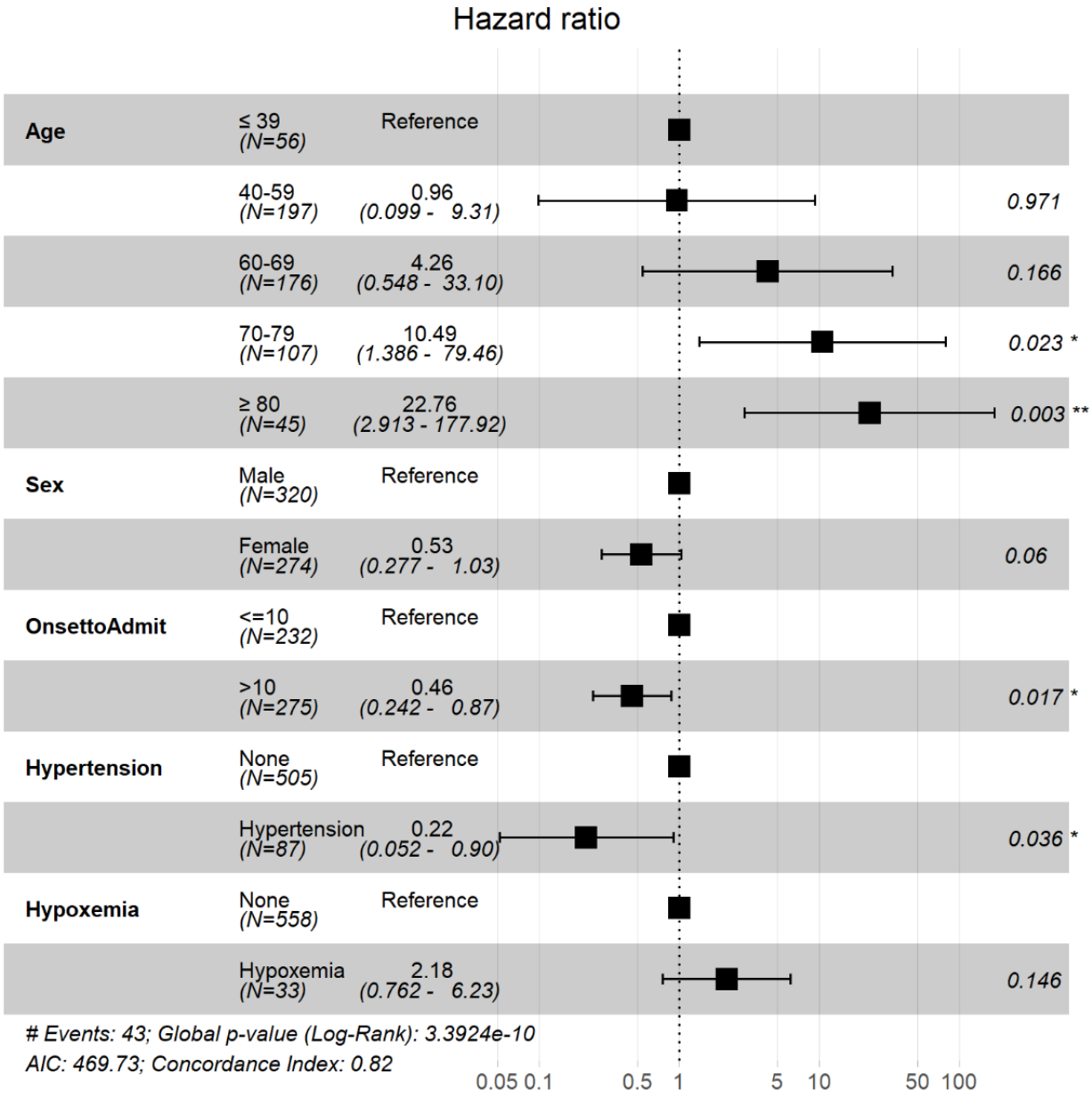


Figure 5. R software output showing the association of computed tomography characteristics with outcomes of patients with coronavirus disease adjusted for time between onset of symptoms and hospital admission. The hazard ratios of each variable were obtained using proportional hazard Cox models after adjustment for age and sex. Onset to admission refers to the number of days between symptom onset and hospital admission. OnsetToAdmit: onset to admission. *P<.05, **P<.01.



Radiographic Findings

The proportions of the 227 patients with unilateral, bilateral, or diffuse pneumonia were 25 (11%), 198 (87.2%), and 4 (1.8%), respectively. Of the 224 patients with positive or negative pleural effusion, 5 (2.2%) had positive effusion; see Table 3.

The Fisher exact tests suggested that lesion distribution differences were significant between patients who died and recovered (P=.045). Of the 25 patients with unilateral lesion distribution, 1 (4.0%) died. The proportions of patients in the bilateral and diffuse groups who died increased to 34/198 (17.2%) and 2/4 (50.0%); see Figure 5. In the pleural effusion positive group, 2/5 patients (40%) died. The proportion of patients in the negative group who died was 33/219 (15.1%, P=.18); see Table 2 and Figure 6.

Univariate Cox regression and Kaplan-Meier analysis also showed that patients with diffuse pneumonia had a significantly higher risk of death (P=.02 and P=.04, respectively); see Figure 7. Multivariate Cox regression adjusted for age and sex showed that only diffuse pneumonia was marginally significant (P=.05); see Figure 8. Kaplan-Meier analysis indicated that pleural effusion was marginally significant, and it was no longer significant when multivariate Cox regression was applied. Figure 9 shows examples of lesion distributions in CT images of patients with COVID-19 pneumonia.

Table 3. Radiographic characteristics of patients with coronavirus disease (n=227), n (%).

Computed tomography characteristic	Patient outcomes			*P* value
	All	Recovered	Died	
Lesion distribution				.045
Total	227 (100.0)	190 (83.7)	37 (16.3)	
Unilateral	25 (11.0)	24 (12.6)	1 (2.7)	
Bilateral	198 (87.2)	164 (86.3)	34 (91.9)	
Diffuse	4 (1.8)	2 (1.1)	2 (5.4)	
Lesion characteristics				.78
Total	222 (97.8)	186 (83.8)	36 (15.9)	
Ground-glass opacity	120 (54.1)	100 (53.8)	20 (55.6)	
Patchy shadowing	27 (12.2)	22 (11.8)	5 (13.9)	
Mixed	70 (31.5)	59 (31.7)	11 (30.6)	
Predominant consolidation	5 (2.3)	5 (2.7)	0 (0)	
Pleural effusion				.18
Total	224 (98.7)	189 (83.3)	35 (15.4)	
Negative	219 (97.8)	186 (98.4)	33 (94.3)	
Positive	5 (2.2)	3 (1.6)	2 (5.7)	

Figure 6. Histogram of lesion distribution among patients with coronavirus disease. NA: not applicable.

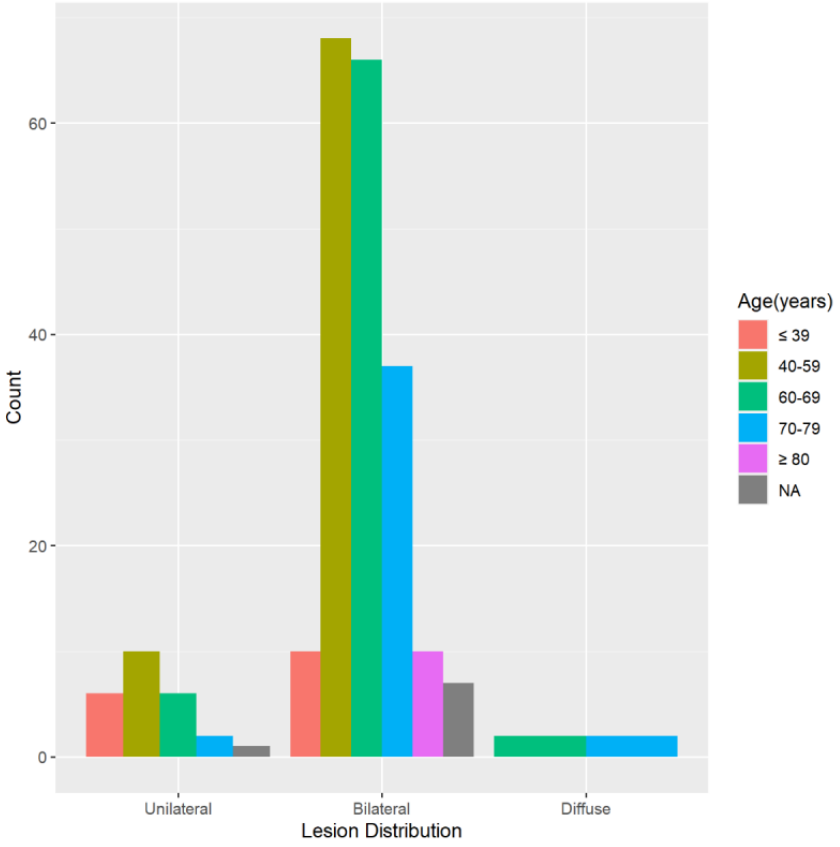


Figure 7. Cumulative incidence of death of patients with grouped by (a) lesion distribution and (b) pleural effusion.

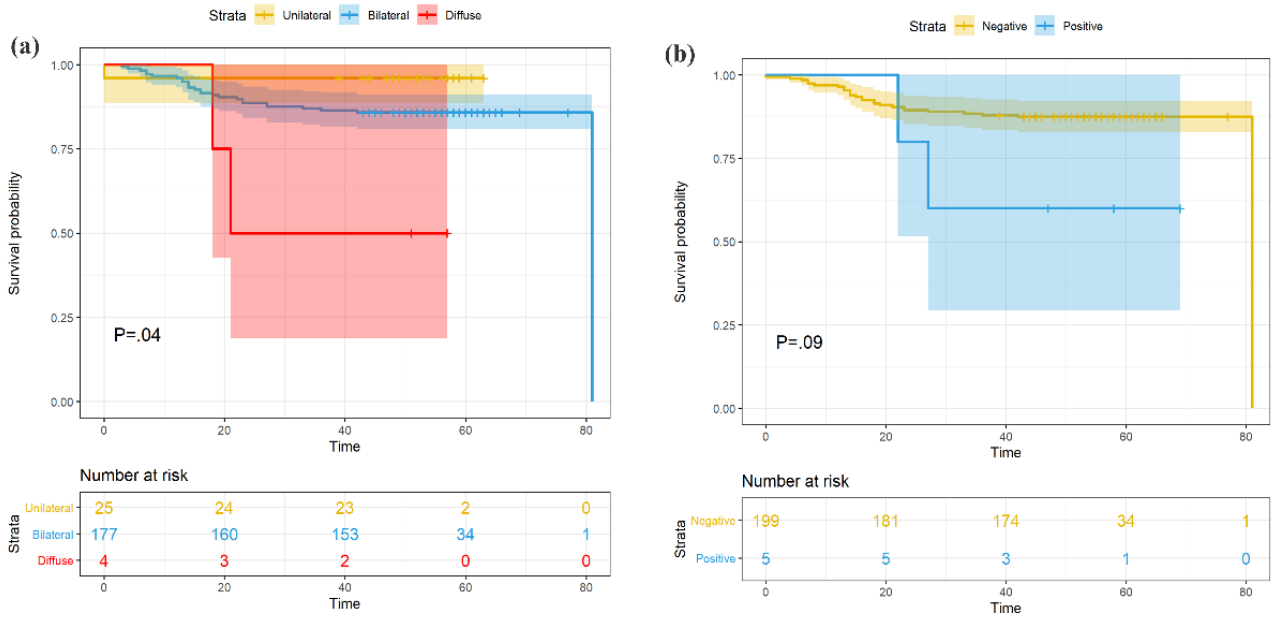


Figure 8. R software output showing the association of computed tomography characteristics with outcomes of patients with coronavirus disease. The hazard ratios of each variable were obtained using proportional hazard Cox models after adjustment for age and sex. ***P<.001.

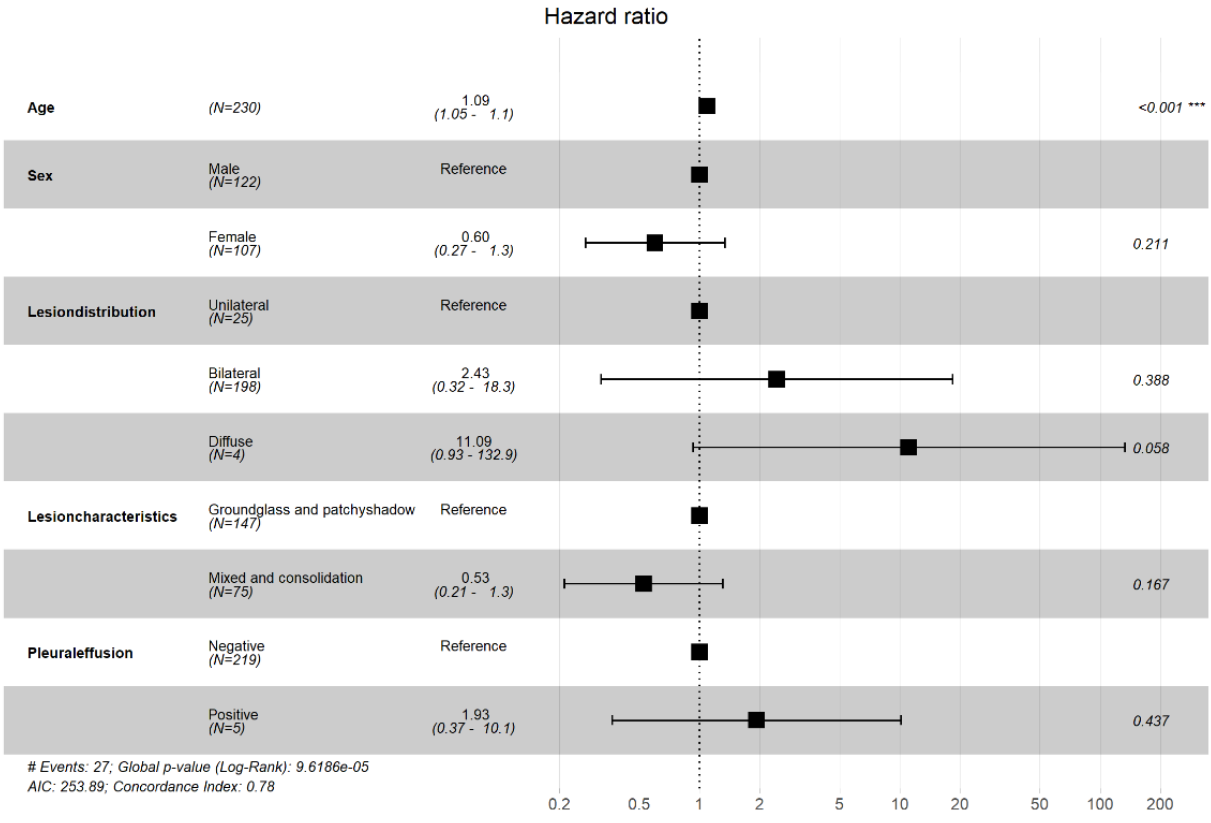
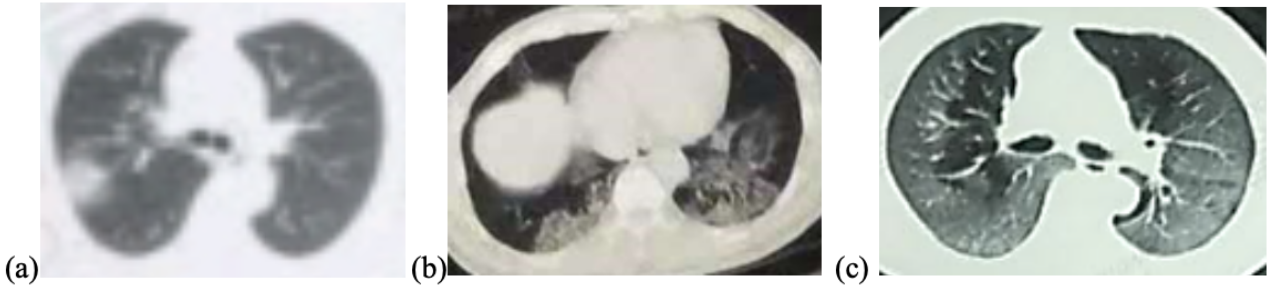


Figure 9. Lesion distribution on computed tomography (CT) images in patients with coronavirus disease pneumonia. (a) 60-year-old woman, unilateral lesion distribution; the axial CT image shows patchy shadowing in the right upper lobe. This patient recovered after 20 days of treatment in hospital. (b) 53-year-old man, bilateral lesion distribution; the axial CT image shows mixed lesions with ground-glass opacity and patchy shadowing in the bilateral lower lobes. This patient recovered after 30 days of treatment in hospital. (c) 68-year-old man, diffuse lesion distribution; the axial CT image shows ground-glass opacity in bilateral lungs. This patient died without treatment in hospital.



Discussion

Principal Findings

In this case series of COVID-19 patients reported on the internet in Wuhan, China, one-third were older people; people aged ≥70 years represent 174/581 (29.9%) of cases. The overall mortality rate was 83/599 (13.9%). Age is one of the most frequently reported prognostic factors in COVID-19; this has been reported consistently in many recent studies worldwide [7-9]. When we stratified the data by age group, the case-fatality rate increased with age; see Figure 3. Mortality was 0/19 (0%) for patients younger than 30 years and 21/45 (47%) for patients older than 80 years. This result was similar to that in previous reports from Italy, China, and the United States [7-9]. The fatality rate was slightly higher in male patients than in female patients in this sample but was only marginally significant. Our results also showed that smoking was not associated with higher mortality risk (P=.54). We also added smoking to the multivariate Cox regression analysis; however, this result was also not significant.

The median time from symptom onset to discharge and the length of hospital stay was much longer than those in the deceased patients, which may be due to the discharge standard of COVID-19 patients in China (afebrile for >3 days; improved respiratory symptoms; pulmonary imaging shows apparent absorption of inflammation; two consecutive negative nucleic acid tests for respiratory tract pathogen with a sampling interval ≥24 hours) [10]. The median time length of the patients' hospital stays was 13 days. Even when intensive and supportive care is provided, rapid progression can result in high mortality shortly after admission [11].

The typical clinical characteristics presented here were consistent with recent studies, except that the incidence of cough was lower (40.8% vs 72.2%) [12]. Because our data were acquired from the internet and were mostly reported by family members rather than the patients, some symptoms may not have been adequately presented. The incidence rate of dyspnea was found to be significantly higher in patients who died of COVID-19 in two previous reports [13,14]; this rate was also higher among patients who died in this study, but the difference was not significant. Consistent with earlier reports, hypoxemia and confusion were more frequent among patients who died [13,14]; these symptoms can be used as indicators of poor prognosis in patients with COVID-19 at baseline. It should be noted that hypoxemia was still significantly associated with death when age and sex were controlled in the multivariable-adjusted Cox regression model. On the other hand, we found that cough, vomiting, and chest distress were more frequent among recovered patients. However, these symptoms should be further confirmed before being used as favorable prognostic factors in patients with COVID-19 because of the nature of these data.

As in other reports [7,9], hypertension was the most common comorbidity in the sample. Although neither Kaplan-Meier analysis nor univariate Cox regression yielded significant results, patients with hypertension had a lower mortality risk when covariates were controlled. Several studies have shown that hypertension is related to higher mortality risk [3,15]; however, neither of these studies controlled confounding variables. Zhang et al [16] found that taking ACEIs or ARBs was associated with lower mortality risk after adjusting for covariates among patients with COVID-19 who had hypertension. The effects of ACEIs and ARBs may be positive [16,17]. Unfortunately, the data of ACEI and ARB use were not collected at first. Only 47/86 (54.7%) of the patients with hypertension answered our medication question. Of these patients, 16/47 (34%) took ACEIs or ARBs; see Multimedia Appendix 1. Although our findings suggest that hypertension and associated medication play roles in good outcomes of patients with COVID-19, their prognostic value cannot be fully verified by our study.

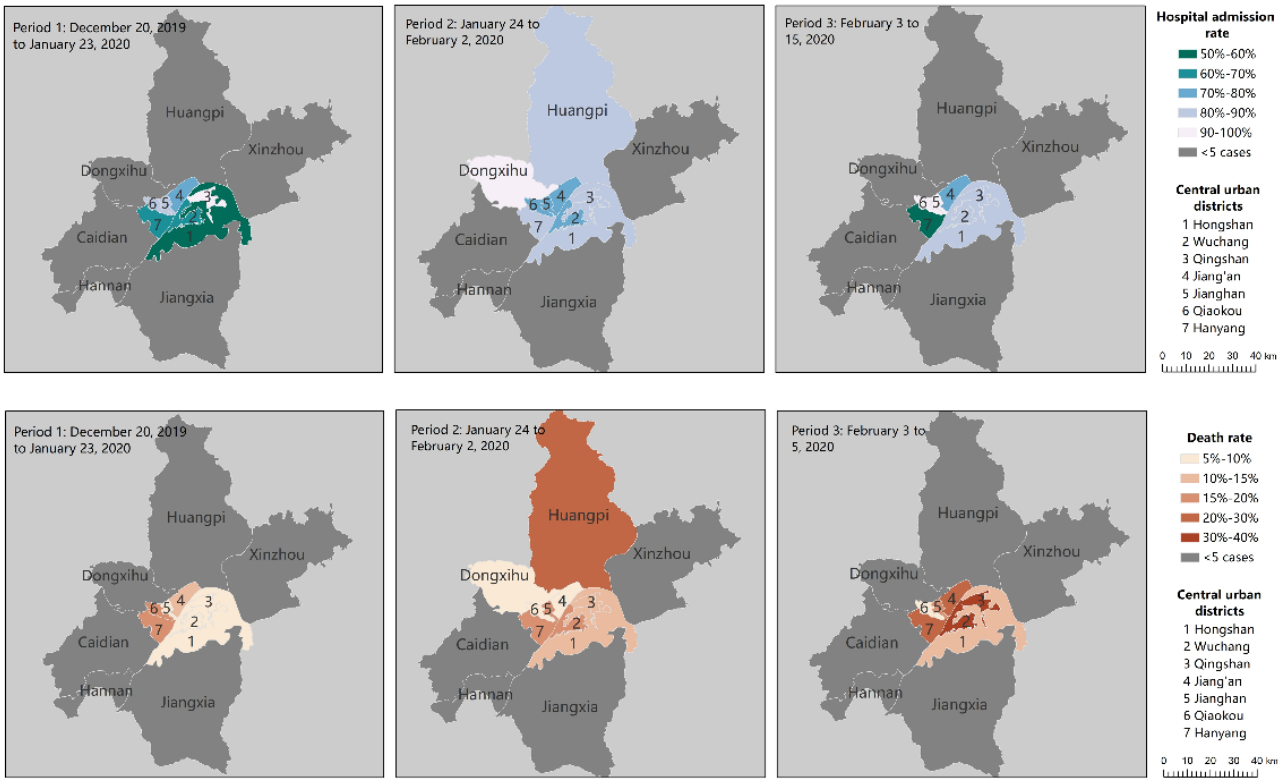
The predominant patterns of abnormality observed were bilateral opacity (198/227, 87.2%) and ground-glass opacity (120/222, 54.1%), which is consistent with another report [18]. We identified diffuse pneumonia as a risk factor for mortality and pleural effusion [19]. Because the CT images in our study were acquired from pictures on the internet and some were not of high quality, these findings should be further confirmed.

Hospital admission and the time from illness onset to hospitalization were significant prognostic factors. Both univariate Cox regression and Kaplan-Meier analysis indicated that hospital admission and disease severity (critically ill or not) were associated with death risk (P<.001). This suggests that in-hospital care can not only help patients control symptoms but can also prevent aggravation of the disease for some patients; see Figure 1 (a) and Figure 4. Although the time from illness onset to hospitalization was not significantly longer in the death group, the Kaplan-Meier analysis of the time from illness onset

to admission (≤10 days or >10 days) implied that patients with delayed medical care had a higher mortality risk; see Figure 1 (b). Together, these findings suggest that timely hospital care contributes to alleviating the severity and improving the prognosis of patients with COVID-19; see Figure 1 (e).

We also plotted the mortality and admission rates of COVID-19 on the map across three periods. According to Pan et al [20], after outbound transportation from Wuhan was blocked and public and vehicular transit were suspended on January 23, 2020, the effective reproduction number (R_t) of COVID-19 was reduced and its spread to other cities was delayed [21]. However, previous reports did not evaluate the effects of city shutdown on hospital admission and death rates. Figure 10 shows the geographic distributions of the mortality and hospital admission rates of COVID-19 cases across three time periods in Wuhan. The three time periods depicted are in reference to Pan et al [20]. The time from December 20, 2019, to January 23, 2020, was considered as the first period, when no COVID-19–specific interventions were imposed. After January 23, the government blocked all outbound transportation from Wuhan and suspended public and vehicular transportation in the city. On February 2, centralized quarantine and a "treatment of all cases" policy were implemented, and the number of hospital beds and medical supplies increased. We found that the mortality rate rose in most districts of Wuhan following the closure of the city. On the other hand, the admission rates fluctuated in different regions in periods 2 and 3. There were noticeable geographic differences in mortality rates, with the highest rates occurring in the suburban districts (see Multimedia Appendix 1). The shutdown of the city may have helped control the disease; however, it may have led to a rise in the mortality rate in a certain period. It should be noted that this result should be explained with caution, as our sample was not random and may not be representative.

Figure 10. Geographic distributions of mortality and hospital admission rates of coronavirus cases across three time periods in Wuhan, China. The mortality and admission rate of the cases were calculated using the number of deaths or admissions divided by the total number of cases reported in the area and time period. The data include all 13 districts of the city of Wuhan; regions with fewer than 5 cases were considered to be nonrepresentative and are plotted in grey.



Policy Implications

Our study suggests that use of social media data can be effective to identify patients at high risk for COVID-19, help coordinate appropriate treatment, and lower the mortality rate. The use of social media can also reduce cross-infection risks by reducing the number repeat visits of low-risk patients to the hospital. In the future, social media can be adopted to effectively help potentially critically ill patients seek timely medical treatment, help patients with low mortality risks to reduce unnecessary cross-infection, screen out critically ill patients in urgent need of hospitalization, and finally to facilitate disease control and hierarchical management. One limitation of this method is that it still requires particular definition and attention from the aspects of law, policy, and ethics; another limitation is that it requires active management and supervision procedures with participation of medical professionals to ensure its accuracy, effectiveness, and reasonableness.

Limitations

This study has several limitations. First, the data were acquired on the internet and followed up via telephone. The nature of the data did not allow us to obtain more detailed information. Second, we did not obtain details regarding the patients'

laboratory characteristics, clinical course, or treatment. Also, some radiological files were not complete.

Conclusions

Hospital admission at an appropriate time is vital for patients with COVID-19, especially those who are critically ill. Older age, hypoxemia, and pleural effusion were related to poor prognosis of mortality. Public health measures such as transportation blocking and city closure should be combined with other measures, such as increasing admission rates and shortening wait times for treatment.

Currently, more than 2.9 billion individuals use social media regularly. Considering the substantially high speed, reach, penetration, and transparency of social media platforms, social media can be used not only to disseminate but also to collect critical information about a sudden outbreak of disease. Individual patients' reports of their symptoms, clinical characteristics, treatment, and clinical outcomes on social media can be aggregated into big data and analyzed in real time to provide valuable insights to accelerate research speed [22].

Acknowledgments

The study is supported by Young Backbone Teachers' Overseas Study program under the State Scholarship Fund (File No. 201806015049) from WYY and Renmin University of China and by Double-First Class Innovation Research Funding from LD (File No. KYGJD2020002).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental materials.
[DOCX File , 995 KB - jmir_v22i8e20108_app1.docx]

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/s1473-3099(20)30120-1]
2. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020 May;8(5):475-481 [FREE Full text] [doi: 10.1016/s2213-2600(20)30079-5]
3. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/s0140-6736(20)30566-3]
4. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, COVID-19 Lombardy ICU Network. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. JAMA 2020 Apr 06;323(16):1574-1581 [FREE Full text] [doi: 10.1001/jama.2020.5394] [Medline: 32250385]
5. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. Kidney Int 2020 May;97(5):829-838 [FREE Full text] [doi: 10.1016/j.kint.2020.03.005] [Medline: 32247631]
6. Guo T, Fan Y, Chen M, Wu X, Zhang L, He T, et al. Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19). JAMA Cardiol 2020 Mar 27;5(7):811-818 [FREE Full text] [doi: 10.1001/jamacardio.2020.1017] [Medline: 32219356]
7. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA 2020 Mar 23;323(18):1775-1776 [FREE Full text] [doi: 10.1001/jama.2020.4683] [Medline: 32203977]
8. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. China CDC Weekly 2020;2(8):113-122 [FREE Full text] [doi: 10.46234/ccdcw2020.032]
9. Wadhera RK, Wadhera P, Gaba P, Figueroa JF, Joynt Maddox KE, Yeh RW, et al. Variation in COVID-19 Hospitalizations and Deaths Across New York City Boroughs. JAMA 2020 Apr 29;323(21):2192-2195 [FREE Full text] [doi: 10.1001/jama.2020.7197] [Medline: 32347898]
10. Interim diagnosis and treatment of 2019 novel coronavirus pneumonia. Article in Chinese. National Health Commission of the People's Republic of China. 2020. URL: http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml [accessed 2020-03-04]
11. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. JAMA Intern Med 2020 Mar 13;180(7):934-943 [FREE Full text] [doi: 10.1001/jamainternmed.2020.0994] [Medline: 32167524]

12. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/nejmoa2002032]
13. Deng Y, Liu W, Liu K, Fang Y, Shang J, Zhou L, et al. Clinical characteristics of fatal and recovered cases of coronavirus disease 2019 in Wuhan, China: a retrospective study. Chin Med J (Engl) 2020 Jul 05;133(11):1261-1267 [FREE Full text] [doi: 10.1097/CM9.0000000000000824] [Medline: 32209890]
14. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. BMJ 2020 Mar 26;368:m1091 [FREE Full text] [doi: 10.1136/bmj.m1091] [Medline: 32217556]
15. Li J, Wang X, Chen J, Zhang H, Deng A. Association of Renin-Angiotensin System Inhibitors With Severity or Risk of Death in Patients With Hypertension Hospitalized for Coronavirus Disease 2019 (COVID-19) Infection in Wuhan, China. JAMA Cardiol 2020 May 23;5(7):825-830 [FREE Full text] [doi: 10.1001/jamacardio.2020.1624] [Medline: 32324209]
16. Gurwitz D. Angiotensin receptor blockers as tentative SARS-CoV-2 therapeutics. Drug Dev Res 2020 Mar 04:1-4 [FREE Full text] [doi: 10.1002/ddr.21656] [Medline: 32129518]
17. Zhang P, Zhu L, Cai J, Lei F, Qin J, Xie J, et al. Association of Inpatient Use of Angiotensin-Converting Enzyme Inhibitors and Angiotensin II Receptor Blockers With Mortality Among Patients With Hypertension Hospitalized With COVID-19. Circ Res 2020 Jun 05;126(12):1671-1681 [FREE Full text] [doi: 10.1161/CIRCRESAHA.120.317134] [Medline: 32302265]
18. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. Lancet Infect Dis 2020 Apr;20(4):425-434 [FREE Full text] [doi: 10.1016/s1473-3099(20)30086-4]
19. Li K, Wu J, Wu F, Guo D, Chen L, Fang Z, et al. The Clinical and Chest CT Features Associated With Severe and Critical COVID-19 Pneumonia. Investigative Radiology 2020;55(6):327-331 [FREE Full text] [doi: 10.1097/rli.0000000000000672]
20. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. JAMA 2020 May 10;323(19):1915-1923 [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]
21. Tian H, Liu Y, Li Y, Wu CH, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science 2020 May 08;368(6491):638-642 [FREE Full text] [doi: 10.1126/science.abb6105] [Medline: 32234804]
22. Merchant R, Lurie N. Social Media and Emergency Preparedness in Response to Novel Coronavirus. JAMA 2020 Mar 23;323(20):2011-2012. [doi: 10.1001/jama.2020.4469] [Medline: 32202611]

Abbreviations

ACEI: angiotensin-converting enzyme inhibitor
ARB: angiotensin receptor blocker
ARD: acute respiratory distress
COVID-19: coronavirus disease
CT: computed tomography
ICU: intensive care unit
R_t: effective reproduction number
RT-PCR: reverse transcription–polymerase chain reaction

Edited by G Eysenbach; submitted 11.05.20; peer-reviewed by A Bhagavathula, N Yamada; comments to author 16.07.20; revised version received 17.07.20; accepted 26.07.20; published 13.08.20.

Please cite as:
Liu D, Wang Y, Wang J, Liu J, Yue Y, Liu W, Zhang F, Wang Z
Characteristics and Outcomes of a Sample of Patients With COVID-19 Identified Through Social Media in Wuhan, China: Observational Study
J Med Internet Res 2020;22(8):e20108
URL: http://www.jmir.org/2020/8/e20108/
doi:10.2196/20108
PMID:32716901

Original Paper

Influence of Social Media Platforms on Public Health Protection Against the COVID-19 Pandemic via the Mediating Effects of Public Health Awareness and Behavioral Changes: Integrated Model

Hani Al-Dmour[1*], PhD; Ra'ed Masa'deh[1*], PhD; Amer Salman[1*], PhD; Mohammad Abuhashesh[2*], PhD; Rand Al-Dmour[1*], PhD

[1]The University of Jordan, Amman, Jordan

[2]Princess Sumaya University for Technology, Amman, Jordan

[*]all authors contributed equally

Corresponding Author:
Hani Al-Dmour, PhD
The University of Jordan
Amman-Jordan
Amman, 11962
Jordan
Phone: 962 795666979
Email: dmourh@ju.edu.jo

Abstract

Background: Despite the growing body of literature examining social media in health contexts, including public health communication, promotion, and surveillance, limited insight has been provided into how the utility of social media may vary depending on the particular public health objectives governing an intervention. For example, the extent to which social media platforms contribute to enhancing public health awareness and prevention during epidemic disease transmission is currently unknown. Doubtlessly, coronavirus disease (COVID-19) represents a great challenge at the global level, aggressively affecting large cities and public gatherings and thereby having substantial impacts on many health care systems worldwide as a result of its rapid spread. Each country has its capacity and reacts according to its perception of threat, economy, health care policy, and the health care system structure. Furthermore, we noted a lack of research focusing on the role of social media campaigns in public health awareness and public protection against the COVID-19 pandemic in Jordan as a developing country.

Objective: The purpose of this study was to examine the influence of social media platforms on public health protection against the COVID-19 pandemic via public health awareness and public health behavioral changes as mediating factors in Jordan.

Methods: A quantitative approach and several social media platforms were used to collect data via web questionnaires in Jordan, and a total of 2555 social media users were sampled. This study used structural equation modeling to analyze and verify the study variables.

Results: The main findings revealed that the use of social media platforms had a significant positive influence on public health protection against COVID-19 as a pandemic. Public health awareness and public health behavioral changes significantly acted as partial mediators in this relationship. Therefore, a better understanding of the effects of the use of social media interventions on public health protection against COVID-19 while taking public health awareness and behavioral changes into account as mediators should be helpful when developing any health promotion strategy plan.

Conclusions: Our findings suggest that the use of social media platforms can positively influence awareness of public health behavioral changes and public protection against COVID-19. Public health authorities may use social media platforms as an effective tool to increase public health awareness through dissemination of brief messages to targeted populations. However, more research is needed to validate how social media channels can be used to improve health knowledge and adoption of healthy behaviors in a cross-cultural context.

(J Med Internet Res 2020;22(8):e19996) doi:10.2196/19996

KEYWORDS

social media platforms; Interventions; public health; awareness; public health protection; coronavirus; COVID-19; pandemic; behavioral change; Jordan; behavior; social media

Introduction

Background

Infection with coronavirus disease (COVID-19) has become a severe public health issue worldwide. COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a novel coronavirus that recently emerged from China. In March 2020, the World Health Organization (WHO) declared that COVID-19 can be characterized as a pandemic. Therefore, it is of utmost importance to prevent further spread of the pandemic in public and health care settings [1,2]. Scholars have reported that evidence of the impact of social media on health knowledge, behavior, and outcomes show that these tools can be effective in meeting individual and population health needs. Most research addresses specific interventions and approaches, which vary widely in focus, target population, theoretical foundations, mode of delivery, functionality, and usability. Due to this wide variation, it is difficult to discover what works and how, and efforts to compare approaches are complicated [3]. General strategies and guidelines include social distancing, testing every suspected case, staying home, avoiding social gatherings, treating patients, and contact tracing [4]. However, some countries are taking stricter measures to contain the pandemic, such as lockdowns and mass testing.

Jordan has been under nearly total lockdown since March 14, 2020. After only a few COVID-19 cases appeared around the country, the Jordanian government took extraordinary measures, including implementing strict emergency laws. Travel restrictions were imposed on passengers in Jordan; all unauthorized travel into and out of the country and between cities was halted. Nonpharmaceutical physical distancing interventions, such as extended university and school closures and workplace distancing, were introduced to reduce the impact of COVID-19. Public awareness and prevention of COVID-19 infection play important roles in disease control; a lack of reasonable knowledge of infectious diseases leads to low detection rates. Therefore, to stop the spread of COVID-19 infection in Jordan, the Jordanian Ministry of Health launched specific national disease control measures, using several media campaigns [5], posters, and advertisements on television and printed media along with other methods to improve the awareness of this pandemic among the general population. The assessment of government websites and social media platforms for public awareness is important because it helps determine the impact of governmental prevention efforts and measures and gauges the need for intervention [6].

Researchers have indicated that most developing countries encounter serious difficulties in preventing the spread of infectious diseases due to inadequate medical facilities and lack of resources [7-10]. In light of the apparent weakness of health care systems in these countries, public awareness of infectious diseases leads to behavioral changes among the public, thereby representing partial treatment; notably, this awareness reduces the pressure and economic burden on medical facilities. To raise public awareness, social media platforms are considered to be effective tools that contribute to the real-time dissemination of information about the current status of the disease and give appropriate advice to the public on how to avoid being infected. Further, according to [11], social media platforms provide beneficial climate and socioeconomic data. Additionally, social media platforms have been shown to represent an essential source of communication that enables the creation and dissemination of information to people through the internet [12,13]. It is worth mentioning here that social media platforms allow groups and individuals to exchange information about all subjects and issues, including members of minority groups or people who have no opportunity to express their opinions using other information sources. Researchers have argued that information and perspectives pertinent to issues associated with human health are revealed by informally using social media platforms away from official medical and health departments [14].

Despite the growing body of literature examining social media in health contexts, including public health communication, promotion, and surveillance [1,2], limited insight has been provided into how the utility of social media may vary depending on the particular public health objectives governing an intervention. For example, the extent to which social media platforms contribute to enhancing public health awareness and prevention during epidemic disease transmission is unknown. Korda et al [3] stated:

> *Evidence about social media's impact on health knowledge, behavior and outcomes shows that these tools can be effective in meeting individual and population health needs. Most research addresses specific interventions and approaches, which vary widely in focus, target population, theoretical foundations, mode of delivery, functionality, and usability. This wide variation makes it difficult to find out what works and how, and complicates efforts to compare approaches.*

Korda et al noted a lack of research focused on the role of social media campaigns on public health awareness of pandemic diseases such as COVID-19, particularly in Jordan. Therefore, in this study, we attempt to answer the following questions:

- Does the use of social media platforms raise public health awareness of COVID-19 as a pandemic disease?
- Does the use of social media platforms increase public behavioral changes toward COVID-19 as a pandemic disease?
- Does the use of social media platforms increase public protection against COVID-19 as a pandemic disease?
- Do public health awareness and behavioral changes play important roles in enhancing the relationship between the use of social media platforms (interventions) and public health protection against COVID-19 as a pandemic disease?

The findings of this study are expected to be helpful and important for public health authorities and governments to understand who receives the intervention (message), the impact of social media platform campaigns, and the extent to which changes in public health behavior and health outcomes can be attributed to the intervention, in addition to determining how the disseminated information is perceived.

Theoretical Background

Social media platforms are widely deployed by the WHO, health care professionals, and regulatory authorities worldwide to address key issues relating to public health [15]. They can be used to educated citizens and health care professionals on a broad range of themes, from the challenges surrounding anti-microbial resistance to topics such as adverse reaction reporting. The core focus of these initiatives is represented by awareness-building campaigns that take advantage of the large scale, breadth of reach, and immediacy of social media platforms to communicate quickly, effectively, and efficiently. Using social media to aid the prevention and control of infectious diseases can be cost-effective [8]. The health sector represents a critical area of government responsibility in most countries; it accounts for a large proportion of national spending, approximately equivalent to 9.9% of the global gross domestic product in 2016 [16]. Like other segments of the public sector, government health departments and national agencies are responsible for monitoring, protecting, and improving the health of residents, and state-funded health care delivery institutions are under increasing pressure to participate in electronic government (e-government) agendas. Many agencies will likely use social media platforms individually to achieve this. Widespread public engagement with social media platforms creates an effective ready-made path to their application in the health care field.

Social media platforms include a wide variety of networking sites (eg, Facebook), information-disseminating platforms (eg, YouTube), and microblogging services (eg, Twitter). These platforms and many others can be used to create and publish knowledge and information about potential health and disease risks and interventions as well as healthy lifestyles and effective health policies and strategies. In contrast to the campaigns occasionally launched by traditional media, campaigns launched through social media platforms often successfully convert knowledge and information on different health topics into daily fruitful web-based discussions and conversations [17]. Another key advantage of web-based social media data, in addition to the availability of an increasingly large volume of data, is that it is highly contextual and networked [2]. For example, there will be robust spatiotemporal sentiment toward a new vaccine, whether positive or negative. Risk factors such as drug abuse, smoking, poor diet, and lack of exercise and their associated diseases are often found to be clustered in a population. A better understanding of social media platforms and their health data will help broaden the utility of social media in public health.

Social media platforms constitute a powerful means of communication that can be used to elevate public awareness of infectious diseases, particularly new ones, in terms of outbreak dates and spreading developments [18]. Members of the public turn to both traditional and social media to obtain information on emerging infectious diseases which represent unprecedented risks to people [6]. The public perceptions of these risks are shaped depending on how information is communicated across social media platforms. This in turn affects people's behavior as well as the decisions they make. In addition to information dissemination through social media platforms, the users of these platforms participate in discussions and conversations by giving their own opinions and presenting their own experiences. However, information disseminated through social media platforms often lacks credibility because it is often generated by the users themselves rather than by medical specialists or professional health care institutions; therefore, this information may lack reliability, accuracy, correctness, or usefulness. As a result, the WHO has called for proactive and effective use of social media platforms to disseminate information on health issues, explicitly on emerging infectious diseases, to unspecialized persons and the general public.

Literature Review

Social media platforms have attracted the interest and attention of researchers and practitioners in the health domain, who use them for different purposes. These include professional training and development of clinicians; formation of health networks and support groups; provision of funding for health institutions; facilitation of cooperation and coordination among health professionals; monitoring of infectious diseases [2].

Even though social media platforms provide professionals in the public health domain with numerous valuable opportunities and benefits, usage of social media platforms by professionals is associated with several challenges, the most important of which are detecting infectious disease outbreaks, monitoring emergencies, predicting disease trends, and measuring the public's awareness and responses. However, many studies have reviewed and explored the potential applications of social media platforms for public health communication. For example, Huebsch et al [19] suggested that social media platforms allow health practitioners to establish a direct relationship with their clients and that health promotion planners must put forward their creative best to integrate social media platforms within their strategies to make full use of the potential of these platforms when marketing their products and services.

In a study conducted by Bennett and Glasgow [20], they examined the issues of how citizens seek to consume medical information and how the World Wide Web transforms the relationship between the public and medical professionals. On the other hand, Moorhead et al [21] reviewed studies that were conducted to investigate the uses, advantages, and limitations of social media platforms in realizing fruitful communication between health professionals, patients, and the general public. Although the advantages of using social media platforms in health communication were identified in some relevant studies, we noticed a lack of studies that discussed the assessment of the effectiveness of social media platforms in shaping and modifying health communication practices in the short and long terms. The use of social media platforms in disease surveillance was reviewed by Ellis et al [22]. Their literature review revealed the effectiveness of social media platforms in speed, accuracy,

and cost performance; therefore, they recommended the use of social media programs to support existing disease surveillance systems. Moreover, a few studies focused on the role of social media platforms in promoting health protection and increasing the relevance of public health messaging, in addition to identifying the lessons learned from social media health campaigns. Among these was an investigation carried out by Collinson et al [23] that indicated the importance of controlling the spread of influenza and reducing the infection effects on a population to public health. Social media campaigns related to epidemics or pandemics can be beneficial in conveying information to the general public, thereby inducing positive attitudes and behaviors that may slow the spread of the disease, such as hand washing and social distancing.

Studies on social media campaigns and healthy behavior have reported that social media campaigns can elicit positive behavior changes and even prevent negative behavior changes in individuals. Social media platforms can be used to reduce the spread of pandemics, thereby lowering the levels of fear and anxiety among the general public. Researchers have argued that social media communication can transfer useful information about infectious diseases based on identifying and tracking users' behavioral patterns [24]. A model was suggested by Misra et al [7] to explore the impact of awareness realized by social media campaigns on infectious disease prevalence, considering that the campaigns launched on social media platforms represent a variable whose growth depends on the number of individuals infected. The proposed model is based on the assumption that social media health campaigns lead to behavioral changes among individuals, causing them to isolate themselves and protect themselves from infection. The findings of the study by Misra et al [7], referred to above, reveal a decrease in the number of infected people with the increasing spread of social media health campaigns. The models discussed above were extended by Samanta et al [25] by assuming that people who are susceptible to a disease but aware of it are less likely to be infected than people who are unaware.

Further, it was assumed that social media health campaigns grow as the mortality rate of the disease increases. Researchers investigated the role of social media platforms in eliminating infectious diseases in the presence of treatment [14]; they found that information conveyed to the public by social media platforms through health campaigns leads to behavior modification [26]. Another study [27] showed the influence of information related to vaccination against infectious diseases on the endemic state; the results indicated that lower social media coverage leads to a globally stable endemic state. In contrast, the endemic state is characterized by instability and fluctuations in the case of higher social media coverage.

Previous research has found that disease prevalence is controlled by social media coverage. Behavioral changes related to infectious diseases through campaigns launched by print media, social media, and the internet are limited to a smaller population of educated people. In comparison, television advertisements tend to be capable of affecting the behavior toward infectious diseases of a larger population that generally consists of less educated people; therefore, television advertisements are more effective than other types of media. It has been found that social media health campaigns are associated with the number of infected individuals, where launching more campaigns on social media platforms reduces the number of infected people. The effectiveness of interventions made by social media networking sites in modifying the health behaviors of individuals was the subject of an investigation conducted by Laranjo et al [12], who uncovered a positive impact of those interventions on the health behaviors of individuals, but with a remarkable level of heterogeneity. Researchers found that internet data contribute to conveying useful information to the public, supporting the existing health surveillance systems and assisting in the fight against disease [8]. An evaluative examination was carried out [28] in which it was found that during outbreaks of infectious diseases, social media platforms negatively affect the quality of disease prediction and detection. The study was conducted on the diffusion of the Ebola virus as a case study. Sharma et al [29] found that most of the information posted on Facebook about the Zika virus was inaccurate and irrelevant. The influence of social media communication on shaping the behaviors of the general public was the concern of a study [30] about the Ebola and Zika viruses. It was found that users' trust in both media and authorities represents an important factor in the relationship under study. Most citizens lack accurate and relevant knowledge about the spread of infectious diseases over time and across space. Therefore, social media platforms can be used to establish a database of disease occurrences in terms of time and space. As useful surveillance tools, information on social media platforms was found to outperform official information about the outbreak and spread of infectious diseases, particularly in timeliness.

The Study Model

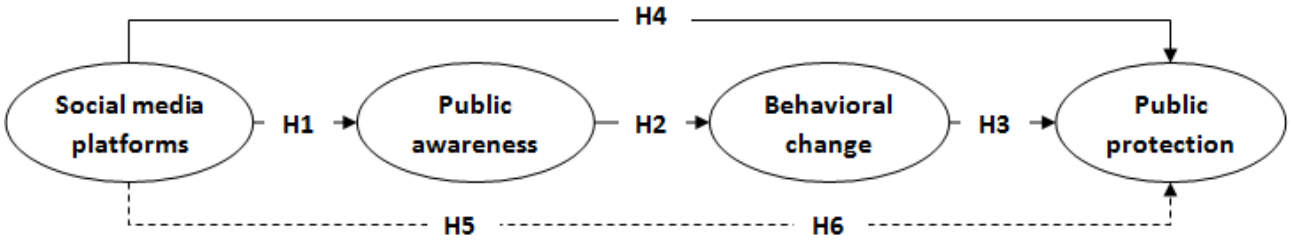
Elevating public health awareness was found to require the incorporation of some theories of behavioral change into social media health interventions [31]. According to [32], behavioral change theories (eg, social cognitive theory and health belief theory) can be beneficial when applied to social media initiatives because they help public health authorities understand the process of changing behaviors and why people behave in a precise manner, thereby enabling the health authorities to evaluate and modify health interventions [33-35]. According to the Health Belief Model, people tend to take preventive actions if they feel that they are seriously threatened. Based on this, health interventions should address the specific perceptions of individuals about susceptibility and benefits [36]. The behavioral change approach has been found to improve health by changing people's lifestyles [37]. It is assumed that individuals must understand basic facts about a specific issue related to their health to be able to change their lifestyles as a result of feeling threatened, especially by an infectious disease. In this context, individuals should learn a group of skills and be granted access to suitable services.

Behavior changes include but are not limited to hand washing, wearing masks, social distancing, avoiding public gatherings, sanitation, and isolation. Interventions aimed at promoting public health can improve the quality of health in society and support the policies and programs run by official health authorities in fighting the outbreak and spread of infectious diseases. If people have trust in these policies and programs, they are likely to

respond positively to public health interventions and participate in the launched health promotion programs in large numbers. Social media health campaigns can induce positive behavioral changes and even eliminate negative ones in individuals. According to Laranjo et al [12], the advantages of social media health interventions are cost-efficiency, ubiquity, and passing geographical barriers. The tremendous growth in social media networking sites has opened the door to more opportunities to disseminate health interventions to the general public in real time and irrespective of geographical location, thus leading to public health promotion and positive behavioral changes.

Therefore, an integrated conceptual model was developed to guide the objectives of this study. It was assumed that social media interventions as tools of health promotion programs would increase public protection and prevention against COVID-19 via the interaction between public awareness and behavior changes as mediating factors. These primary constructs are mainly derived from a theoretical background, relevant previous studies, and behavior change theory (health belief theory). In this study, the variables of primary interest (the independent variables) were social media platforms. The influence of the independent variables on the variance in the dependent variable (public protection against COVID-19) via the mediating factors (public awareness and behavioral changes) was studied. The expected relationships among these constructs are illustrated in Figure 1.

Figure 1. Diagram of the study model.



Based upon the above model, the following hypotheses were formulated concerning the role of social media campaigns in increasing public awareness of COVID-19 as a pandemic disease in Jordan:

Hypothesis 1 (H_1): The use of social media platforms is significantly increasing public health awareness.

Hypothesis 2 (H_2): Public health awareness is significantly contributing to public health behavioral change.

Hypothesis 3 (H_3): Public health behavioral change is significantly increasing public health protection.

Hypothesis 4 (H_4): The use of social media platforms is significantly increasing public health protection.

Hypothesis 5 (H_5): Public health awareness is significantly mediating the relationship between social media platforms and public health awareness.

Hypothesis 6 (H_6): Public health behavioral change is significantly mediating the relationship between the use of social media platforms and public health awareness.

Methods

Study Design

We employed a quantitative method with an exploratory and descriptive design. To confirm the conceptual model of the research and to investigate the research hypotheses, a survey questionnaire was employed to collect data (Multimedia Appendix 1). The target population of this research consisted of all followers on any social media platform in Jordan. To reach them, a web link to the questionnaire was sent to potential respondents during the period between March 15 and March 30, 2020. The questionnaire was prepared in the Arabic and English languages, and 2555 social media users were sampled to collect the data. The content of the questionnaire (constructs and measures) was mainly selected and adopted from previous relevant studies [2,10,29,30,36-38] using a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5). Table 1 summarizes these constructs and their related measurement items. To perform construct validation, the questionnaire content was modified to the practice of Jordanian business culture context based on the results of a pilot study and feedback from six professional academic staff members in this field. The survey instrument was reviewed by a panel of six academic researchers in the areas of marketing and management information systems to guarantee face validity. Consequently, several questions were modified, and the revised questionnaire was used for pilot testing on citizens who lived in Jordan during the COVID-19 pandemic. Indeed, a pretest was conducted with 25 citizens to check the understandability of the questions. Some revisions were made, resulting in an easily understandable survey questionnaire.

Table 2 shows that 1283 of the 2555 respondents in this study (50.2%) were female, and 1823 (71.3%) were aged 18 to 44 years. Additionally, 1352 of the 2555 respondents (52.9%) have a bachelor's degree, and 1219 (47.7%) live in Amman.

Table 1. Variables and measurement items.

Construct and measurement items	Description
Social media platforms (SMP)	
SMP1	Facebook helps me to recognize COVID-19[a].
SMP2	Instagram helps me to recognize COVID-19.
SMP3	Twitter helps me to recognize COVID-19.
SMP4	WhatsApp helps me to recognize COVID-19.
SMP5	YouTube helps me to recognize COVID-19.
Public awareness (PAW)	
PAW1	Facebook contributes to increasing my awareness/knowledge of how to prevent COVID-19.
PAW2	Instagram contributes to increasing my awareness/knowledge of how to prevent COVID-19.
PAW3	Twitter contributes to increasing my awareness/knowledge of how to prevent COVID-19.
PAW4	WhatsApp contributes to increasing my awareness/knowledge of how to prevent COVID-19.
PAW5	YouTube contributes to increasing my awareness/knowledge of how to prevent COVID-19.
Public behavioral change (PBC)	
PBC1	Facebook contributes to changes in my behavior to prevent COVID-19 by taking various preventive measures (such as not shaking hands or kissing, not leaving the house, eating healthy food and vitamins, general hygiene, lack of anxiety and fear of disease, and increasing religious belief).
PBC2	Instagram contributes to changes in my behavior to prevent COVID-19 by taking various preventive measures (such as not shaking hands or kissing, not leaving the house, eating healthy foods and vitamins, general hygiene, lack of anxiety and fear of disease, and increasing religious belief).
PBC3	Twitter contributes to changes in my behavior to prevent COVID-19 by taking various preventive measures (such as not shaking hands or kissing, not leaving home, eating healthy foods and vitamins, general hygiene, lack of anxiety and fear of disease, and increasing religious belief).
PBC4	WhatsApp contributes to changes in my behavior to prevent COVID-19 by taking various preventive measures (such as not shaking hands or kissing, not leaving home, eating healthy food and vitamins, general hygiene, lack of anxiety and fear of the disease, and increasing religious belief).
PBC5	YouTube contributes to changes in my behavior to prevent COVID-19 by taking various preventive measures (such as not shaking hands or kissing, not leaving home, eating healthy food and vitamins, general hygiene, lack of anxiety and fear of the disease, and increasing religious belief).
Public Protection (PPR)	
PPR1	Social media platforms contribute to behavioral changes to protect me from infection with COVID-19.
PPR2	Social media platforms contribute to behavioral changes to protect others from infection with COVID-19.
PPR3	Social media platforms contribute to behavioral changes in educating others about infection with COVID-19.

[a]COVID-19: coronavirus disease.

Table 2. Sample profile (N=2555), n (%).

Characteristic	Value
Gender	
Male	1272 (49.8)
Female	1283 (50.2)
Age (years)	
18-33	1196 (46.8)
34-43	627 (24.5)
44-53	447 (17.5)
54-63	206 (8.1)
≥64	79 (3.1)
Educational level	
High school or less	112 (4.4)
Diploma	262 (10.3)
Bachelor's degree	1352 (52.9)
Master's degree	421 (16.5)
PhD	520 (20.4)
Governorate	
Irbid	310 (12.1)
Balqa	123 (4.8)
Jerash	37 (1.4)
Zarqa	158 (6.2)
Tafilah	17 (0.7)
Ajloun	29 (1.1)
Aqaba	276 (10.8)
Amman	1219 (47.7)
Karak	227 (8.9)
Madaba	52 (2.0)
Maan	39 (1.5)
Mafraq	68 (2.7)

Descriptive Analysis

To illustrate the respondents' attitudes towered each question, they were asked in the assessment, and the mean and SD were accordingly calculated for all the measurements. The descriptive statistics offered in Table 3 point to a positive disposition toward the items measured. The level of every item was calculated by the follow method: (highest point in Likert scale – lowest point in Likert scale)/the number of the levels used = (5 – 1)/5 = 0.80, where a level of 1 to 1.80 was considered very low, 1.81 to 2.60 was low, 2.61 to 3.40 was moderate, 3.41 to 4.20 was high, and 4.21 to 5 was very high. After that, the items were ordered by their means [39,40]. Tables 3 and 4 show the results of the calculations.

Table 3. Descriptive statistics of the research items and variables.

Category	Mean (SD)	Level	Order
Social Media Platforms (SMP)			
SMP1	3.84 (1.111)	High	1
SMP2	3.07 (1.132)	Moderate	5
SMP3	3.17 (1.133)	Moderate	4
SMP4	3.54 (1.304)	High	3
SMP5	3.81 (1.151)	High	2
Public Awareness (PAW)			
PAW1	3.86 (1.117)	High	1
PAW2	3.17 (1.148)	Moderate	5
PAW3	3.18 (1.131)	Moderate	4
PAW4	3.58 (1.251)	High	3
PAW5	3.71 (1.144)	High	2
Public Behavioral Changes (PBC)			
PBC1	3.94 (1.120)	High	1
PBC2	3.28 (1.159)	Moderate	4
PBC3	3.26 (1.139)	Moderate	5
PBC4	3.67 (1.216)	High	3
PBC5	3.72 (1.130)	High	2
Public Protection (PPR)			
PPR1	3.99 (1.065)	High	1
PPR2	3.99 (1.052)	High	1
PPR3	3.96 (1.067)	High	2

Table 4. Overall means, SDs, levels, and orders of the study variables.

Type and variable	Mean (SD)	Level	Order
Independent			
Social media platforms	3.4849 (0.84353)	High	4
Mediating			
Public awareness	3.5011 (0.88427)	High	3
Public behavioral change	3.5754 (0.90706)	High	2
Dependent			
Public protection	3.9808 (1.02517)	High	1

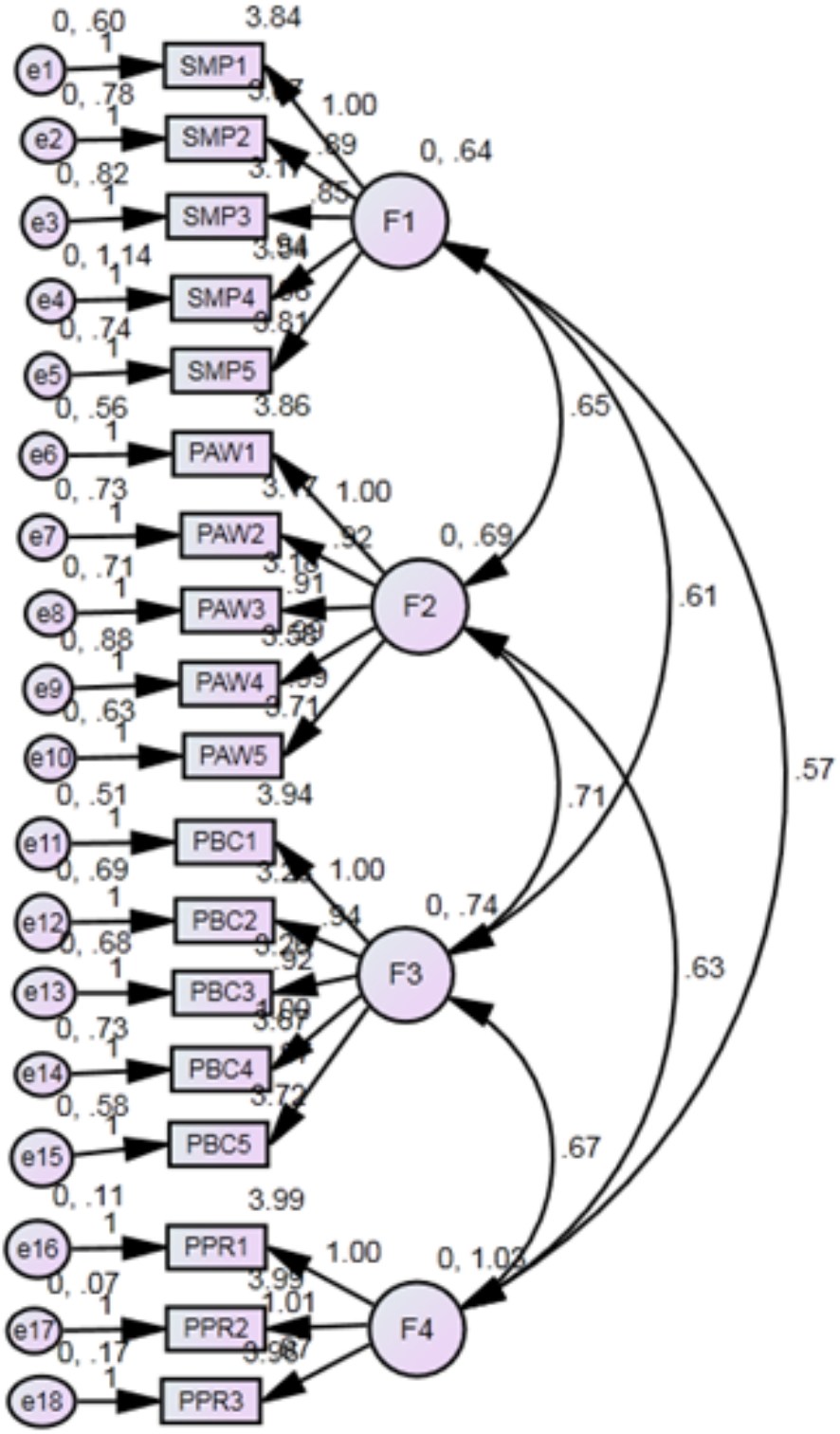
Exploratory Factor Analysis

Exploratory factor analysis is often used to gather information about research variables [39]. The outcome of the Kaiser-Meyer-Olkin test was 0.894, and all items were higher than 0.60; consequently, all items were used in the data analysis to capture the investigated latent variables. Furthermore, for the multicollinearity issue, the results indicate that the variance inflation factor for each variable was below 3, suggesting that multicollinearity was not an issue. The findings also indicate the absence of common method bias in that the first factor did not account for the majority of the variance, and no single factor occurred from the factor analysis [41].

Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) was performed to confirm the properties of the research items. Scholars have reported that the measurement model shows how latent variables or hypothetical variables are evaluated under the conditions of observed variables, representing how the validity and reliability of the observed variables answers for the latent variables [42,43]. Figure 2 shows the measurement model and the correlations among the four research variables; it can be seen that all variables were correlated.

Figure 2. Measurement model showing the correlations among the four research variables.



The initial CFA model provided an acceptable fit without eliminating any items to achieve an enhanced fitting measurement model. The goodness of fit indices of the evaluation of the initial research model indicated that the findings of the initial model could be deemed as the final model. CFA showed that for the model, χ^2_{1544}=4281.7 (P<.001), which implies that the measurement model fitted the data. Also, χ^2/df (4281.742/1544) = 2.773; this is an absolute fit index with a threshold of <3 for a serious viewpoint or <5 for adequate criteria. The incremental fit index of 0.89, the Tucker-Lewis index of 0.86, the comparative fit index of 0.87, and the root mean square error of approximation of 0.052 all meet the threshold of <1 for sufficient criteria [43]. Based on these fit indices, the measurement model indicated a good fit of the sample data.

To determine the reliability and validity of the research model, factor loadings, Cronbach alpha, composite reliability, and average variance extracted (AVE) for the variables were

calculated. The entire research indicator (ie, factor loadings) surpassed 0.50 [42,44]; this confirmed the convergent validity. All the composite reliability values surpassed 0.60, representing a high level of internal consistency for the latent variables. Also, each AVE value surpassed 0.50 [43]; thus, the convergent validity was proved (Multimedia Appendix 2 [45]).

Results

As shown in Table 6, the structural equation modelling analysis showed that H_1, H_2, H_3, and H_4 were supported.

Table 6. Path analysis results for hypotheses 1 to 4. For all hypotheses, $P<.001$.

Hypothesis	Path	Standardized effect (β)	Robust *t* (df)	Result
1	SMP[a]→PAW[b]	.823	64.128 (1544)	Supported
2	PAW→PBC[c]	.704	39.096 (1544)	Supported
3	PBC→PPR[d]	.465	16.134 (1544)	Supported
4	SMP→PPR	.149	5.301 (1544)	Supported

[a]SMP: social media platforms.

[b]PAW: public awareness.

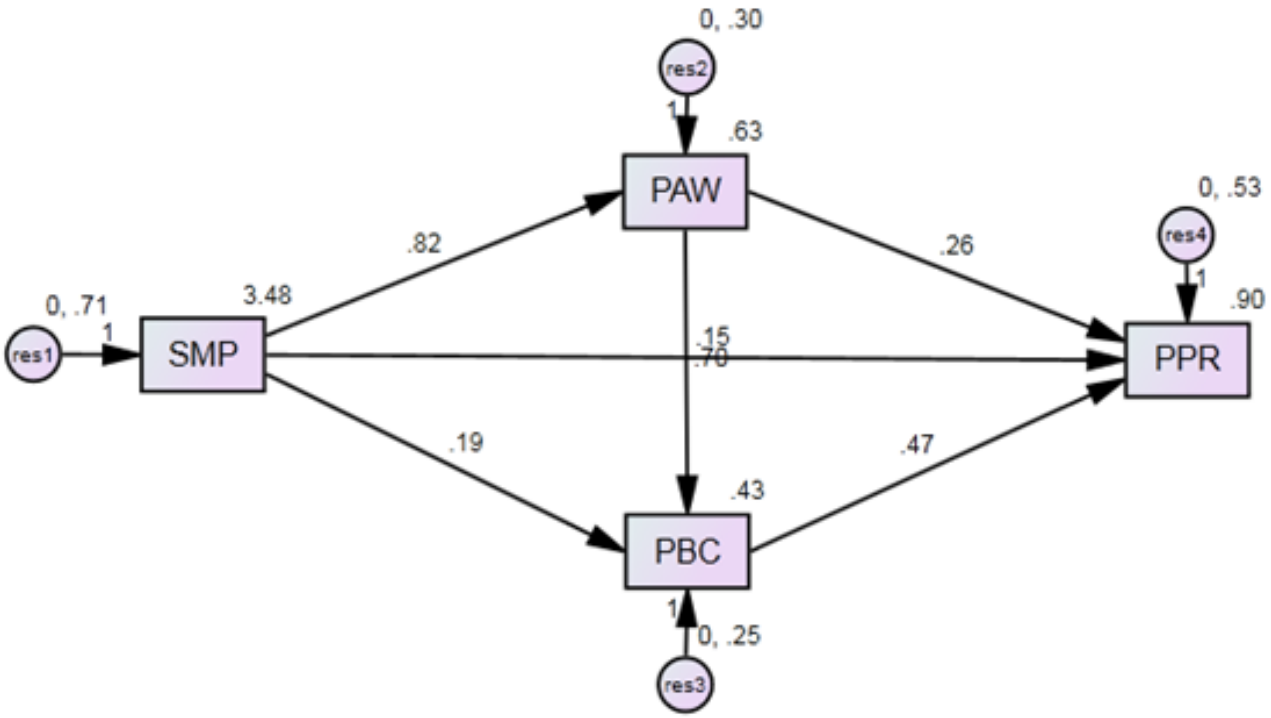
[c]PBC: public behavioral change.

[d]PPR: public protection.

To examine the mediating effects of public awareness, public behavioral change, and social media platforms on public protection, we considered both direct and indirect effects. It was found that public awareness and public behavioral change significantly affected public protection both directly (α=.149 for both H_5 and H_6) and indirectly (α_{H5}=.571, α_{H6}=.579), resulting in total effect sizes of α=.720 and α=.728 for H_5 and H_6, respectively; as a result, the data support partial mediation.

Further, to investigate the structural model, path analysis was conducted. We examined the research hypotheses via the statistical significance of the standardized regression weights (ie, *t* value; see Table 6) and the coefficient of determination R^2 for the endogenous research variables. The coefficients of determination for public awareness, public behavioral change, and public protection were 0.62, 0.70, and 0.50, respectively; these results show that the model strongly accounts for the variation of the research model. Figure 3 shows the estimated path values for the hypothesized model.

Figure 3. Estimated path values for the hypothesized structural model.



Discussion

Principal Findings

In this study, we aimed to explore the impact of using social media platform applications on health and safety during the COVID-19 pandemic through public health awareness and behavioral changes as mediating factors in Jordan. To achieve the study objectives and conduct the research using a systematic approach, a conceptual framework was developed based on a literature review and health belief change theory. The potential benefits of using social media platforms in public health protection against pandemic diseases include dissemination of public health interventions, enhanced public awareness, promotion of healthy behavior, improved health outcomes, and provision of health information to the community [8,24,46-50]. The analysis provides empirical evidence regarding the impact of using social media platforms on public health awareness, public health behavior changes, and health protection against COVID-19 in hypotheses 1, 3, and 4, respectively. These three hypotheses significantly and positively supported the linkage between use of social media platforms and public health awareness, public behavioral changes, and health protection. Numerous research studies have explored the relationship between the use of social media platforms and public health [2,10,12,29,30,36-38,47]. Furthermore, the analysis provided empirical evidence regarding the effectiveness of public health awareness on public health behavioral changes, as proposed in H_4. The results showed that the effect was positive and significant. Therefore, H_4 agreed with the findings of Lunn et al [31].

The fifth and sixth hypotheses (H_5 and H_6) were developed to determine whether there were mediating effects of public health awareness and public health behavioral changes on the relationship between the use of social media platforms and health protection against COVID-19. The results clearly showed that public health awareness and public health behavioral changes mediated the effects of social media platform use on health protection; however, the mediating effect was partial. Additionally, the results indicated a significant and positive indirect effect of social media platform use on health protection against COVID-19 through public health awareness and public health behavioral changes, with standardized indirect effects of 0.571 and 0.579 and *P* values <.001. The results showed a significant and statistical effect of social platforms use on health protection against COVID-19 without the mediating effects of public health awareness and public health behavioral changes. The total standardized effect was 0.149, which was significant (*P*<.001) but weak.

This discussion shows that the standardized direct effects between social media platform use and public health protection with public health awareness and public health behavioral change as mediators increased to .720 and .728, respectively, and the standardized direct effect of the same relationship in the absence of mediators was .149. This indicates that the mediation was partial. Thus, this result supports H_5 and H_6. Moreover, the indirect effects of social media platform use on health protection through public health awareness and public health behavioral change as mediators were both positive and significant, with a standardized indirect effect at a *P* value <.001. Therefore, the mediating effects of public health awareness and public health behavioral changes between usage of social media platforms and public health protection may be a new relationship. The results indicated that social media use had a significant and direct positive effect on public health protection. Social media platform use also had a direct effect on public health awareness; this effect was also significant, as was the effect of public health awareness on public health protection against COVID-19. However, no previous empirical research studies have examined the mediating effects of public health awareness and public health behavioral changes on the relationship between social media use and public health protection.

Therefore, the results confirm that public awareness and public health behavioral changes have vital mediating effects on the relationship between the use of social media platforms and public health protection, and the degree of mediation was partial. This supports H_5 and H_6. Furthermore, the main statistical results supported the predictive validity of the conceptual model of the study. Overall, the study validated the use of social media platforms to improve public health protection through public awareness. Therefore, we conclude that social media campaigns should be used to inform the public so that behavior changes can result.

Theoretical Contributions and Implications

This study fills the gap within the literature regarding a comprehensive understanding of the relationships between social media platform use, public protection against COVID-19 during the pandemic, public awareness, and public health behavioral changes. It also significantly contributes to supporting health belief theory by supporting the links between social media platform use and public protection against COVID-19 via the mediating roles of public awareness and public health behavioral changes. This study provides many theoretical contributions to the literature on social media platform applications for public health care and protection against COVID-19 as a pandemic disease, one of which is to validate the research framework applied in Jordan. Moreover, this research supports the application of social media platforms in Jordan, particularly in public health awareness.

The results endorse the mediating effect of public awareness and public health behavioral changes on the relationship between the independent variable, social media use, and the dependent variable, public health protection, which is another gap addressed by this research. Furthermore, this study has extended the literature that considers social media applications by providing the following:

First, we examined an unexplored connection between social media applications and public health protection against COVID-19 as a pandemic disease. Previous research examined and linked the effects of social media application campaigns on public health awareness and the relationship between public awareness and behavioral health changes.

Second, we validated and tested the impact of the role of public awareness and public health behavioral changes as a mediating factor between the effects of the use of social media applications and public health protection, which can be considered as another contribution to the literature.

The research findings of this study also have significant implications for governmental health officials, health care professionals and practitioners, and other decision makers in health organizations. First, they should be fully aware of the importance of the use of social media campaigns (interventions) for public health awareness and of behavioral health changes to protect their communities and nations from the spread of pandemic diseases such as COVID-19. Second, the role of social media (innervations) to enhance public health awareness and behavioral health changes should be adequately considered in any choice of strategic health promotion plan. Social media campaigns should be considered as critical components of comprehensive approaches to improving public health behaviors.

Limitations

This study is not without limitations. The application of the conclusions of the study may be limited by time and geographical location. This study is a cross-sectional survey; the underlying identified associations may contrast across divisions and countries or may even lose their meaning over time. Most importantly, these data are self-reported by self-selected participants; also, the lockdown period was a constraint to gathering more representative data.

Conclusion

Our findings suggest that the use of social media platforms can positively influence awareness of public health behavioral changes and public protection against COVID-19. Public health authorities may use social media platforms as useful tools to increase public health awareness through the dissemination of brief messages to targeted populations. More research is needed to validate how social media channels can be used to improve health knowledge and adopt healthy behaviors in a cross-cultural context.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire (English version).
[DOCX File , 26 KB - jmir_v22i8e19996_app1.docx]

Multimedia Appendix 2
Properties of the final measurement model.
[DOCX File , 16 KB - jmir_v22i8e19996_app2.docx]

References

1. Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T. Social media and internet-based data in global systems for public health surveillance: a systematic review. Milbank Q 2014 Mar 06;92(1):7-33 [FREE Full text] [doi: 10.1111/1468-0009.12038] [Medline: 24597553]
2. Giustini DM, Ali SM, Fraser M, Kamel Boulos MN. Effective uses of social media in public health and medicine: a systematic review of systematic reviews. Online J Public Health Inform 2018 Sep 21;10(2):e215 [FREE Full text] [doi: 10.5210/ojphi.v10i2.8270] [Medline: 30349633]
3. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan 10;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
4. Chan JF, Yuan S, Kok K, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020 Feb;395(10223):514-523. [doi: 10.1016/s0140-6736(20)30154-9]
5. Epidemiological situation. Website in Arabic. Jordanian Ministry of Health. URL: https://corona.moh.gov.jo/ar [accessed 2020-08-12]
6. Allgaier J, Svalastog AL. The communication aspects of the Ebola virus disease outbreak in Western Africa--do we need to counter one, two, or many epidemics? Croat Med J 2015 Oct;56(5):496-499 [FREE Full text] [doi: 10.3325/cmj.2015.56.496] [Medline: 26526888]
7. Misra A, Sharma A, Shukla J. Stability analysis and optimal control of an epidemic model with awareness programs by media. Biosystems 2015 Dec;138:53-62. [doi: 10.1016/j.biosystems.2015.11.002] [Medline: 26551557]
8. Al-Surimi K, Khalifa M, Bahkali S, Ashraf EM, Househ M. The potential of social media and internet-based data in preventing and fighting infectious diseases: from the internet to twitter. Adv Exp Med Biol 2017;972:131-139. [doi: 10.1007/5584_2016_132]
9. Kumar A, Srivastava PK, Takeuchi Y. Modeling the role of information and limited optimal treatment on disease prevalence. J Theor Biol 2017 Feb 07;414:103-119. [doi: 10.1016/j.jtbi.2016.11.016] [Medline: 27890574]

10. Misra AK, Rai RK, Takeuchi Y. Modeling the control of infectious diseases: Effects of TV and social media advertisements. Math Biosci Eng 2018 Dec 01;15(6):1315-1343 [FREE Full text] [doi: 10.3934/mbe.2018061] [Medline: 30418788]
11. Gluskin RT, Johansson MA, Santillana M, Brownstein JS. Evaluation of Internet-based dengue query data: Google Dengue Trends. PLoS Negl Trop Dis 2014 Feb 27;8(2):e2713 [FREE Full text] [doi: 10.1371/journal.pntd.0002713] [Medline: 24587465]
12. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
13. Kouri P, Rissanen M, Weber P, Park H. Competences in Social Media Use in the Area of Health and Healthcare. Stud Health Technol Inform 2017;232:183-193. [Medline: 28106596]
14. Dubey B, Dubey P, Dubey US. Role of media and treatment on an SIR model. NAMC 2016 Mar 25;21(2):185-200. [doi: 10.15388/na.2016.2.3]
15. Benetoli A, Chen T, Aslani P. How patients' use of social media impacts their interactions with healthcare professionals. Patient Educ Couns 2018 Mar;101(3):439-444. [doi: 10.1016/j.pec.2017.08.015] [Medline: 28882545]
16. Jowett M, Brunel M, Flores G, Cylus J. Spending targets for health: no magic number. World Health Organization. 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/250048/WHO-HIS-HGF-HFWorkingPaper-16.1-eng.pdf?sequence=1 [accessed 2020-08-12]
17. Porumbescu GA. Linking public sector social media and e-government website use to trust in government. Gov Inf Q 2016 Apr;33(2):291-304. [doi: 10.1016/j.giq.2016.04.006]
18. Freberg K, Palenchar MJ, Veil SR. Managing and sharing H1N1 crisis information using social media bookmarking services. Public Relations Review 2013 Sep;39(3):178-184. [doi: 10.1016/j.pubrev.2013.02.007]
19. Huesch MD, Galstyan A, Ong MK, Doctor JN. Using Social Media, Online Social Networks, and Internet Search as Platforms for Public Health Interventions: A Pilot Study. Health Serv Res 2016 Jun 10;51 Suppl 2:1273-1290 [FREE Full text] [doi: 10.1111/1475-6773.12496] [Medline: 27161093]
20. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009 Apr;30(1):273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
21. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
22. Ellis LA, Collin P, Davenport TA, Hurley PJ, Burns JM, Hickie IB. Young men, mental health, and technology: implications for service design and delivery in the digital age. J Med Internet Res 2012 Nov 22;14(6):e160 [FREE Full text] [doi: 10.2196/jmir.2291] [Medline: 23171827]
23. Collinson S, Khan K, Heffernan JM. The Effects of Media Reports on Disease Spread and Important Public Health Measurements. PLoS One 2015 Nov 3;10(11):e0141423 [FREE Full text] [doi: 10.1371/journal.pone.0141423] [Medline: 26528909]
24. Hassan Zadeh A, Zolbanin HM, Sharda R, Delen D. Social Media for Nowcasting Flu Activity: Spatio-Temporal Big Data Analysis. Inf Syst Front 2019 Jan 5;21(4):743-760. [doi: 10.1007/s10796-018-9893-0]
25. Samanta S, Rana S, Sharma A, Misra A, Chattopadhyay J. Effect of awareness programs by media on the epidemic outbreaks: A mathematical model. Appl Math Comput 2013 Feb;219(12):6965-6977. [doi: 10.1016/j.amc.2013.01.009]
26. Ferguson N. Capturing human behaviour. Nature 2007 Apr 12;446(7137):733-733. [doi: 10.1038/446733a] [Medline: 17429381]
27. Buonomo B, D'Onofrio A, Lacitignola D. Global stability of an SIR epidemic model with information dependent vaccination. Math Biosci 2008 Nov;216(1):9-16. [doi: 10.1016/j.mbs.2008.07.011] [Medline: 18725233]
28. Ye X, Li S, Yang X, Lee J, Wu L. The Fear of Ebola: A Tale of Two Cities in China. In: Ye X, Li S, Yang X, editors. Big Data Support of Urban Planning and Management. Advances in Geographic Information Science. Cham, Switzerland: Springer; 2018:112-132.
29. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic-analysis of Facebook as a social media health information platform. Am J Infect Control 2017 Mar 01;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
30. Villar ME, Marsh E. Social Media and Infectious Disease Perceptions in a Multicultural Society. In: Bilge N, Marino MI, editors. Reconceptualizing New Media and Intercultural Communication in a Networked Society. Pennsylvania, US: IGI Global; 2018:350.
31. Lunn PD, Belton CA, Lavin C, McGowan FP, Timmons S, Robertson DA. Using Behavioral Science to help fight the Coronavirus. JBPA 2020 Mar 29;3(1):online. [doi: 10.30636/jbpa.31.147]
32. Coronavirus disease (COVID-2019) situation reports. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-08-12]
33. Commitments to Responsible Use of Antimicrobials in Humans. World Health Organization. 2014 Nov. URL: https://www.who.int/drugresistance/events/Oslomeeting/en/ [accessed 2020-08-12]

34. Theory at a Glance – A Guide For Health Promotion Practice (Second Edition). U.S. Department of Health and Human Services, National Institutes of Health. 2015. URL: http://sbccimplementationkits.org/demandrmnch/ikitresources/theory-at-a-glance-a-guide-for-health-promotion-practice-second-edition/ [accessed 2020-08-12]
35. Behaviour change and antibiotic prescribing in healthcare settings: Literature review and behavioural analysis. Public Health, England. 2015. URL: https://www.gov.uk/government/publications/antibiotic-prescribing-and-behaviour-change-in-healthcare-settings [accessed 2020-08-12]
36. Laranjo L. Social Media and Health Behavior Change. In: Syed-Abdul S, Gabarron E, Lau AYS, Househ M, editors. Participatory Health Through Social Media. Cambridge, MA: Academic Press; 2016:83-111.
37. Laverack G. The Challenge of Behaviour Change and Health Promotion. Challenges 2017 Oct 17;8(2):25. [doi: 10.3390/challe8020025]
38. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16(1):442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
39. Pallant J. SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows (Version 12). Chicago, IL: Open University Press; Dec 01, 2004.
40. Sekaran U, Bougie R. Research Methods for Business: A Skill-Building Approach. New York, NY: Wiley; 2013.
41. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
42. Bagozzi R, Yi Y. On the evaluation of structural equation models. JAMS 1988 Mar;16(1):74-94. [doi: 10.1007/BF02723327]
43. Hair J, Black W, Babin B, Anderson R, Tatham R. Multivariate Data Analysis, 6th Edition. Upper Saddle River, NJ: Prentice-Hall; 2006.
44. Creswell WJ. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (3rd ed.). Newbury Park, CA: Sage Publications; 2009.
45. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. J Mark Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
46. Pagoto S, Waring ME, Xu R. A Call for a Public Health Agenda for Social Media Research. J Med Internet Res 2019 Dec 19;21(12):e16661 [FREE Full text] [doi: 10.2196/16661] [Medline: 31855185]
47. Ahmad AR, Murad HR. The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
48. Alqutob R, Al Nsour M, Tarawneh MR, Ajlouni M, Khader Y, Aqel I, et al. COVID-19 Crisis in Jordan: Response, Scenarios, Strategies, and Recommendations. JMIR Public Health Surveill 2020 Jul 07;6(3):e19332 [FREE Full text] [doi: 10.2196/19332] [Medline: 32407289]
49. Fagherazzi G, Goetzinger C, Rashid MY, Aguayo GA, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
50. Lin Y, Hu Z, Alias H, Wong LP. Influence of Mass and Social Media on Psychobehavioral Responses Among Medical Students During the Downward Trend of COVID-19 in Fujian, China: Cross-Sectional Study. J Med Internet Res 2020 Jul 20;22(7):e19982 [FREE Full text] [doi: 10.2196/19982] [Medline: 32584779]

Abbreviations

AVE: average variance extracted
CFA: confirmatory factor analysis
COVID-19: coronavirus disease
e-government: electronic government
H1: hypothesis 1
H2: hypothesis 2
H3: hypothesis 3
H4: hypothesis 4
H5: hypothesis 5
H6: hypothesis 6
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach, G Fagherazzi; submitted 08.05.20; peer-reviewed by DA Hamed, DR Habash, DR Brghuthi, PA Ahmed, P Weber; comments to author 13.07.20; revised version received 16.07.20; accepted 23.07.20; published 19.08.20.

Please cite as:
Al-Dmour H, Masa'deh R, Salman A, Abuhashesh M, Al-Dmour R
Influence of Social Media Platforms on Public Health Protection Against the COVID-19 Pandemic via the Mediating Effects of Public Health Awareness and Behavioral Changes: Integrated Model
J Med Internet Res 2020;22(8):e19996
URL: http://www.jmir.org/2020/8/e19996/
doi:10.2196/19996
PMID:32750004

Original Paper

Applying and Extending the FITT Framework to Identify the Challenges and Opportunities of Successful eHealth Services for Patient Self-Management: Qualitative Interview Study

Sari Kujala[1], PhD; Elske Ammenwerth[2], MET, Dr sc hum; Heta Kolanen[3], MHS; Minna Ervast[4], MHS, MSc Econ

[1]Department of Computer Science, Aalto University, Espoo, Finland

[2]UMIT – Private University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria

[3]HUS IT Management, HUS Helsinki University Hospital, Helsinki, Finland

[4]Development Unit, Hospital District of South-West Finland, Turku, Finland

Corresponding Author:
Sari Kujala, PhD
Department of Computer Science
Aalto University
Konemiehentie 2
Espoo
Finland
Phone: 358 50 3862768
Email: sari.kujala@aalto.fi

Abstract

Background: The number of public eHealth services that support patient self-management is rapidly increasing. However, the implementation of these eHealth services for self-management has encountered challenges.

Objective: The purpose of this paper was to analyze the challenges and opportunities of implementing eHealth services for self-management by focusing on the fit between the technical solution and clinical use.

Methods: We performed in-depth interviews with 10 clinical project coordinators and managers who were responsible for developing and implementing various eHealth services for self-management interventions in five university hospitals in Finland. The results were analyzed using content analysis and open coding. The Fit between Individuals, Task, and Technology (FITT) framework was used to interpret the findings.

Results: The implementation of self-management services involved many challenges related to technical problems, health professional acceptance, patient motivation, and health organization and management. The implementers identified practices to manage the identified challenges, including improving the design of the technology, supporting health professionals in the adoption of the eHealth services, changing the work processes and tasks, involving patients, and collectively planning the implementation inside an organization. The findings could be mostly attributed to the dimensions of the FITT framework.

Conclusions: The FITT framework helped to analyze the challenges related to the implementation, and most of them were related to poor fit. The importance of patients as stakeholders in eHealth services for patient self-management needs to be highlighted. Thus, we propose that patients should be added as a different type of individual dimension to the FITT framework. In addition, the framework could be extended to include organization and management in a new context dimension.

(J Med Internet Res 2020;22(8):e17696) doi:10.2196/17696

KEYWORDS

interview; implementation; adoption; patient self-management; organization

Introduction

eHealth for Self-Management

In many countries, the number of public eHealth services that support patient self-management is rapidly increasing. Barlow et al [1] define self-management as an individual's ability to manage the symptoms, treatment, physical and psychosocial consequences, and life changes inherent in living with a chronic condition. According to their review, self-management

interventions benefit participants' well-being. Diabetes and heart failure interventions seem to be particularly effective [2].

However, many studies also report that the implementation of eHealth services that support self-management has encountered challenges, such as motivating patients and health professionals [3-6]. The implementation challenges have also led to limited adoption by patients and health professionals and their use of the services [3,4,7].

Health care professionals have many concerns related to self-management services and their professional roles in these new situations [8,9]. They especially worry about whether patients are willing and able to use these new services. Thus, they may not be willing to introduce new eHealth services to patients, but their endorsement increases patients' trust in a technical solution [10] and greatly impacts patients' initial adoption and continued use of eHealth services [11].

Several literature reviews have identified factors that facilitate or hinder the successful implementation of eHealth services [12-14]. While these studies have identified many good implementation practices, such as the importance of leadership support, the suitability of the practices and approaches depends on the context [12,15]. The self-management context has rarely been studied. Thus, this study sought to gain deeper knowledge in this area through a qualitative interview study.

Fit Between Individuals, Task, and Technology Model

Several implementation models can be used to analyze barriers and facilitators that occur during implementation [16]. We applied the *Fit between Individuals, Task, and Technology* (FITT) framework developed by Ammenwerth et al [17], as it helps analyze the factors that influence the success or failure of information technology (IT) implementation in a health care setting. The FITT framework has already been shown to be useful in several case studies [18-21].

The FITT framework is based on the idea that IT adoption in a clinical environment depends on the optimal fit or interaction between the attributes of three fit dimensions: the individual users, the technology, and the clinical tasks and processes [17]. An individual represents a single user or a user group. Technology includes the interaction of the various tools needed to accomplish a given task. The task comprises all working processes and tasks that need to be completed. Organizational aspects are included either as part of the individual dimension or part of the task dimension. One of the FITT framework's specific strengths is its focus on the interaction between the user and the task [20], as issues related to IT support of professionals' workflows are the most frequently reported failure factors of eHealth interventions [22].

Compared to other frameworks, such as the Consolidated Framework for Implementation Research (CFIR) [23], the FITT framework differs in its addition of the interaction aspect. While the CFIR suggests that the characteristics of individuals and interventions influence implementation, the FITT framework considers the interactions between the characteristics of individuals and technical interventions. According to the FITT framework, the influence of technological interventions also depends on the individuals' motivation, knowledge, and training. Tsiknakis and Kouroubali [18] used the FITT framework in their case study and reported that the model provided a structured way to explain the reasons for the success or failure of IT systems and eHealth services. Prgomet et al [19] also found that the health professionals' use of technology was related to the fit between users, tasks, and technology. However, they identified additional environmental factors, such as the temporal rhythms of a ward or space limitation, and proposed that the FITT framework should be extended to include an environment dimension as well. In addition, the FITT framework has been successfully used to analyze different stakeholders' perceptions of eHealth, such as those of nurses [24], case managers [20], and patients [25].

In summary, the FITT framework has been useful in analyzing implementation barriers and facilitators in different case studies. However, self-management eHealth services are novel for patients and health care professionals and require complex changes in clinical care [6]. Both patients and health care professionals need to be motivated to use these services, and there remains a limited understanding of how these services should be implemented.

This paper applies the FITT framework to analyze the implementation experiences of 10 clinical project coordinators and managers who were responsible for implementing digital care paths supporting patient self-management in five university hospitals. In a study by Murray et al [26], staff charged with implementing eHealth initiatives had a deep understanding of the barriers to and facilitators of success. Thus, we specifically collected information from the clinical project coordinators and managers who were responsible for developing and implementing various eHealth services for self-management. As the implementation of eHealth services was in the early phase and health care professionals had a key role in endorsing and engaging patients [9,10], project coordinators focused more on introducing the services to health care professionals than to patients.

Study Aims

The aims of this study were (1) to identify the specific challenges to implementing eHealth services for self-management and opportunities to manage these challenges and (2) to evaluate how well the FITT framework explains the identified challenges in the context of self-management eHealth services. The findings provide a better understanding of the factors that influence the implementation of eHealth services for self-management interventions and extend the FITT framework to explain the success of eHealth service adoption.

Methods

Study Setting

In Finland, the objective of the national eHealth and eSocial Strategy 2020 is to support the active role of citizens in promoting their own well-being, preventing health problems, self-assessing the need for services, and independent coping [27]. As a part of the strategy, an eHealth portal, HealthVillage.fi, was developed by the joint Virtual Hospital 2.0 project among five Finnish university hospitals. The project

was funded by the hospitals and the Ministry of Social Affairs and Health. The Virtual Hospital 2.0 project was funded from 2016 to 2019 and was coordinated by the HUS Helsinki University Hospital (referred to simply as HUS). The goal was to raise the quality of specialist health care services and improve their accessibility with the use of digital technology.

The coordinating HUS developed the technical platform for developing eHealth services for citizens, patients, and professionals. The joint project provided guidelines for planning the content and implementation. Using the guidelines, the services were developed by a multi-professional team usually including physicians and nurses.

The eHealth portal was developed in two phases. First, an open-access eHealth portal was developed to offer information, advice, self-care instructions, and symptom navigators for Finnish citizens and patients. The portal includes over 20 eHealth services, called hubs, that focus on specific patient or disease groups, such as neurological diseases.

Second, hospital-specific digital care paths only open to invited patients of a care unit were developed [28], and the first two paths were opened in November 2017. The digital care paths were designed to supplement and offer alternatives to the traditional treatment paths. The functionalities depended on the patient group, but they included patient instructions, exercises, self-monitoring and symptom assessment, and secure messaging. A digital path could be for short-term treatment, such as surgery, or for longer-term care of a chronic disease. The team planned the functionalities and developed the content. The project coordinators were trained for their positions, and they were responsible for adding the content to the platform and implementing the new service. Each project coordinator was supported by a developer partner from an IT organization.

Table 1 summarizes the digital care paths and their functions for which the participants were responsible for implementing in their organizations. Three of the digital care paths were to support patients during preparation for a surgery—mitral valve surgery, cervical spine surgery, and pacemaker surgery—and during postsurgical care. Three were short-term digital care paths for couples receiving in vitro fertilization treatment, pregnant women, and women with gestational diabetes. Two were long-term digital care paths for patients with spinal cord disability and rheumatism.

According to the project coordinators, the main goals of the new digital care paths were to improve the quality of service provided to patients and to support self-management and communication between patients and care personnel. Patients were expected to be better informed and require less guidance when using the paths. Nurses could also receive information from patients, monitor them, and perform preventive interventions when needed. Organizations also aimed to minimize costs by reducing the number of phone calls and moving the communication to digital messaging.

As the digital care paths had been used from 1 to 10 months, the technical platform was still under development, the number of patients using a service remained low, and all services were in the early stage of implementation. The development of the technical platform was continuously developed based on professionals' and patients' feedback using an agile approach.

Table 1. Description of digital care paths.

Digital care path	Functions	Starting date (first patient entered)	Estimated potential number of patients	Real number of patients between the starting date and 11/3/2019 (path ended or still ongoing)
Mitral valve surgery	Information Health questionnaires	4/2018	150 per year	177
Cervical spine surgery	Messaging Self-management information Appointments Anamneses forms	10/2018 piloting	500 per year	244
Pacemaker surgery	Information before pacemaker implantation and answers to the most frequently asked questions Messaging	12/2018	300 per year	214
In vitro fertilization	Appointments and questionnaires Information and instructions Messaging	11/2017	400 per year	597
Pregnancy	Appointments Information on practicalities and fetal screening	11/2017	10,000 per year	17,229
Gestational diabetes	Messaging, self-management information, diary, tasks, and tests Mobile app	1/2019	200 per year	239
Spinal cord disability	Messaging and sending pictures Anamneses forms Ability-to-function forms Symptom diaries Remote consultations	10/2018 piloting	200 per year	200
Rheumatology	Information about the clinic and care and answers to the most frequently asked questions Messaging	1/2019	Thousands per year	1951

Study Participants

The participants were selected by purposive sampling. The goal was to cover a variety of experiences of different eHealth services and contexts from a project management point of view. The inclusion criterion was a responsibility to implement a digital care path in a care unit. Implementers were chosen for this study, as they have experience planning and managing implementation, and they observe factors that influence implementation [26]. The participants were recruited by a development manager of HUS who had contact with project coordinators and managers.

A total of 10 participants were interviewed from five hospitals (see Table 2). In total, 7 of them were nurse project coordinators who were responsible for the practical implementation of the digital care paths in their unit. In addition, 1 was a physician project manager who was leading an implementation project alongside her clinical work. As one of the hospitals had decided to postpone the development and implementation of the digital care paths, their development manager and 1 technical manager were also selected to be interviewed to reveal their experiences.

Table 2. Details about the study participants.

No.	Role	Expertise	Responsibility	Hospital ID
1	Development manager	Economics, change management, and implementation	Managing digitalization of health services	1
2	Technical product owner and project coordinator	Software engineering	Project manager of digital care paths	1
3	Physician project manager	Medicine	Planning and implementing a digital care path	2
4	Nurse project coordinator	Nursing	Planning and implementing a digital care path	2
5	Nurse project coordinator	Nursing	Planning and implementing a digital care path	2
6	Nurse project coordinator	Nursing	Planning and implementing a digital care path	2
7	Nurse project coordinator	Nursing	Planning and implementing a digital care path	2
8	Nurse project coordinator	Nursing	Planning and implementing a digital care path	3
9	Nurse project coordinator	Nursing	Planning and implementing a digital care path	4
10	Nurse project coordinator	Nursing	Planning and implementing a digital care path	5

All the participants were women, and their age ranged from 33 to 53 years. None of the project coordinators or the project manager had previous experience developing or implementing eHealth services.

Data Collection and Analysis

One interviewer completed semistructured interviews with each participant. In total, 2 participants wanted to have a pair interview. The interviewer met the participants in their office or performed the interviews through videoconferencing. The interview included questions from two main themes:

1. The challenges of the implementation.
2. Opportunities to manage these implementation challenges.

In addition, background information about the interviewee; information about the digital care path, planning, and implementation; and patient feedback were collected.

The interviews were conducted by the first author (SK) from May 2018 to November 2019. The interviews lasted from 1 to 2 hours, and they were audio recorded and transcribed for analysis. In addition to interview data, a documented patient feedback survey report was used as an information source.

After each interview, the main observations were recorded as notes. The analysis was performed in two stages. In stage 1, open coding was used to identify themes in the data. Using in vivo coding, the respondents' words were used to define the themes to ensure that the themes represented the original meaning of the respondents. One of the authors created a coding scheme using a subset of four interviews. The coding scheme was checked by two other authors before it was used to code the rest of the interviews. Any new themes that emerged in subsequent coding were added to the coding scheme. Finally, the number of interviewees mentioning a theme was calculated. As the development manager and technical manager were interviewed together and shared experiences in the same hospital that postponed the implementation of digital care paths, the analysis of their responses was combined.

In stage 2, the FITT framework [17] was used as a deductive coding framework [29] to place the identified themes in a theory context. The themes identified in stage 2 were categorized into the FITT framework dimensions of task, technology, and individuals, as shown in Tables 3 and 4. The attributes of the dimensions identified by Ammenwerth et al [17] were used to support this categorization.

Table 3. Challenges of implementing eHealth services for self-management.

Dimension and themes	Mentions (n=9), n (%)
Individual-technology fit: health professionals	
Problems with usability, technical problems, and missing functionalities	8 (89)
Resistance, lack of use, and difficulty changing professionals' work processes	7 (78)
Professionals not informing patients about the eHealth services	3 (33)
Negative previous experiences with information systems	1 (11)
Lack of training	1 (11)
Lack of technical support	1 (11)
Individual-technology fit: patients	
Problems with usability and missing functionalities	7 (78)
Lack of use	4 (44)
Lack of active patient participation in planning	1 (11)
Health professional–task fit	
Extra work caused by insufficient interoperability	3 (33)
Bad fit with the work processes	2 (22)
Organization and management–technology fit	
Lack of knowledge about the possible technical functionalities	3 (33)
Lack of resources	3 (33)
Lack of management support	2 (22)
Unclear roles during implementation	1 (11)
Failure of the initial technical platform to fit the organization's goals and processes	1 (11)

Table 4. Practices for managing the challenges of implementation.

Dimension and practices	Mentions (n=9), n (%)
Individual-technology fit: health professionals	
Involving all the stakeholders, professional groups, and a technical expert in planning the service	2 (22)
Testing and piloting the eHealth services before implementation	4 (44)
Repeatedly informing health professionals about the implementation, changes, and the influence of the new services to their work for an extended period via unit meetings, training, personal contacts, and laminated instructions	4 (44)
Proactively responding to health professionals' concerns	1 (11)
Involving frontline leaders and health professionals in planning the services and implementation is needed to create buy-in	3 (33)
Providing adequate introductory knowledge, repeated training, and personal guidance as well as a test environment, which is required to train professionals	4 (44)
Training a superuser to encourage health professionals and support implementation	1 (11)
Proceeding slowly and gradually, so professionals have time to adjust to and practice using the new services	2 (22)
Reserving extra personnel resources and time for the changing tasks	1 (11)
Providing technical support with a responsible person when needed	1 (11)
Individual-technology fit: patients	
Identifying a patient group that can benefit from an eHealth service and having a patient point of view	2 (22)
Involving patients early on and creating new methods needed to motivate patient participation	2 (22)
Informing patients in an interesting way and providing leaflets or other marketing materials	2 (22)
Collecting constant feedback from patients and improving the service through the use of questionnaires, contacting patients for further details to create a partnership, and request for feedback from patients that did not use the service	3 (33)
Encouraging health professionals to discuss the digital service when meeting a patient and to reserve digital appointments with the patient	2 (22)
Offering technical support during problems	1 (11)
Health professional–task fit	
Identifying the current work processes and needs	1 (11)
Fitting the eHealth service plans to current care processes to support and ease health professionals' work, reduce the number of phone calls, increase remote work, and increase interoperability of the systems so that health professionals can view patient information from one system and do not need to record the same information twice	5 (56)
Ensuring the service is quick, easy, and effortless to request and use	5 (56)
Planning changes in the work processes well in advance and piloting the services to test the new processes and show the benefits of the service	1 (11)
Organization and management–technology fit	
Identifying the needs early	1 (11)
Fitting the eHealth service plans to the technical possibilities, including demonstrations and examples of existing services to help illustrate the possibilities	2 (22)
Evaluating the work cost of implementation, a responsible person, resources needed, and the potential benefits	2 (22)
Planning the implementation carefully in advance and defining the roles and responsibilities of the participants, changes, ways of relieving resistance, and solutions to problems	5 (56)
Involving an active multi-professional team	1 (11)
Indicating more than one person as a spokesperson to support the implementation, especially frontline leaders, who were considered important in influencing their subordinates' commitment and providing resources for implementation	2 (22)
Involving and informing top management to provide support and resources and ensure the availability of spaces and devices	2 (22)
Identifying measures of impacts and making baseline measurements in the very beginning	1 (11)

Ethics

The interviewees received oral and written information about the study and its voluntary nature before the interviews. Written informed consent was obtained from all participants. The study protocol was reviewed and approved by the Ethical Review Board of Aalto University, Finland.

Results

Challenges of Implementing eHealth Services for Self-Management

Table 3 summarizes the challenges that the participants experienced in implementing the eHealth services. We categorized the challenges according to the FITT model dimensions. Most of the challenges were related to poor fit between individuals and technology, such as health professionals suffering usability problems, technical problems, and missing functionalities. Health professionals were reported to be critical of new services because they had had negative experiences with information systems. One of the first digital paths had a challenging start, as nurses needed to solve the usability problems that patients faced. Nurses had no training on how to use the service, and technical support was not yet available.

Many of the health professionals were not willing to use the new eHealth services. One of the participants described, "The resistance over changing practices surprised me most and how long it has been continued." She also said that not all the nurses understood how this service was going to help with their work; hence, their motivation to use it was low. Only 3 of the participants did not mention any professionals' resistance or lack of use, but their eHealth services were only used by the developing team, which consisted of 2-10 health professionals. Thus, it seemed to be more challenging to introduce a service to larger user groups. In total, 3 participants mentioned that the group of health professionals who had not participated in the development team demonstrated the most resistance.

One interviewee described that it was very challenging to reach and inform all the health care professionals as they worked in three shifts. Physicians forgot to use the digital path as a tool and tell their patients about the new service. In addition, another participant said that physicians used the old paper-based approach and were reluctant to use the digital path, which appeared to be slower. Nurses also forgot to reserve digital appointments for the patients.

The services included a patient feedback survey, and all except one implementation project received feedback from patients through the survey. The feedback provided was positive, and the services were evaluated to be useful. For example, the information received was seen to support preparation for an operation. Patients gave negative feedback related to difficult registration, problems in use, slow or confusing services, and missing functionalities.

In the documented survey, 15 out of 17 patients (88%) rated that they were *satisfied* or *rather satisfied* with the service, and all of them agreed that the service was useful. However, they reported missing instructions, impractical registration, slowness, and other difficulties that made using the service cumbersome or disrupted the service.

The number of patient users was also low in some cases. Health professionals often did not actively use the new eHealth services, and they were also passive in introducing the services to patients and motivating them to use the services. Patients did not provide much feedback, and the number of patients who completed the questionnaires was low. One of the interviewees said that they were not able to find patients who were willing to participate in the design workshops. Thus, in the context of online self-management intervention, patients need to be engaged in both using and designing the services. We, therefore, divided the individual dimension to separately cover health professionals and patients.

Some participants identified fit problems between technology and tasks. Mostly, the lack of interoperability created extra work for health professionals, but in 2 cases, participants identified that the process required by the eHealth service did not fit the work processes.

In addition to the FITT model's dimensions, there were challenges related to poor fit between the organization and management and the technology. For example, it was unclear how the technical platform could be used in the organization and what benefits exist. In some cases, there was insufficient management support or available resources for implementation. For example, one of the interviewees found it challenging to support others in the implementation, as she did not have reserved work hours for it. She felt that management should also have provided resources for implementation and use and not only for the development of the digital path.

The development of the technical platform was in one hospital at the beginning; the initial version of the platform did not fit the organization's goals and processes. The usability and interoperability of the customer management system was considered to be poor. Consequently, the implementation of any new eHealth services using the technical platform in this hospital was halted until the platform was further developed.

Opportunities to Manage the Challenges of Implementing eHealth Services for Self-Management

Table 4 shows the practices that the participants found successful or recommended for managing the identified challenges and supporting the implementation of eHealth services. Many of the practices were related to improving the design of the eHealth services by involving all the stakeholders, testing and piloting the service, and collecting feedback. Management practices, such as informing health professionals, responding to their concerns, and supporting them in the change, reduced health professional resistance. In addition, involving and informing patients and collecting their feedback was considered important.

Participants found that task-technology fit could be improved by aligning the new service to the current work processes and needs and designing a service that is usable and effortless to use. To improve the fit between the organization and the technology, participants saw a need to identify a relevant problem that the technology was able to solve and the benefits that could be produced and to evaluate the costs related to

implementation. Many participants also learned that implementation needed to be well planned and organized collectively inside the organization.

Discussion

Principal Findings

The study shows three major findings:

1. The implementation of self-management services involved many challenges related to technical problems, health professional acceptance, patient motivation, and health organization and management.
2. The implementers had identified practices to manage the identified challenges by improving the design of the technology, supporting health professionals in the technology adoption, changing the work processes and tasks, involving patients, and planning the implementation collectively inside the organization.
3. The challenges and practices could be mostly attributed to the dimensions of the FITT framework. However, the findings suggest that patients should be added as a different type of individual dimension, and organizations and management should be added as new dimensions to the FITT framework.

Challenges of Implementing eHealth Services for Self-Management

In line with an earlier case study [6], it was challenging to introduce self-management services to health care professionals, change their work practices, and motivate patients to use the service. The poor fit of tasks, the technology, and individuals troubled health care professionals. As shown by previous studies, these issues are remarkably common barriers to eHealth interventions [22].

Furthermore, the technology needs also fit the organization, and management has an important role, as identified in previous studies [12,30]. The technical solution fit the strategy and operation model of the organization, and successful commitment requires commitment and resources from management. Many of the challenges were interrelated. For example, if the new service did not fit the work processes of the health professionals, they did not inform patients about the service, which resulted in a lack of use among both professionals and patients.

Opportunities to Manage the Challenges of Implementing eHealth Services for Self-Management

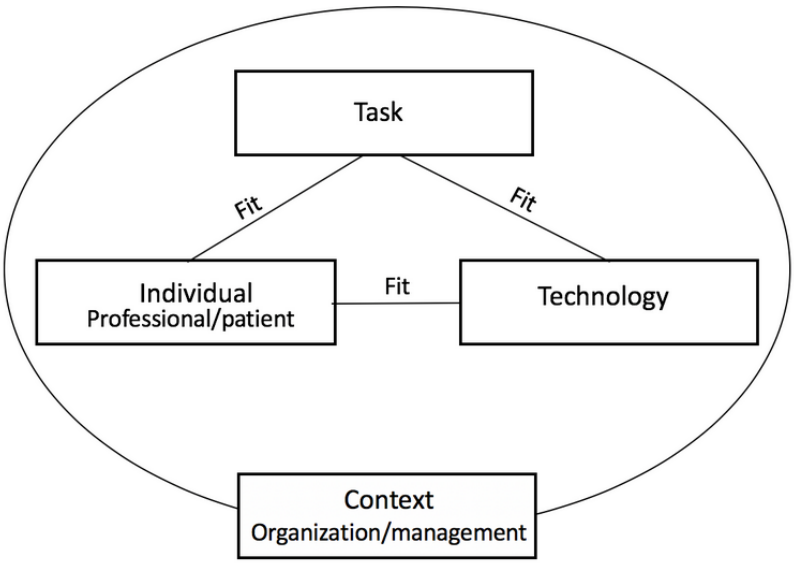
Most of the identified opportunities to handle the challenges were related to improving the fit between tasks, the technology, and individuals. Better fit can be achieved by designing better technical solutions and usability to support patients, health professionals, and tasks. Therefore, it is important to foster stakeholder involvement and user-centered design in future projects [31-33]. Current work processes and the needs of users should be understood at the beginning of the design. Different stakeholders also need to be involved to identify their needs and encourage buy-in. In the studied cases, health professionals served as designers and implementers. Thus, health professionals were involved, but this situation also created new challenges, as the professionals had no experience in user-centered design or implementations.

In addition, many of the opportunities to handle the identified challenges were related to the management and the organization. As shown by previous studies, the way eHealth services are introduced is also important for implementation success [8,12]. Top management support and a shared strategic vision of the new technology are often reported to support implementation [30], but the results from this study also pointed out the need for the collective support of frontline leaders and other spokespersons.

Extension of the FITT Framework to Better Cover the Patient as an Individual and Organizational Context

Overall, the findings were well interpreted with the dimensions of the FITT framework (see Figure 1). However, the role of patient stakeholders in self-management needs to be highlighted, and we propose that both health professionals and patients are represented in the individual dimension.

Figure 1. Extended FITT (Fit between Individuals, Task, and Technology) framework.



In the original FITT framework, the organization is mentioned as one of the attributes of the individual level. However, previous research shows that in complex implementation projects, there are also higher-level contextual factors, such as the health care system, the social climate, and economic and political issues, that are important to consider [34,35]. Additionally, many of the identified challenges and opportunities in this study were not related to individuals but to organizations and management in a broader context. For example, if management did not provide resources for implementation, training, and planning, individuals were not supported to change their work processes. Furthermore, our results imply that it is challenging for an individual to manage implementation. Instead, the implementation work and introduction of a service to patients need to be supported collectively.

Thus, we propose that organizations and management be represented by a separate context dimension, as shown in Figure 1. Unlike the findings of Prgomet et al [19], our findings did not demonstrate environmental issues, which is probably because we did not observe professionals' behaviors in detail. However, we suggest that the environment is a part of the new context dimension.

The extensions we proposed to the FITT model are not specific to self-management services. However, these kinds of services influence health professionals' and patients' interactions in a novel way; thus, the extensions especially support the analysis of the organizational efforts to engage patients in care.

The combination of the FITT framework and the Clinical Adoption Meta-Model [36] may help to further explain the progress of the implementation from one phase to the next over time. In this study, one organization did not proceed to the initial phase, where the service was available to professionals and patients, because the technology fit poorly with the organization's strategy and operational model, as interoperability and information safety were not considered acceptable. Other organizations made the new services available, but there were challenges in initiating use of the services due to usability and interoperability reasons. One of the services was also challenging to use, as it did not fit the work processes and needed to be redesigned. Without sufficient use, there were no clinical or health behavior changes or positive outcomes.

Comparison With Prior Work

Our results highlight the complex relationship between different factors and stakeholders that influence the success of implementation. The larger the number of users, the more complicated the implementation was in this study. In a literature review, Ludwick and Doucette [37] suggested that large organizations should use an incremental approach due to the complexities associated with their size. In addition to complexity, the larger number of users made participation in planning the new services and the related work processes more difficult. As Granja et al [22] identified in their literature review, this study observed that it was especially challenging to change health professionals' work processes. Our results suggest that the participation of health professionals in the change of their work processes could support the change process.

In the case of eHealth services for self-management, patients are important stakeholders. Urowitz et al [3] found that patients were not always active in an online diabetes self-management portal, and their use declined over time. Thus, patients should be involved early to ensure the quality of the eHealth services and should be supported during adoption of the services. Health professionals play an important role in endorsing the new services to patients [10].

Our results imply that although many of the challenges of implementation are similar to those of different eHealth innovations [14,38], new approaches are clearly needed to handle the challenges and lack of fit in practice. In the self-management context, the fit between individuals and the technology needs to be improved both from health care professionals' and patients' points of view. Both groups need to be involved in the design, and they should be well informed. Their feedback should also be constantly collected. In improving the fit between the health care professionals and the tasks, the goal should be to improve the current work processes and ease health care professionals' work.

Limitations

A limitation of this study is that we only studied the perspectives of project coordinators and managers who were responsible for implementing the new online self-management services. As our study relied on the implementers' reporting of health care professionals' reactions, the health care professionals' characteristics influencing the implementation were not identified, as they were identified in an interview study by Ciere et al [6]. To gain a more comprehensive understanding of the challenges and practices, other stakeholders, such as patients and health care professionals, should be studied in the future as well.

The number of interviewees was relatively low. However, the goal was to gain qualitative insight into the challenges and opportunities in the implementation projects, and data saturation was reached, meaning that not much new information (eg, a single new theme) could be identified in an additional interview [39]. The participants were also relatively inexperienced in implementing eHealth services, and the hospitals were all in Finland. Implementers working in other organizations may face different implementation challenges, and the generalizability of this study is restricted to this specific context.

The level of health care professionals' resistance can vary during implementation [40-42]. Our study provided only a cross-section of the implementation process, and a better understanding of the timing of the challenges and best implementation practices are needed. Further studies should clarify how health care professionals and especially patients can be better supported in the adoption of eHealth services.

Acknowledgments

The authors are grateful to the participating health organizations and interviewees. This work was supported by the Strategic Research Council at the Academy of Finland (decision number 327147).

Conflicts of Interest

None declared.

References

1. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: A review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
2. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: A systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
3. Urowitz S, Wiljer D, Dupak K, Kuehner Z, Leonard K, Lovrics E, et al. Improving diabetes management with a patient portal: A qualitative study of diabetes self-management portal. J Med Internet Res 2012 Nov 30;14(6):e158 [FREE Full text] [doi: 10.2196/jmir.2265] [Medline: 23195925]
4. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: Health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15 Suppl 2:183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
5. Harvey J, Dopson S, McManus RJ, Powell J. Factors influencing the adoption of self-management solutions: An interpretive synthesis of the literature on stakeholder experiences. Implement Sci 2015 Nov 13;10(1):159 [FREE Full text] [doi: 10.1186/s13012-015-0350-x] [Medline: 26566623]
6. Ciere Y, van der Vaart R, van der Meulen-De Jong A, Maljaars P, van Buul A, Koopmans J, et al. Implementation of an eHealth self-management care path for chronic somatic conditions. Clin eHealth 2019;2:3-11. [doi: 10.1016/j.ceh.2019.04.001]
7. Fiks AG, DuRivage N, Mayne SL, Finch S, Ross ME, Giacomini K, et al. Adoption of a portal for the primary care management of pediatric asthma: A mixed-methods implementation study. J Med Internet Res 2016 Jun 29;18(6):e172 [FREE Full text] [doi: 10.2196/jmir.5610] [Medline: 27357835]
8. Kujala S, Hörhammer I, Kaipio J, Heponiemi T. Health professionals' expectations of a national patient portal for self-management. Int J Med Inform 2018 Sep;117:82-87 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.06.005] [Medline: 30032968]
9. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: A state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
10. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: A systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
11. Wald JS. Variations in patient portal adoption in four primary care practices. In: Proceedings of the American Medical Informatics Association (AMIA) Annual Symposium. 2010 Nov 17 Presented at: American Medical Informatics Association (AMIA) Annual Symposium; November 13-17, 2010; Washington, DC p. 837-841 URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041333/pdf/amia-2010_sympproc_0837.pdf
12. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: An interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
13. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
14. Lluch M. Healthcare professionals' organisational barriers to health information technologies: A literature review. Int J Med Inform 2011 Dec;80(12):849-862. [doi: 10.1016/j.ijmedinf.2011.09.005] [Medline: 22000677]
15. Abbott PA, Foster J, de Fatima Marin H, Dykes PC. Complexity and the science of implementation in health IT: Knowledge gaps and future visions. Int J Med Inform 2014 Jul;83(7):e12-e22. [doi: 10.1016/j.ijmedinf.2013.10.009] [Medline: 24444700]
16. Kooij L, Groen WG, van Harten WH. Barriers and facilitators affecting patient portal implementation from an organizational perspective: Qualitative study. J Med Internet Res 2018 May 11;20(5):e183 [FREE Full text] [doi: 10.2196/jmir.8989] [Medline: 29752253]
17. Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: A fit framework and a case study. BMC Med Inform Decis Mak 2006 Jan 09;6:3 [FREE Full text] [doi: 10.1186/1472-6947-6-3] [Medline: 16401336]

18. Tsiknakis M, Kouroubali A. Organizational factors affecting successful adoption of innovative eHealth services: A case study employing the FITT framework. Int J Med Inform 2009 Jan;78(1):39-52. [doi: 10.1016/j.ijmedinf.2008.07.001] [Medline: 18723389]
19. Prgomet M, Georgiou A, Callen J, Westbrook J. Fit between individuals, tasks, technology, and environment (FITTE) framework: A proposed extension of FITT to evaluate and optimise health information technology use. Stud Health Technol Inform 2019 Aug 21;264:744-748. [doi: 10.3233/SHTI190322] [Medline: 31438023]
20. Schnall R, Smith AB, Sikka M, Gordon P, Camhi E, Kanter T, et al. Employing the FITT framework to explore HIV case managers' perceptions of two electronic clinical data (ECD) summary systems. Int J Med Inform 2012 Oct;81(10):e56-e62 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.07.002] [Medline: 22841702]
21. Cho H, Flynn G, Saylor M, Gradilla M, Schnall R. Use of the FITT framework to understand patients' experiences using a real-time medication monitoring pill bottle linked to a mobile-based HIV self-management app: A qualitative study. Int J Med Inform 2019 Nov;131:103949 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.08.009] [Medline: 31561192]
22. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: Systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
23. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
24. Koivunen M, Hätönen H, Välimäki M. Barriers and facilitators influencing the implementation of an interactive internet-portal application for patient education in psychiatric hospitals. Patient Educ Couns 2008 Mar;70(3):412-419. [doi: 10.1016/j.pec.2007.11.002] [Medline: 18079085]
25. Zulman DM, Jenchura EC, Cohen DM, Lewis ET, Houston TK, Asch SM. How can eHealth technology address challenges related to multimorbidity? Perspectives from patients with multiple chronic conditions. J Gen Intern Med 2015 Aug;30(8):1063-1070 [FREE Full text] [doi: 10.1007/s11606-015-3222-9] [Medline: 25691239]
26. Murray E, Burns J, May C, Finch T, O'Donnell C, Wallace P, et al. Why is it difficult to implement e-health initiatives? A qualitative study. Implement Sci 2011 Jan 19;6:6 [FREE Full text] [doi: 10.1186/1748-5908-6-6] [Medline: 21244714]
27. Information to Support Well-Being and Service Renewal: eHealth and eSocial Strategy 2020. Helsinki, Finland: Ministry of Social Affairs and Health; 2015. URL: http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/74459/URN_ISBN_978-952-00-3575-4.pdf [accessed 2018-06-21] [WebCite Cache ID 70L10R44oISBN9789520035747]
28. This is Health Village. Health Village. 2019. URL: https://www.terveyskyla.fi/en/about-health-village/this-is-health-village [accessed 2020-07-06]
29. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
30. Ingebrigtsen T, Georgiou A, Clay-Williams R, Magrabi F, Hordern A, Prgomet M, et al. The impact of clinical leadership on health information technology adoption: Systematic review. Int J Med Inform 2014 Jun;83(6):393-405. [doi: 10.1016/j.ijmedinf.2014.02.005] [Medline: 24656180]
31. Atkinson NL, Massett HA, Mylks C, McCormack LA, Kish-Doto J, Hesse BW, et al. Assessing the impact of user-centered research on a clinical trial eHealth tool via counterbalanced research design. J Am Med Inform Assoc 2011;18(1):24-31 [FREE Full text] [doi: 10.1136/jamia.2010.006122] [Medline: 21169619]
32. Kujala S. User involvement: A review of the benefits and challenges. Behav Inf Technol 2003 Jan;22(1):1-16. [doi: 10.1080/01449290301782]
33. Marien S, Legrand D, Ramdoyal R, Nsenga J, Ospina G, Ramon V, et al. A user-centered design and usability testing of a web-based medication reconciliation application integrated in an eHealth network. Int J Med Inform 2019 Jun;126:138-146. [doi: 10.1016/j.ijmedinf.2019.03.013] [Medline: 31029255]
34. Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. Med J Aust 2004 Mar 15;180(S6):S57-S60. [Medline: 15012583]
35. Chaudoir SR, Fisher JD, Simoni JM. Understanding HIV disclosure: A review and application of the Disclosure Processes Model. Soc Sci Med 2011 May;72(10):1618-1629 [FREE Full text] [doi: 10.1016/j.socscimed.2011.03.028] [Medline: 21514708]
36. Price M, Lau F. The clinical adoption meta-model: A temporal meta-model describing the clinical adoption of health information systems. BMC Med Inform Decis Mak 2014 May 29;14:43 [FREE Full text] [doi: 10.1186/1472-6947-14-43] [Medline: 24884588]
37. Ludwick DA, Doucette J. Adopting electronic medical records in primary care: Lessons learned from health information systems implementation experience in seven countries. Int J Med Inform 2009 Jan;78(1):22-31. [doi: 10.1016/j.ijmedinf.2008.06.005] [Medline: 18644745]
38. Kruse CS, Kothman K, Anerobi K, Abanaka L. Adoption factors of the electronic health record: A systematic review. JMIR Med Inform 2016 Jun 01;4(2):e19 [FREE Full text] [doi: 10.2196/medinform.5525] [Medline: 27251559]

39. Guest G, Bunce A, Johnson L. How many interviews are enough? Field methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]
40. Kirkendall ES, Goldenhar LM, Simon JL, Wheeler DS, Spooner SA. Transitioning from a computerized provider order entry and paper documentation system to an electronic health record: Expectations and experiences of hospital staff. Int J Med Inform 2013 Nov;82(11):1037-1045. [doi: 10.1016/j.ijmedinf.2013.08.005] [Medline: 24041453]
41. Ward MM, Vartak S, Schwichtenberg T, Wakefield DS. Nurses' perceptions of how clinical information system implementation affects workflow and patient care. Comput Inform Nurs 2011 Sep;29(9):502-511. [doi: 10.1097/NCN.0b013e31822b8798] [Medline: 21825972]
42. Lapointe L, Rivard S. Getting physicians to accept new information technology: Insights from case studies. CMAJ 2006 May 23;174(11):1573-1578 [FREE Full text] [doi: 10.1503/cmaj.050281] [Medline: 16717265]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
FITT: Fit between Individuals, Task, and Technology
HUS: HUS Helsinki University Hospital
IT: information technology

Edited by G Eysenbach; submitted 05.01.20; peer-reviewed by M Twomey, R Zhang, A George; comments to author 10.03.20; revised version received 24.04.20; accepted 03.06.20; published 12.08.20.

Please cite as:
Kujala S, Ammenwerth E, Kolanen H, Ervast M
Applying and Extending the FITT Framework to Identify the Challenges and Opportunities of Successful eHealth Services for Patient Self-Management: Qualitative Interview Study
J Med Internet Res 2020;22(8):e17696
URL: https://www.jmir.org/2020/8/e17696
doi:10.2196/17696
PMID:32784175

Review

Evidence on Virtual Reality–Based Therapies for Psychiatric Disorders: Meta-Review of Meta-Analyses

Laura Dellazizzo[1,2], MSc; Stéphane Potvin[1,2], PhD; Mimosa Luigi[1,2], BSc; Alexandre Dumais[1,2,3], MD, PhD, FRCPC

[1]Research Center of the Institut Universitaire en Santé Mentale de Montréal, Montreal, QC, Canada

[2]University of Montreal, Montreal, QC, Canada

[3]Institut national de psychiatrie légale Philippe-Pinel, Montreal, QC, Canada

Corresponding Author:
Alexandre Dumais, MD, PhD, FRCPC
Research Center of the Institut Universitaire en Santé Mentale de Montréal
7331 Hochelaga
Montreal, QC, H1N 3V2
Canada
Phone: 1 514 251 4000 ext 3925
Email: alexandre.dumais@umontreal.ca

Abstract

Background: Among all diseases globally, mental illnesses are one of the major causes of burden. As many people are resistant to conventional evidence-based treatments, there is an unmet need for the implementation of novel mental health treatments. Efforts to increase the effectiveness and benefits of evidence-based psychotherapy in psychiatry have led to the emergence of virtual reality (VR)–based interventions. These interventions have shown a wide range of advantages over conventional psychotherapies. Currently, VR-based interventions have been developed mainly for anxiety-related disorders; however, they are also used for developmental disorders, severe mental disorders, and neurocognitive disorders.

Objective: This meta-review aims to summarize the current state of evidence on the efficacy of VR-based interventions for various psychiatric disorders by evaluating the quality of evidence provided by meta-analytical studies.

Methods: A systematic search was performed using the following electronic databases: PubMed, PsycINFO, Web of Science, and Google Scholar (any time until February 2020). Meta-analyses were included as long as they quantitatively examined the efficacy of VR-based interventions for symptoms of a psychiatric disorder. To avoid overlap among meta-analyses, for each subanalysis included within this meta-review, only one analysis provided from one meta-analysis was selected based on the best quality of evidence.

Results: The search retrieved 11 eligible meta-analyses. The quality of evidence varied from very low to moderate quality. Several reasons account for the lower quality evidence, such as a limited number of randomized controlled trials, lack of follow-up analysis or control group, and the presence of heterogeneity and publication bias. Nonetheless, evidence has shown that VR-based interventions for anxiety-related disorders display overall medium-to-large effects when compared with inactive controls but no significant difference when compared with standard evidence-based approaches. Preliminary data have highlighted that such effects appear to be sustained in time, and subjects may fare better than active controls. Neurocognitive disorders also appear to improve with VR-based approaches, with small effects being found for various clinical outcomes (eg, cognition, emotion). Finally, there are insufficient data to classify VR-based interventions as an evidence-based practice for social skills training in neurodevelopmental disorders and compliance among patients with schizophrenia.

Conclusions: VR provides unlimited opportunities by tailoring approaches to specific complex problems and individualizing the intervention. However, VR-based interventions have not shown superiority compared with usual evidence-based treatments. Future VR-based interventions should focus on developing innovative approaches for complex and treatment-resistant symptoms that are difficult to address with traditional treatments. Future research should also aim to gain a better understanding of the potential factors that may mediate VR outcomes to improve treatment.

(J Med Internet Res 2020;22(8):e20889) doi:10.2196/20889

KEYWORDS

systematic review; virtual reality; therapy; mental disorders; meta-analysis

Introduction

Mental illnesses are one of the predominant causes of burden among all diseases globally [1]. It has been estimated that over 15% of adults in the United States have lived with a psychiatric disorder in the past year, including mental, behavioral, or emotional disorders [2]. Anxiety disorders and depression generally display the highest prevalence rates [2-5]. Furthermore, approximately 4.5% of adults have reported being affected by a severe mental disorder resulting in functional impairment that affects or limits major life activities [2]. In addition, 1 in 6 youths in the United States aged between 6 and 17 years will experience a mental disorder every year [6]. With half of the mental health conditions manifesting by the age of 14 years and three-quarter by mid-20s, youth remains to be an important period for the emergence of a mental disorder [7]. Given the elevated prevalence of mental health problems, psychiatric disorders represent a substantial socioeconomic burden for patients, caregivers, health care providers, and the overall society, with associated costs including informal care, productivity loss, and premature death [1,8,9]. The global direct and indirect economic costs of mental disorders have been estimated at approximately US $2.5 trillion [10]. For those seeking treatment, conventional mental health care for most psychiatric disorders typically includes pharmacological and psychological options. However, there is an enduring discussion as to whether an individual option or a combination of options should be used to treat psychiatric symptoms. Huhn et al [11] conducted a systematic overview of the efficacy of pharmacotherapies and psychotherapies for major psychiatric disorders and concluded that there remains room for amelioration. Among meta-analyses specifically comparing pharmacotherapy with psychotherapy head-to-head, there was a trend in favor of psychotherapy for relapse prevention in depression and bulimia and pharmacological interventions for schizophrenia and dysthymia. Although pharmacological treatments have received more attention, they may likely be less acceptable to patients, and a proportion of individuals will experience adverse effects or will not respond adequately to this approach [12-14]. Evidence-based psychosocial interventions (eg, psychoeducation, interpersonal psychotherapy, cognitive behavioral therapy), offered as the sole or adjunctive treatment, have shown promising results and allow patients to learn skills to overcome and better cope with their symptoms while also preventing relapse [15,16]. Nevertheless, the effect sizes of psychotherapies for mental disorders are moderate at best with dropout rates as high as 30%, and treatment gains not always being maintained for a long term [12,17]. Thus, with underscored inadequacies of conventional treatment, there remains an unmet need for the implementation of novel treatments. Efforts to increase the effectiveness, acceptance, and access to evidence-based psychotherapies have led to the emergence of technology-assisted psychological interventions. A prime example is the virtual reality (VR)–based approach that may enhance conventional face-to-face approaches. Generally, VR techniques are based on similar principles as those used in traditional cognitive behavioral approaches; however, they also increase the possibility of transferring the learning achieved during VR sessions to patients' everyday lives. These interventions enable the manipulation of the virtual environment and can be used to recreate environmental triggers that elicit distress in patients with mental health problems, thereby allowing them to learn to better manage their difficulties in real time [18,19]. Although VR approaches display additional treatment costs and may lead to cybersickness in some patients [18,20], the literature has nonetheless shown the wide range of advantages of its use, that is, reduced ecological impact, personalized treatment, high level of control over exposure parameters, and better acceptability of and adherence to treatment [18,19,21,22].

VR-based treatments have been developed for many psychopathologies, particularly for anxiety-related disorders, and for developmental disorders, severe mental disorders, and neurocognitive disorders [23,24]. As the field is relatively new, many of these studies have been impacted by methodological issues (ie, small sample size, limited number of randomized trials with strong methodologies including blinding and allocation concealment). Nonetheless, several meta-analyses have been conducted to summarize the evidence of these VR interventions. Statistical meta-analyses are frequently used by clinicians as a resource to determine the best evidence-based treatment options for their patients [25]. Considering the increasing number of meta-analyses on the efficacy of VR-based interventions in psychiatric disorders, we conducted a meta-review to summarize the magnitude of the effects of VR for the treatment of various mental disorders and to evaluate the quality of evidence provided by the meta-analyses. This is to help create recommendations for the use of VR-based approaches for clinicians and policy makers and to guide future research on novel VR interventions.

Methods

Search Strategy

A search was independently conducted by 2 graduate students (LD and ML) on PubMed, PsycINFO, Web of Science, and Google Scholar electronic databases, from each database's inception to February 2020. Search terms were chosen to be inclusive of VR (eg, "virtual," "virtual reality," "VR"), mental disorders (eg, "mental illness," "anxiety," "post-traumatic stress disorder," "autism," "attention deficit hyperactivity disorder," "neurodevelopmental disorder," "severe mental disorder," "depression," "schizophrenia," "dementia," "substance use disorder"), and interventions (eg, "intervention," "therapy"). The search syntax was tailored for each database. See Multimedia Appendix 1 for the specific search strategy adapted for each database. Only meta-analytical study designs were selected. No setting, date, or geographical restrictions were applied. Searches were limited to English or French language sources. The authors of the articles to which we had restricted access were contacted.

Study Eligibility

Meta-analyses were included as long as they quantitatively examined the efficacy of VR-based interventions for the symptoms of psychiatric disorders. To maximize the number of meta-analyses, we did not restrict the search to any specific psychiatric population or any age group. It is noteworthy that

a problem with meta-analyses is that they may overlap when many have been conducted on a particular disorder and a particular type of subanalysis for the disorder (ie, pre-post efficacy, comparison with inactive or active control, long-term effects). To avoid this issue, for each subanalysis included within this meta-review, only one analysis provided from one meta-analysis was selected based on the best quality of evidence. The inclusion of the meta-analyses was generally based on (1) the year of publication, (2) the number of included studies, and (3) the quality of the included studies (ie, randomized controlled trials [RCTs]). To ensure consensus, discussions on the inclusion of meta-analyses were held with a senior researcher (SP). As a meta-analysis only requires a minimum of 2 studies [26], we chose to include meta-analyses that analyzed at least 2 studies per symptom. However, it should be noted that increasing the number of included studies tends to enhance the generalizability of results [27]. Studies were excluded if they (1) combined several treatment modalities (eg, other computerized approaches such as internet-based therapies) and did not have an effect size for VR specifically or (2) combined disorders together (eg, overall anxiety disorders).

Data Extraction

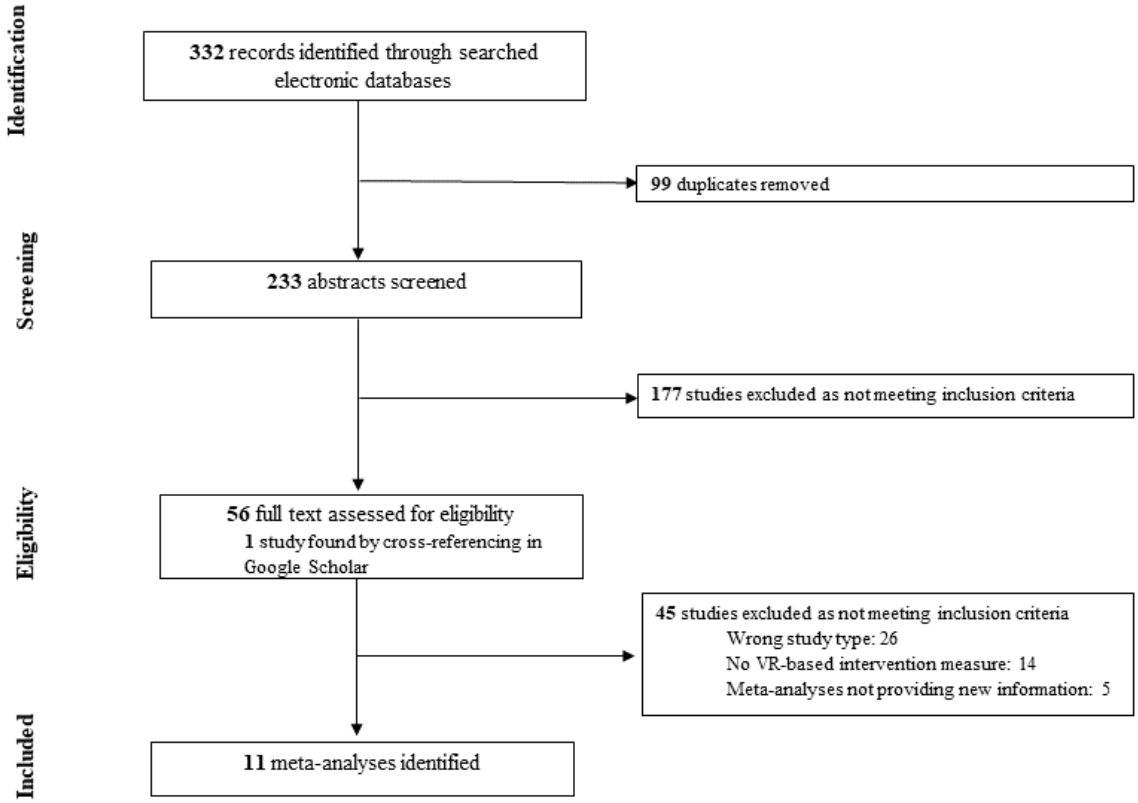
Data were extracted using a standardized form by LD and ML. Key information related to the sample, effect sizes (ie, Cohen d, Hedges g, standardized mean difference), outcome measured, control group, timeline (ie, posttreatment, follow-up), confounding factors (ie, moderator analyses), heterogeneity (ie, Q statistics, I^2 index), and publication bias (ie, funnel plot examination, Egger test) were recorded. Refer to Multimedia Appendix 2 [28-38] for an overview of the extracted data. The effect sizes were categorized as small (0.2), medium (0.5), and large (>0.8) effects [39]. Data were independently extracted by LD and ML, and all queries were resolved in discussions with SP. Furthermore, LD and SP independently undertook quality assessments for the effect sizes reported in the meta-analyses using a set of criteria based on the Grading of Recommendation, Assessment, Development, and Evaluation checklist [40-43]. Higher scores were assigned to analyses that suggested more precision (ie, a smaller range of 95% CIs around the effect size [under 0.5 absolute effect size]), analyzed follow-ups, included only controlled trials, conducted moderator analyses, reported no heterogeneity and publication bias, and included an outcome principally targeted by the intervention. Studies were assigned to be of high, moderate-to-high, moderate, moderate-to-low, low, and very low quality. To achieve a high standard of reporting data, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed (Multimedia Appendix 3) [44].

Results

Description of Studies

The literature search identified 233 potential articles that were screened for eligibility after removing duplicates. One additional meta-analysis was identified by cross-referencing on Google Scholar. Among these articles, 11 meta-analyses were selected that provided 41 effect sizes. The PRISMA flowchart for the inclusion of studies in the meta-review is shown in Figure 1. The psychiatric disorders were categorized based on the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition [45], as (1) anxiety disorders (acrophobia, arachnophobia, aviophobia, panic disorder, and social anxiety), (2) trauma- and stress-related disorders (posttraumatic stress disorder), (3) severe mental disorders (depressive disorder and schizophrenia spectrum), (4) neurodevelopmental disorders (autism), and (5) neurocognitive disorders (mild cognitive impairment and dementia). When several analyses were conducted within the meta-analyses, we retrieved one specific effect size estimate for (1) the pre-post efficacy of VR-based interventions, (2) the comparison of VR-based interventions with inactive control, (3) the comparison of VR-based interventions with active controls, and (4) the long-term effects of VR-based interventions after the follow-up. Refer to Multimedia Appendix 2 [28-38] for a summary of the quality of evidence provided by the included meta-analyses. Each meta-analysis included 2-16 studies, with samples ranging between 30 and 454 individuals.

Figure 1. Flow-chart depicting the search strategy employed to find the meta-analyses to include in this review.



Anxiety Disorders

Anxiety disorders, particularly specific phobias (eg, fear of flying, fear of heights), have become typical in VR implementation as exposure is undeniably a key element that must be addressed in these disorders [28,46-48]. There are 2 predominant theoretical models related to learning, which may explain how exposure therapy reduces anxiety [49]: emotional processing theory [50,51] and inhibitory learning model [52]. Both theories claim that exposure allows patients to learn corrective information about a stimulus that is feared. In VR-based interventions, the sense of presence or the feeling of *being there* has been considered as the principle mechanism that leads to the experience of anxiety [53,54]. In this sense, the feeling of presence experienced in VR offers the opportunity to immerse patients to their feared stimuli in the VR environment, which is customized to match specific aspects of their fear [55,56].

Specific Phobias

Fear of Heights (Acrophobia)

A meta-analysis by Parsons et al [29] found an average random effect size of 0.93 (95% CI 0.44 to 1.43) for acrophobia. Graded as low-quality evidence from 4 studies with different control groups (2 were compared with a waitlist, 1 had no control group, and 1 was compared with in vivo interventions), the results of this meta-analysis suggest a statistically large overall effect for this specific phobia. However, heterogeneity and publication bias were not evaluated, and confounding factors were not considered.

Fear of Spiders (Arachnophobia)

A meta-analysis by Parsons et al [29], comprising 4 studies with mixed designs, found an overall large effect of 0.92 (95% CI 0.25 to 1.59) for VR interventions. The evidence was graded as low quality, notably owing to the inclusion of studies with mixed designs and the lack of consideration of heterogeneity, publication bias, and confounding factors. Better quality evidence was provided by Opris et al [28], including RCTs that compared VR-based interventions with active controls specifically. The authors retrieved 2 studies that showed no significant posttreatment (d=−0.12; 95% CI −0.31 to 0.06) and no long-term (d=−0.20; 95% CI −0.49 to 0.08) differences in primary arachnophobia outcomes.

Fear of Flights (Aviophobia)

Evidence based on the meta-analysis by Cardoş et al [30] including RCTs for aviophobia was evaluated to be of low-to-moderate to moderate quality. First, regarding the efficacy of VR-based interventions based on a large sample size of 454 participants, 16 study arms were included at posttreatment and 15 at follow-up. Statistically significant medium effect sizes were observed (g=0.592; 95% CI 0.327 to 0.858; g=0.588; 95% CI 0.216 to 0.960, respectively). As both analyses presented statistically significant heterogeneity, moderator analyses were conducted to explain the divergences among studies. The quality of randomized trials and the mean age of patients were significant moderators at posttreatment, whereas the number of patients and follow-up intervals were significant moderators at follow-up. Moreover, examination of the funnel plot showed asymmetry, suggesting publication bias. Second, when compared with the inactive control groups, the

results showed large statistically significant effects at posttreatment (g=1.350; 95% CI 0.664 to 2.037) and medium statistically significant effects at follow-up (g=0.583; 95% CI 0.108 to 1.058). Heterogeneity was observed at posttreatment, and there was presence of funnel plot asymmetry. Third, when comparing VR with classical evidence-based interventions, the results showed a small significant effect for VR-based interventions (g=0.353; 95% CI 0.152 to 0.555), whereas follow-up studies indicated a moderate significant effect (g=0.615; 95% CI 0.179 to 1.052). Heterogeneity was evident at follow-up, with the number of participants and follow-up period as significant moderators. Furthermore, follow-up studies pointed toward publication bias. Fourth, a lack of difference between VR-based interventions and other exposure-based interventions was observed at posttreatment (g=0.122; 95% CI −0.225 to 0.469). A moderate-to-large significant effect was still found at follow-up, in favor of VR-based interventions (g=0.697; 95% CI 0.101 to 1.292). Significant heterogeneity was observed, revealing 3 moderators (number of exposure sessions, outcome type, and follow-up intervals), in addition to the presence of publication bias.

In summary, meta-analytical evidence shows that VR-based interventions may be effective for specific phobias. The quality of evidence ranged from low to moderate quality, with better quality evidence provided for aviophobia, which comprised a larger number of RCTs with a larger sample size. The presence of heterogeneity and publication bias was evaluated for aviophobia. When compared with active controls, the results suggested better aviophobia outcomes for VR-based therapies than classical evidence-based interventions, with no significant superiority over other exposure-based therapies. However, for arachnophobia, significant superiority was found for VR when compared with active controls. Finally, the effects of VR for aviophobia remained stable in time, indicating that VR might fare better than active controls in the long term.

Panic Disorder With or Without Agoraphobia

First, the meta-analysis by Parsons et al [29] included 3 studies and observed very large significant overall effects (d=1.79) for VR-based interventions at posttreatment. The evidence was graded to be of low quality, notably owing to the inclusion of studies with mixed designs and the lack of consideration for heterogeneity, publication bias, and confounding factors. Second, Fodor et al [31] observed large effects at posttreatment for VR-based interventions in comparison with inactive controls (g=1.80; 95% CI 1.01 to 2.60). The evidence provided by 2 RCTs was graded as low-to-moderate quality; the authors did not observe any heterogeneity, although there was a publication bias for their entire study sample. Third, Fodor et al [31] found no significant difference at posttreatment between VR and other psychological therapies (g=−0.05; 95% CI −0.32 to 0.21). This evidence graded as moderate quality was provided from 6 RCTs, and the data displayed no heterogeneity. Finally, Opris et al [28] analyzed the long-term effect of VR-based therapies, specifically in comparison with active controls. Their analysis found a small significant effect favoring VR (d=0.18; 95% CI 0.10 to 0.26). Evidence was graded as low-to-moderate quality based on 2 RCTs. However, heterogeneity and publication bias were not reported.

In summary, these meta-analyses on panic disorder with or without agoraphobia showed that VR-based interventions are efficient. Evidence has been evaluated to be of low to moderate quality, with the quality of evidence being lower owing to a lack of consideration of heterogeneity, publication bias, and moderating factors. Better quality evidence was provided for comparison with classical evidence-based interventions, which showed that VR was no better than these interventions at posttreatment. At follow-up, there was a small superiority observed favoring VR over standard interventions.

Social Anxiety

First, the meta-analysis by Kampmann et al [32] observed large overall effects for VR-based interventions at posttreatment (g=1.09; 95% CI 0.80 to 1.39) of social anxiety symptoms. Evidence provided by 3 RCTs was graded as low-to-moderate quality, with heterogeneity, publication bias, and moderators not being examined. Second, in comparison with inactive controls, the meta-analysis by Carl et al [33] found a large posttreatment effect (g=0.97; 95% CI 0.62 to 1.31). Evidence from 7 studies included randomized controls, with a total sample of 236 individuals, and was evaluated to be of low-to-moderate quality. Although not specifically for this subanalysis, the authors did observe moderate heterogeneity and possible presence of publication bias in their overall study. Third, as for the comparison with active controls, Chesham et al [34] found no significant difference between VR-based interventions and standard treatments using in vivo or imaginal approaches (g=−0.01; 95% CI −0.30 to 0.28). Evidence from 7 well-controlled trials (n=340) with moderate heterogeneity and no presence of publication bias was graded as low-to-moderate quality. Fourth, in terms of follow-up assessments, Kampmann et al [32] observed that the large overall effect for VR was maintained in time (g=0.93 for less than 5 months and g=1.20 for over 5 months). However, the effect was not different from the effect of active controls. Evidence was evaluated as low-to-moderate quality as the authors did not examine heterogeneity or publication bias owing to the limited number of trials included in their analyses.

In summary, overall evidence was evaluated as low-to-moderate quality: most meta-analyses included a limited number of trials and moderator analyses, and did not report heterogeneity or publication bias. Medium to large effects were observed for VR-based interventions for social phobia. Nevertheless, no significant difference existed between the VR-based interventions and standard treatment. The overall beneficial effect of VR interventions was maintained in the long term, although no significant difference was observed with active controls.

Trauma- and Stressor-Related Disorders

Trauma- and stress-related disorders (such as posttraumatic stress disorder) may develop by directly experiencing, witnessing, or repeating exposure to aversive elements of a traumatic event (eg, combat, sexual assault). Although many show resilience following exposure, up to one-third of those confronted with a traumatic event will subsequently develop clinically relevant posttraumatic symptoms (eg, reexperiencing, avoidance) [57]. It is worth noting that VR exposure therapy

has potential efficacy in the treatment of posttraumatic stress disorder for different types of trauma and that this technology can compensate for the shortcomings of traditional therapy (ie, inherent avoidance of traumatic memory) [58,59]. VR may ease the emotional engagement of patients during exposure to traumatic stimuli by eschewing avoidance symptoms and facilitating therapeutic control [60].

First, a meta-analysis by Deng et al [35] found a small superiority of VR interventions for posttraumatic stress disorder symptoms in comparison with inactive and active controls combined. This effect was significant (g=0.327; 95% CI 0.105 to 0.550). A similar significant effect of VR-based interventions was observed when considering only studies that used intention-to-treat analyses or reported complete outcome data (g=0.584; 95% CI 0.318 to 0.850). Evidence was evaluated to be of moderate quality, provided from 10 RCTs (n=309) showing moderate heterogeneity and no publication bias. Second, the same authors observed moderate effects for the superiority of VR-based interventions relative to inactive controls alone (g=0.567; 95% CI 0.270 to 0.863) [35]. Evidence was evaluated to be of low-to-moderate quality based on 5 RCTs (n=175) with no publication bias; heterogeneity for this specific subanalysis was not provided. Third, no significant difference was found between VR and active controls [35]. Evidence from 6 RCTs (n=239) was also evaluated to be of low-to-moderate quality with no presence of publication bias; heterogeneity for this specific subanalysis was similarly not provided. Finally, as for follow-up effects, Deng et al [35] found moderate-to-large improvements for VR-based interventions in comparison with the combination of inactive and active controls (g=0.697 and g=0.848 for short- and long-term effects, respectively). Evidence provided by 9 and 11 RCTs was evaluated to be of low-to-moderate quality. Moderator analyses, heterogeneity, and publication bias were not reported.

In summary, the meta-analyses on posttraumatic stress disorder showed small-to-moderate effects for VR. Evidence was generally graded as low-to-moderate to moderate quality. No significant difference was found with standard evidence-based interventions. Moreover, improvements in VR were maintained in time.

Severe Mental Disorders

VR-based treatments for the symptoms of individuals with severe mental disorders have multiplied in recent years. Although there are very limited studies on the effects of VR for those with mood disorders [61,62], this innovative tool may nonetheless be used to deliver psychoeducation and to induce relaxation and enhance positive emotions [63]. Moreover, VR scenarios have been used to treat symptoms of other severe mental disorders such as schizophrenia by enabling patients to practice social skills (eg, vocational skill training) and learn to cope with distress associated with psychotic symptoms [64,65]. As those with severe mental disorders also experience difficulties with activities in everyday life, VR may also be used to test and support their performance using an environment that simulates real-life activities and increase compliance with treatment [36].

Depressive Disorder

It is worth noting that VR interventions included in the meta-analyses did not target depression as a diagnosis per se and did not generally aim at the reduction of depressive symptoms as a main outcome. First, an overall small effect was observed for VR-based interventions at posttreatment in the analysis by Kampmann et al [32] (g=0.44; 95% CI 0.02 to 0.87). Evidence evaluated to be of low-to-moderate quality was based on 2 randomized trials (n=119); moderator analyses, heterogeneity, and publication bias were not reported. Second, when compared with inactive controls, Fodor et al [31] found a significant moderate effect for VR interventions (g=0.73; 95% CI 0.25 to 1.21). Evidence graded as low-to-moderate quality was based on 10 RCTs showing high heterogeneity; there was also the presence of publication bias on their overall analyses. Third, the same authors observed no significant difference between the VR and active controls at posttreatment. Evidence was evaluated as moderate quality based on 13 RCTs showing low heterogeneity. Fourth, in the follow-up assessment by Fodor et al [31], which retrieved 5 RCTs, there was no significant difference in comparison with active controls. Moderate heterogeneity was observed, and evidence was similarly graded as moderate quality.

In summary, evidence from these meta-analyses on depressive symptoms highlighted that overall VR-based interventions may reduce comorbid depressive symptoms. Evidence was graded as low-to-moderate quality. However, the effect did not seem to be different from standard evidence-based interventions. No significant long-term differences were found at follow-up compared with active controls.

Schizophrenia Spectrum Disorders

A meta-analysis by Valimaki et al [36] investigated RCTs on the effects of VR to support treatment compliance among patients with schizophrenia spectrum disorders. Treatment compliance was defined as loss to follow-up and withdrawal by the trialist. Overall, 3 short-term trials (n=156) with a duration of 5 to 12 weeks were retrieved, which were aimed at delivering skill training (ie, social skills and vocational skills). The authors assessed the quality of the included trials as low quality. Findings showed that there was a nonsignificant effect of VR on compliance (risk difference=0.02; 95% CI −0.08 to 0.12). Evidence provided by this meta-analysis was evaluated to be of moderate quality, showing no heterogeneity yet a moderate risk of bias. Comparison with active controls has not been reported.

In summary, at present, there are insufficient quality data to classify VR as an evidence-based practice for treatment compliance in patients with schizophrenia.

Neurodevelopmental Disorders

Autism spectrum disorder has received interest in the field of VR. VR technologies have been promising by supporting learning for children and adults with autism, who may find social interactions difficult. Several VR environments have been developed, such as virtual cafes, schools, or job interviews [66]. VR allows role play and practice skills without the threat of real-world consequences [67].

One meta-analysis by Barton et al [37] evaluated the effects of technology-aided support in comparison with a control condition on improving a mix of primary skills (ie, communication, academic, engagement or task completion, social, emotion recognition, and adaptive). For VR-based interventions specifically, 2 studies using group designs amounting to a small sample size of 30 individuals were included. One study comprised children with high-functioning autism for social interaction training and the other study comprised adults with autism spectrum for job interview training. Evidence was thus evaluated to be of very low quality, with heterogeneity and publication bias not being reported for the subanalysis. The meta-analysis yielded a nonsignificant estimated effect size of 0.37 (95% CI −1.71 to 2.46). Moreover, follow-up results were not evaluated.

In summary, there are insufficient quality data to classify VR as an evidence-based practice among individuals with autism.

Neurocognitive Disorders

Individuals with neurocognitive disorders (ie, mild cognitive impairment or dementia) may benefit from VR-based interventions that promote simulations of functional learning, the transfer of learned functions to daily life, and relaxation [38].

A meta-analysis by Kim et al [38] analyzed the effects of different VR-based intervention platforms for individuals with mild cognitive impairment and dementia. The authors found an overall small effect size for VR, including executive, emotion, fitness, and cognitive outcomes (d=0.29; 95% CI 0.16 to 0.42). Larger improvements were found for patients with mild cognitive impairment compared with patients with dementia or mixed samples. With regard to their subanalysis for experimental and control group allocation, random allocation (d=0.36) and no randomization (d=0.4) showed small-to-moderate effects, which were larger than those with a one-group design (d=0.15). When subdividing the different outcomes, the effect sizes for cognitive functions (d=0.42) were higher and significant in comparison with emotion (d=0.14) and executive functions (d=0.07). Overall, evidence graded as low to low-to-moderate quality was based on a mix of impairments provided from 11 studies with mixed designs with significant heterogeneity; the authors stated that publication bias was not a concern. There was also a lack of follow-up assessments and no comparison with active controls.

In summary, low to low-to-moderate quality evidence indicated that VR interventions may positively affect various clinical outcomes among patients with cognitive impairment and thus improve cognitive and routine functions. However, these VR-based interventions were not compared with active controls.

Discussion

This meta-review aimed to summarize the current state of evidence on the efficacy of VR-based interventions for psychiatric disorders by evaluating the data provided by meta-analytical studies. Cumulating evidence on various anxiety disorders and posttraumatic stress disorder showed that VR-based interventions displayed overall medium-to-large effects in comparison with inactive controls. However, there was globally no significant difference in comparison with standard evidence-based approaches at posttreatment, apart from significant differences with classical evidence-based interventions (g=0.353 in favor of VR) for aviophobia. With limited evidence on the superiority of one over the other, these findings suggest that both VR-based and standard evidence-based therapies are as effective for anxiety-related disorders at posttreatment. Furthermore, although results may not be dissimilar among interventions in the short term, preliminary data on aviophobia and panic disorder have highlighted that the effects of VR appear to be sustained in time, and subjects may fare better in the long term than with active controls. This suggests that although the effects of conventional treatments diminish in time, the effects of VR appear to be maintained, leading to longer-lasting positive outcomes. Such differential outcomes may be explained by the advantages of VR over classical and in vivo exposure-based interventions, comprising a more flexible and personalized approach where the therapist can better control the content of exposure (eg, including turbulence in the exposure of flight phobia), exposure rhythm, and repetition of scenarios [68-70].

VR interventions have also shown promise in the treatment of other disorders included in this meta-review. First, although there are only a few VR interventions that have been developed specifically for individuals with mood disorders [61,62], VR-based therapies have been reported to be effective in the short term to reduce depressive symptoms comorbid with anxiety-related disorders. Second, neurocognitive disorders seem to benefit from VR-based interventions with overall small effects on clinical outcomes such as cognition and emotion. The interaction provided in VR environments may therefore improve well-being, routine functions, and cognition among patients with cognitive impairments by stimulating them. However, it is unknown if VR fares better than conventional treatments for this population. As for autism, given the core impairments in social communication and interaction, it is contended that VR may have a potential for training in highly controlled social scenarios, allowing patients to rehearse interactions or responses [66]. However, no support was observed for the efficacy of VR for neurodevelopmental disorders on social skills training, but this was based on limited low-quality studies on autism [66]. Similarly, VR environments have been created to enable skills training of everyday tasks among patients with schizophrenia, such as to help support treatment follow-up and medication taking. However, the retrieved meta-analysis for compliance among patients with schizophrenia yielded no significant effects based on the low-quality studies included [36].

As observed, meta-analyses serve as a useful tool to provide a global overview of the benefits of VR for patients affected by psychiatric disorders. The quality of evidence was evaluated as being quite variable, ranging from very low to moderate quality. Several reasons account for the lower quality of evidence. First, many meta-analyses included a limited number of RCTs within their analyses, thereby also lacking large sample sizes. Meta-analyses with a larger number of randomized trials were provided for aviophobia [30], posttraumatic stress disorder [35], and depressive symptoms [31]. Among these meta-analyses,

Cardoş et al [30] conducted moderator analyses and observed that the quality of RCTs was a significant moderator, with lower quality trials yielding larger effect sizes. Although few highlighted this as a concern, it remains of importance because an RCT with methodological issues, such as lack of blinding, is insufficient to create evidence-based practice. Thus, the quality of the studies should be taken into account to understand the efficacy of interventions. Second, for certain psychiatric disorders such as autism and mild cognitive impairment and dementia, no analysis was conducted evaluating long-term effects and comparing VR-based interventions with active controls. Moreover, the outcomes were less well-defined in both disorders. Third, when they were reported, results often displayed moderate-to-high heterogeneity, which may suggest the presence of subgroups of patients that may better respond to VR than others. Unfortunately, most meta-analyses did not report heterogeneity. Finally, although several meta-analyses did not report publication bias, many noted the presence of publication bias, which may suggest the possibility of either overestimating or underestimating the results. This reinforces the importance of registering the conducted studies.

Furthermore, numerous psychiatric symptoms and disorders that are treatable by VR interventions have not been examined by meta-analytical investigation. For example, although we retrieved only one meta-analysis on compliance for schizophrenia [36], the past decade has seen an emergence in VR treatment for other symptoms such as positive symptoms of psychosis (ie, delusions and auditory verbal hallucinations). These therapies have shown important benefits in psychotic symptomatology, with large effects being observed for both delusions and auditory verbal hallucinations in several trials [71-74]. As a second area of interest, traditional psychological interventions in the field of addiction generally teach individuals new skills to avoid high-risk situations, to refuse substance offers, and to ultimately better cope with cue- and stress-related craving. However, these conventional treatments, such as imaginal cue exposure therapy, have provided mixed findings [75-78], which may be improved by using VR [79]. VR technology may add effectiveness to standard treatments (ie, cue exposure treatment) owing to its capacity to induce greater subjective and physiological craving, which may prompt the generalization of treatment effects to real-life daily activities [80]. A limited amount of research has been conducted to date on the efficacy of VR-based cue exposure approaches for addiction, which aim to extinguish craving and prevent relapses. Promising results from case reports and small trials on subjective and physiological outcomes have emerged for nicotine [81-83], alcohol use disorder [80,84,85], and pathological gambling [86]. Furthermore, research on treatments for eating disorders has paralleled the methods used in the treatment of addiction and adapted them for food cues and environmental settings key to eating behaviors [87]. These interventions have aimed to improve eating disorders, with outcomes including craving, weight regain, and eating patterns [88-91]. Finally, VR may show potential for the treatment of more deviant behaviors such as violence-related outcomes in psychiatric samples. VR may provide a solution to the shortcomings of conventional interventions for violence (ie, clinicians cannot ethically place offenders in at-risk situations) by enabling individuals to be immersed into virtual simulations of real-life events under the control of the clinician [92,93]. Preliminary studies in at-risk populations have shown reductions in anger and impulsivity, improvements in conflict resolution skills and empathy levels, and decreases in aggression [94]. Hitherto, clinical research with novel VR development could make an important contribution to patient care [64], mostly when traditional face-to-face interventions may be more limited or cannot be conducted. Although no meta-analytical evidence was available for the disorders stated above at the time of our literature search, there is preliminary support for the use of VR-based interventions to improve the treatment of symptoms of other psychiatric disorders. Nevertheless, research remains to be generally limited by fewer studies, small samples, lack of control groups (mainly standard evidence-based interventions), and lack of follow-up. In this sense, future research using strong methodology (ie, single-blinded RCTs with large samples) is required to determine whether VR approaches yield additional benefits over standard treatment and whether these effects last over time.

In the above efficacy studies, some key aspects remain to be further investigated. With the rise of personalized medicine, future research should be encouraged to achieve a better understanding of factors that may play a role in VR outcomes and help explain different effects from usual treatment. These factors may include patient characteristics (eg, age, gender, and personality traits) and the severity of the disorder (eg, comorbidities, treatment resistance); certain patients may indeed be more susceptible to better respond to these VR approaches. For instance, a meta-analysis by Cardoş et al [30] on the symptoms of aviophobia found that the age of the participants was a significant moderator, explaining the difference in efficacy of VR-based interventions at posttreatment, with greater effects among younger individuals. In addition, the design of the virtual environments and exposure approach of the therapy may have a role in the therapeutic outcome, which warrants further investigation. Hence, it may be suggested that patients who fully experience the VR paradigm as realistic (ie, higher level of immersion and sense of presence) may respond better to the intervention [31,64]. This may be possible with the use of more recent technologies, which are more immersive and closely resemble the real world. Improved engagement with the virtual environment, with the inclusion of social dynamic interactions via tailored avatars, may similarly have a role in the efficacy of the intervention and heighten the sense of presence and immersion [79]. These dynamic interactions may enable patients to engage with the VR environment in a more naturalistic and intuitive way [95,96]. It is noteworthy that the sense of immersion may be increased by incorporating senses other than vision into the VR environment, such as hearing and smell. Supplemental studies are needed to evaluate the effects of these factors to possibly improve the efficacy of VR-based treatments.

Conclusions

VR provides opportunities to go over and beyond traditional interventions and allows tailoring approaches to each individual, thereby possibly improving efficacy and the maintenance of skills. With variable quality of evidence, meta-analytical literature suggests positive outcomes in the VR treatment of

psychiatric conditions, mainly anxiety-related disorders. VR-based interventions are better than inactive controls and generally show similar effects when compared with evidence-based approaches for these disorders. Preliminary findings also suggest that the effects of VR may be long-lasting. Furthermore, VR has shown efficacy for the treatment of depressive symptoms and neurocognitive disorders. However, support for the use of VR in the treatment of social skills in autism and compliance in schizophrenia is lacking. There are also numerous VR studies that were not included in meta-analyses that targeted other psychiatric symptoms and disorders (ie, psychotic symptoms, addiction); these have also shown prefatory beneficial outcomes. Nevertheless, more research is necessary in the field of psychiatry to establish high-quality evidence with the use of gold-standard evidence from well-designed RCTs comprising large samples. As current VR treatments have not clearly shown superiority over conventional treatments, future VR-based interventions should focus on developing innovative approaches for complex and treatment-resistant symptoms that are difficult to address with traditional treatment. Research is also warranted to evaluate the aspects enabling the better use of VR and examine the specificity of VR-based interventions. As soon as more studies become available, systematic meta-regression analyses could statistically examine the influence of certain variables on the efficacy of VR for improving personalized patient care.

Acknowledgments

No funding was provided for this study. Nonetheless, LD and ML are holders of scholarships from the Fonds de Recherche du Québec en Santé. SP is the holder of the Eli Lilly Canada Chair in schizophrenia research. AD is the holder of a junior 2 salary award from the Fonds de Recherche du Québec en Santé.

Authors' Contributions

LD, SP, and AD contributed to study planning and design. LD and ML conducted the literature search. LD wrote the paper. All authors provided critical revisions for the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Electronic search strategy for the meta-review conducted.
[DOCX File , 21 KB - jmir_v22i8e20889_app1.docx]

Multimedia Appendix 2
Details of the retrieved studies included in the meta-review.
[DOCX File , 51 KB - jmir_v22i8e20889_app2.docx]

Multimedia Appendix 3
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Checklist.
[DOCX File , 25 KB - jmir_v22i8e20889_app3.docx]

References

1. Charlson F, Baxter A, Dua T, Degenhardt L, Whiteford H, Vos T. Excess mortality from mental, neurological, substance use disorders in the global burden of disease study 2010. In: Patel V, Chisholm D, Dua T, Laxminarayan R, Medina-Mora ME, editors. Disease Control Priorities, Third Edition (Volume 4): Mental, Neurological, and Substance Use Disorders. Washington, DC: The World Bank; 2016.
2. Bose J, Hedden SL, Lipari RN, Park-Lee E. Key Substance Use and Mental Health Indicators in the United States: Results from the 2017 National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration. 2018. URL: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHFFR2017/NSDUHFFR2017.htm [accessed 2020-08-11]
3. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012 Sep;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
4. Bandelow B, Michaelis S. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin Neurosci 2015 Sep;17(3):327-335 [FREE Full text] [Medline: 26487813]
5. Pratt LA, Brody DJ. Depression in the US household population, 2009-2012. NCHS Data Brief 2014 Dec(172):1-8 [FREE Full text] [Medline: 25470183]

6. Whitney DG, Peterson MD. US national and state-level prevalence of mental health disorders and disparities of mental health care use in children. JAMA Pediatr 2019 Apr 1;173(4):389-391 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.5399] [Medline: 30742204]
7. Kessler RC, Angermeyer M, Anthony JC, de Graaf R, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's world mental health survey initiative. World Psychiatry 2007 Oct;6(3):168-176 [FREE Full text] [Medline: 18188442]
8. Baranne ML, Falissard B. Global burden of mental disorders among children aged 5-14 years. Child Adolesc Psychiatry Ment Health 2018;12:19 [FREE Full text] [doi: 10.1186/s13034-018-0225-4] [Medline: 29682005]
9. Rehm J, Shield KD. Global burden of disease and the impact of mental and addictive disorders. Curr Psychiatry Rep 2019 Feb 7;21(2):10. [doi: 10.1007/s11920-019-0997-0] [Medline: 30729322]
10. Trautmann S, Rehm J, Wittchen H. The economic costs of mental disorders: do our societies react appropriately to the burden of mental disorders? EMBO Rep 2016 Sep;17(9):1245-1249 [FREE Full text] [doi: 10.15252/embr.201642951] [Medline: 27491723]
11. Huhn M, Tardy M, Spineli LM, Kissling W, Förstl H, Pitschel-Walz G, et al. Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders: a systematic overview of meta-analyses. JAMA Psychiatry 2014 Jun;71(6):706-715. [doi: 10.1001/jamapsychiatry.2014.112] [Medline: 24789675]
12. Harvey AG, Gumport NB. Evidence-based psychological treatments for mental disorders: modifiable barriers to access and possible solutions. Behav Res Ther 2015 May;68:1-12 [FREE Full text] [doi: 10.1016/j.brat.2015.02.004] [Medline: 25768982]
13. Stafford MR, Mayo-Wilson E, Loucas CE, James A, Hollis C, Birchwood M, et al. Efficacy and safety of pharmacological and psychological interventions for the treatment of psychosis and schizophrenia in children, adolescents and young adults: a systematic review and meta-analysis. PLoS One 2015;10(2):e0117166 [FREE Full text] [doi: 10.1371/journal.pone.0117166] [Medline: 25671707]
14. Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. J Am Med Assoc 2010 Jan 6;303(1):47-53 [FREE Full text] [doi: 10.1001/jama.2009.1943] [Medline: 20051569]
15. Barbui C, Purgato M, Abdulmalik J, Acarturk C, Eaton J, Gastaldon C, et al. Efficacy of psychosocial interventions for mental health outcomes in low-income and middle-income countries: an umbrella review. Lancet Psych 2020 Feb;7(2):162-172. [doi: 10.1016/s2215-0366(19)30511-5] [Medline: 31948935]
16. Menon J, Kandasamy A. Relapse prevention. Indian J Psychiatry 2018;60(Suppl 4):S473-S478. [doi: 10.4103/psychiatry.IndianJPsychiatry_36_18] [Medline: 29540916]
17. Swift JK, Greenberg RP. A treatment by disorder meta-analysis of dropout from psychotherapy. J Psychother Integr 2014;24(3):193-207. [doi: 10.1037/a0037512]
18. Rizzo A, Kim GJ. A SWOT analysis of the field of virtual reality rehabilitation and therapy. Presence-Teleop Virt 2005 Apr;14(2):119-146. [doi: 10.1162/1054746053967094]
19. Fernández-Alvarez J, Colombo D, Botella C, García-Palacios A, Riva G. Virtual reality for anxiety and stress-related disorders: a SWOT analysis. In: Cipresso P, Serino S, Villani D, editors. Pervasive Computing Paradigms for Mental Health. New York, USA: Springer; 2019.
20. Rebenitsch L, Owen C. Review on cybersickness in applications and visual displays. Virtual Reality 2016 Apr 26;20(2):101-125. [doi: 10.1007/s10055-016-0285-9]
21. Fernández-Álvarez J, Rozental A, Carlbring P, Colombo D, Riva G, Anderson PL, et al. Deterioration rates in virtual reality therapy: an individual patient data level meta-analysis. J Anxiety Disord 2019 Jan;61:3-17. [doi: 10.1016/j.janxdis.2018.06.005] [Medline: 30057346]
22. Riva G. Virtual reality in psychotherapy: review. Cyberpsychol Behav 2005 Jun;8(3):220-30; discussion 231. [doi: 10.1089/cpb.2005.8.220] [Medline: 15971972]
23. Park MJ, Kim DJ, Lee U, Na EJ, Jeon HJ. A literature overview of virtual reality (VR) in treatment of psychiatric disorders: recent advances and limitations. Front Psychiatry 2019;10:505 [FREE Full text] [doi: 10.3389/fpsyt.2019.00505] [Medline: 31379623]
24. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: an systematic review of recent evidence. Psychiatry Res 2016 Feb 28;236:189-195. [doi: 10.1016/j.psychres.2016.01.015] [Medline: 26795129]
25. Petitti D. Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine. New York, USA: Oxford University Press; 2000.
26. Valentine JC, Pigott TD, Rothstein HR. How many studies do you need? J Educ Behav Stat 2010 Apr;35(2):215-247. [doi: 10.3102/1076998609346961]
27. Schmidt F, Hunter J. Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. Thousand Oaks, CA: Sage Publications; 2015.
28. Opriş D, Pintea S, García-Palacios A, Botella C, Szamosközi S, David D. Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis. Depress Anxiety 2012 Feb;29(2):85-93. [doi: 10.1002/da.20910] [Medline: 22065564]

29. Parsons TD, Rizzo AA. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. J Behav Ther Exp Psychiatry 2008 Sep;39(3):250-261. [doi: 10.1016/j.jbtep.2007.07.007] [Medline: 17720136]
30. Cardoş RA, David OA, David DO. Virtual reality exposure therapy in flight anxiety: a quantitative meta-analysis. Comput Hum Behav 2017 Jul;72:371-380. [doi: 10.1016/j.chb.2017.03.007]
31. Fodor LA, Cote CD, Cuijpers P, Szamoskozi S, David D, Cristea IA. The effectiveness of virtual reality based interventions for symptoms of anxiety and depression: a meta-analysis. Sci Rep 2018 Jul 9;8(1):10323 [FREE Full text] [doi: 10.1038/s41598-018-28113-6] [Medline: 29985400]
32. Kampmann IL, Emmelkamp PM, Morina N. Meta-analysis of technology-assisted interventions for social anxiety disorder. J Anxiety Disord 2016 Aug;42:71-84. [doi: 10.1016/j.janxdis.2016.06.007] [Medline: 27376634]
33. Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, et al. Virtual reality exposure therapy for anxiety and related disorders: a meta-analysis of randomized controlled trials. J Anxiety Disord 2019 Jan;61:27-36. [doi: 10.1016/j.janxdis.2018.08.003] [Medline: 30287083]
34. Chesham RK, Malouff JM, Schutte NS. Meta-analysis of the efficacy of virtual reality exposure therapy for social anxiety. Behav Change 2018 Jun 18;35(3):152-166. [doi: 10.1017/bec.2018.15]
35. Deng W, Hu D, Xu S, Liu X, Zhao J, Chen Q, et al. The efficacy of virtual reality exposure therapy for PTSD symptoms: a systematic review and meta-analysis. J Affect Disord 2019 Oct 1;257:698-709. [doi: 10.1016/j.jad.2019.07.086] [Medline: 31382122]
36. Välimäki M, Hätönen HM, Lahti M, Kurki M, Hottinen A, Metsäranta K, et al. Virtual reality for treatment compliance for people with serious mental illness. Cochrane Database Syst Rev 2014 Oct 8(10):CD009928. [doi: 10.1002/14651858.CD009928.pub2] [Medline: 25300174]
37. Barton EE, Pustejovsky JE, Maggin DM, Reichow B. Technology-aided instruction and intervention for students with ASD: a meta-analysis using novel methods of estimating effect sizes for single-case research. Remedial Spec Educ 2017 Oct 6;38(6):371-386. [doi: 10.1177/0741932517729508]
38. Kim O, Pang Y, Kim J. The effectiveness of virtual reality for people with mild cognitive impairment or dementia: a meta-analysis. BMC Psychiatry 2019 Jul 12;19(1):219 [FREE Full text] [doi: 10.1186/s12888-019-2180-x] [Medline: 31299921]
39. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, USA: Academic Press; 2013.
40. Guyatt G, Oxman A, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol 2011 Apr;64(4):407-415. [doi: 10.1016/j.jclinepi.2010.07.017] [Medline: 21247734]
41. Guyatt G, Oxman A, Kunz R, Woodcock J, Brozek J, Helfand M, GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol 2011 Dec;64(12):1303-1310. [doi: 10.1016/j.jclinepi.2011.04.014] [Medline: 21802903]
42. Guyatt G, Oxman A, Kunz R, Woodcock J, Brozek J, Helfand M, GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J Clin Epidemiol 2011 Dec;64(12):1294-1302. [doi: 10.1016/j.jclinepi.2011.03.017] [Medline: 21803546]
43. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. J Clin Epidemiol 2011 Dec;64(12):1277-1282. [doi: 10.1016/j.jclinepi.2011.01.011] [Medline: 21802904]
44. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
45. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5. Washington, DC: American Psychiatric Pub; 2013.
46. Wiederhold BK, Riva G. Virtual reality therapy: emerging topics and future challenges. Cyberpsychol Behav Soc Netw 2019 Jan;22(1):3-6. [doi: 10.1089/cyber.2018.29136.bkw] [Medline: 30649958]
47. Morina N, Ijntema H, Meyerbröker K, Emmelkamp PM. Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. Behav Res Ther 2015 Nov;74:18-24. [doi: 10.1016/j.brat.2015.08.010] [Medline: 26355646]
48. Powers MB, Emmelkamp PM. Virtual reality exposure therapy for anxiety disorders: a meta-analysis. J Anxiety Disord 2008;22(3):561-569. [doi: 10.1016/j.janxdis.2007.04.006] [Medline: 17544252]
49. Abramowitz JS. The practice of exposure therapy: relevance of cognitive-behavioral theory and extinction theory. Behav Ther 2013 Dec;44(4):548-558. [doi: 10.1016/j.beth.2013.03.003] [Medline: 24094780]
50. Foa EB, Kozak MJ. Emotional processing of fear: exposure to corrective information. Psychol Bull 1986 Jan;99(1):20-35. [Medline: 2871574]
51. Foa E, Huppert J, Cahill S. Emotional processing theory: an update. In: Pathological Anxiety: Emotional Processing in Etiology and Treatment. New York, USA: The Guilford Press; 2006.
52. Craske MG, Kircanski K, Zelikowsky M, Mystkowski J, Chowdhury N, Baker A. Optimizing inhibitory learning during exposure therapy. Behav Res Ther 2008 Jan;46(1):5-27. [doi: 10.1016/j.brat.2007.10.003] [Medline: 18005936]

53. Slater M, Steed A, McCarthy J, Maringelli F. The influence of body movement on subjective presence in virtual environments. Hum Factors 1998 Sep;40(3):469-477. [doi: 10.1518/001872098779591368] [Medline: 9849105]
54. Wiederhold B, Wiederhold M. Advances in Virtual Reality and Anxiety Disorders. Washington, DC: American Psychological Association; 2005.
55. Lindner P, Miloff A, Hamilton W, Reuterskiöld L, Andersson G, Powers MB, et al. Creating state of the art, next-generation virtual reality exposure therapies for anxiety disorders using consumer hardware platforms: design considerations and future directions. Cogn Behav Ther 2017 Sep;46(5):404-420. [doi: 10.1080/16506073.2017.1280843] [Medline: 28270059]
56. Krijn M, Emmelkamp P, Olafsson R, Biemond R. Virtual reality exposure therapy of anxiety disorders: a review. Clin Psychol Rev 2004 Jul;24(3):259-281. [doi: 10.1016/j.cpr.2004.04.001] [Medline: 15245832]
57. Cusack K, Jonas DE, Forneris CA, Wines C, Sonis J, Middleton JC, et al. Psychological treatments for adults with posttraumatic stress disorder: a systematic review and meta-analysis. Clin Psychol Rev 2016 Feb;43:128-141. [doi: 10.1016/j.cpr.2015.10.003] [Medline: 26574151]
58. Botella C, García-Palacios A, Guillen V, Baños RM, Quero S, Alcaniz M. An adaptive display for the treatment of diverse trauma PTSD victims. Cyberpsychol Behav Soc Netw 2010 Feb;13(1):67-71. [doi: 10.1089/cyber.2009.0353] [Medline: 20528295]
59. Gerardi M, Cukor J, Difede J, Rizzo A, Rothbaum BO. Virtual reality exposure therapy for post-traumatic stress disorder and other anxiety disorders. Curr Psychiatry Rep 2010 Aug;12(4):298-305. [doi: 10.1007/s11920-010-0128-4] [Medline: 20535592]
60. Gonçalves R, Pedrozo AL, Coutinho ES, Figueira I, Ventura P. Efficacy of virtual reality exposure therapy in the treatment of PTSD: a systematic review. PLoS One 2012;7(12):e48469 [FREE Full text] [doi: 10.1371/journal.pone.0048469] [Medline: 23300515]
61. Shah LB, Torres S, Kannusamy P, Chng CM, He H, Klainin-Yobas P. Efficacy of the virtual reality-based stress management program on stress-related variables in people with mood disorders: the feasibility study. Arch Psychiatr Nurs 2015 Feb;29(1):6-13. [doi: 10.1016/j.apnu.2014.09.003] [Medline: 25634868]
62. Gould RA, Mueser KT, Bolton E, Mays V, Goff D. Cognitive therapy for psychosis in schizophrenia: an effect size analysis. Schizophr Res 2001 Mar 30;48(2-3):335-342. [doi: 10.1016/s0920-9964(00)00145-6] [Medline: 11295385]
63. Lindner P, Hamilton W, Miloff A, Carlbring P. How to treat depression with low-intensity virtual reality interventions: perspectives on translating cognitive behavioral techniques into the virtual reality modality and how to make anti-depressive use of virtual reality-unique experiences. Front Psychiatry 2019;10:792 [FREE Full text] [doi: 10.3389/fpsyt.2019.00792] [Medline: 31736809]
64. Maples-Keller JL, Bunnell BE, Kim S, Rothbaum BO. The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. Harv Rev Psychiatry 2017;25(3):103-113 [FREE Full text] [doi: 10.1097/HRP.0000000000000138] [Medline: 28475502]
65. Rus-Calafell M, Garety P, Sason E, Craig TJ, Valmaggia LR. Virtual reality in the assessment and treatment of psychosis: a systematic review of its utility, acceptability and effectiveness. Psychol Med 2018 Feb;48(3):362-391. [doi: 10.1017/S0033291717001945] [Medline: 28735593]
66. Parsons S. Authenticity in virtual reality for assessment and intervention in autism: a conceptual review. Educ Res Rev 2016 Nov;19:138-157. [doi: 10.1016/j.edurev.2016.08.001]
67. Parsons S, Mitchell P. The potential of virtual reality in social skills training for people with autistic spectrum disorders. J Intellect Disabil Res 2002 Jun;46(Pt 5):430-443. [doi: 10.1046/j.1365-2788.2002.00425.x] [Medline: 12031025]
68. Botella C, Fernández-Álvarez J, Guillén V, García-Palacios A, Baños R. Recent progress in virtual reality exposure therapy for phobias: a systematic review. Curr Psychiatry Rep 2017 Jul;19(7):42. [doi: 10.1007/s11920-017-0788-4] [Medline: 28540594]
69. Emmelkamp PM. Technological innovations in clinical assessment and psychotherapy. Psychother Psychosom 2005;74(6):336-343. [doi: 10.1159/000087780] [Medline: 16244509]
70. Krijn M, Emmelkamp PM, Olafsson RP, Bouwman M, van Gerwen LJ, Spinhoven P, et al. Fear of flying treatment methods: virtual reality exposure vs cognitive behavioral therapy. Aviat Space Environ Med 2007 Feb;78(2):121-128. [Medline: 17310883]
71. Freeman D, Bradley J, Antley A, Bourke E, DeWeever N, Evans N, et al. Virtual reality in the treatment of persecutory delusions: randomised controlled experimental study testing how to reduce delusional conviction. Br J Psychiatry 2016 Jul;209(1):62-67 [FREE Full text] [doi: 10.1192/bjp.bp.115.176438] [Medline: 27151071]
72. Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry 2018 Jan;5(1):31-40. [doi: 10.1016/s2215-0366(17)30427-3]
73. Leff J, Williams G, Huckvale MA, Arbuthnot M, Leff AP. Computer-assisted therapy for medication-resistant auditory hallucinations: proof-of-concept study. Br J Psychiatry 2013 Jun;202:428-433. [doi: 10.1192/bjp.bp.112.124883] [Medline: 23429202]

74. du Sert OP, Potvin S, Lipp O, Dellazizzo L, Laurelli M, Breton R, et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: a pilot clinical trial. Schizophr Res 2018 Jul;197:176-181. [doi: 10.1016/j.schres.2018.02.031] [Medline: 29486956]
75. Conklin C, Tiffany S. Applying extinction research and theory to cue-exposure addiction treatments. Addiction 2002 Feb;97(2):155-167. [doi: 10.1046/j.1360-0443.2002.00014.x] [Medline: 11860387]
76. McHugh RK, Hearon BA, Otto MW. Cognitive behavioral therapy for substance use disorders. Psychiatr Clin North Am 2010 Sep;33(3):511-525 [FREE Full text] [doi: 10.1016/j.psc.2010.04.012] [Medline: 20599130]
77. Jhanjee S. Evidence based psychosocial interventions in substance use. Indian J Psychol Med 2014 Apr;36(2):112-118 [FREE Full text] [doi: 10.4103/0253-7176.130960] [Medline: 24860208]
78. Carroll KM, Onken LS. Behavioral therapies for drug abuse. Am J Psychiatry 2005 Aug;162(8):1452-1460 [FREE Full text] [doi: 10.1176/appi.ajp.162.8.1452] [Medline: 16055766]
79. Hone-Blanchet A, Wensing T, Fecteau S. The use of virtual reality in craving assessment and cue-exposure therapy in substance use disorders. Front Hum Neurosci 2014;8:844 [FREE Full text] [doi: 10.3389/fnhum.2014.00844] [Medline: 25368571]
80. Lee J, Kwon H, Choi J, Yang B. Cue-exposure therapy to decrease alcohol craving in virtual environment. Cyberpsychol Behav 2007 Oct;10(5):617-623. [doi: 10.1089/cpb.2007.9978] [Medline: 17927528]
81. Choi J, Park S, Lee J, Jung H, Lee H, Jin C, et al. The effect of repeated virtual nicotine cue exposure therapy on the psychophysiological responses: a preliminary study. Psychiatry Investig 2011 Jun;8(2):155-160 [FREE Full text] [doi: 10.4306/pi.2011.8.2.155] [Medline: 21852993]
82. Pericot-Valverde I, García-Rodríguez O, Ferrer-García M, Secades-Villa R, Gutiérrez-Maldonado J. Virtual reality for smoking cessation: a case report. Stud Health Technol Inform 2012;181:292-296. [Medline: 22954874]
83. Park C, Choi J, Park SM, Lee J, Jung HY, Seol J, et al. Comparison of the effectiveness of virtual cue exposure therapy and cognitive behavioral therapy for nicotine dependence. Cyberpsychol Behav Soc Netw 2014 Apr;17(4):262-267 [FREE Full text] [doi: 10.1089/cyber.2013.0253] [Medline: 24555521]
84. Son JH, Lee SH, Seok JW, Kee BS, Lee HW, Kim HJ, et al. Virtual reality therapy for the treatment of alcohol dependence: a preliminary investigation with positron emission tomography/computerized tomography. J Stud Alcohol Drugs 2015 Jul;76(4):620-627. [doi: 10.15288/jsad.2015.76.620] [Medline: 26098039]
85. Lee SH, Han DH, Oh S, Lyoo IK, Lee YS, Renshaw PF, et al. Quantitative electroencephalographic (qEEG) correlates of craving during virtual reality therapy in alcohol-dependent patients. Pharmacol Biochem Behav 2009 Jan;91(3):393-397. [doi: 10.1016/j.pbb.2008.08.014] [Medline: 18771681]
86. Bouchard S, Robillard G, Giroux I, Jacques C, Loranger C, St-Pierre M, et al. Using virtual reality in the treatment of gambling disorder: the development of a new tool for cognitive behavior therapy. Front Psychiatry 2017;8:27 [FREE Full text] [doi: 10.3389/fpsyt.2017.00027] [Medline: 28286486]
87. Bordnick PS, Carter BL, Traylor AC. What virtual reality research in addictions can tell us about the future of obesity assessment and treatment. J Diabetes Sci Technol 2011 Mar 1;5(2):265-271 [FREE Full text] [doi: 10.1177/193229681100500210] [Medline: 21527092]
88. de Carvalho M, Dias T, Duchesne M, Nardi A, Appolinario J. Virtual reality as a promising strategy in the assessment and treatment of bulimia nervosa and binge eating disorder: a systematic review. Behav Sci (Basel) 2017 Jul 9;7(3):43 [FREE Full text] [doi: 10.3390/bs7030043] [Medline: 28698483]
89. Gutiérrez-Maldonado J, Wiederhold BK, Riva G. Future directions: how virtual reality can further improve the assessment and treatment of eating disorders and obesity. Cyberpsychol Behav Soc Netw 2016 Feb;19(2):148-153. [doi: 10.1089/cyber.2015.0412] [Medline: 26378982]
90. Manzoni GM, Cesa GL, Bacchetta M, Castelnuovo G, Conti S, Gaggioli A, et al. Virtual reality-enhanced cognitive-behavioral therapy for morbid obesity: a randomized controlled study with 1 year follow-up. Cyberpsychol Behav Soc Netw 2016 Feb;19(2):134-140 [FREE Full text] [doi: 10.1089/cyber.2015.0208] [Medline: 26430819]
91. Clus D, Larsen ME, Lemey C, Berrouiguet S. The use of virtual reality in patients with eating disorders: systematic review. J Med Internet Res 2018 Apr 27;20(4):e157 [FREE Full text] [doi: 10.2196/jmir.7898] [Medline: 29703715]
92. Hubal RC, Fishbein DH, Sheppard MS, Paschall MJ, Eldreth DL, Hyde CT. How do varied populations interact with embodied conversational agents? Findings from inner-city adolescents and prisoners. Comput Human Behav 2008 May;24(3):1104-1138 [FREE Full text] [doi: 10.1016/j.chb.2007.03.010] [Medline: 19412316]
93. Rovira A, Swapp D, Spanlang B, Slater M. The use of virtual reality in the study of people's responses to violent incidents. Front Behav Neurosci 2009;3:59 [FREE Full text] [doi: 10.3389/neuro.08.059.2009] [Medline: 20076762]
94. Dellazizzo L, Potvin S, Bahig S, Dumais A. Comprehensive review on virtual reality for the treatment of violence: implications for youth with schizophrenia. NPJ Schizophr 2019 Jul 23;5(1):11 [FREE Full text] [doi: 10.1038/s41537-019-0079-7] [Medline: 31337763]
95. Rizzo AA, Buckwalter JG, Neumann U. Virtual reality and cognitive rehabilitation: a brief review of the future. J Head Trauma Rehab 1997;12(6):1-15. [doi: 10.1097/00001199-199712000-00002]
96. Anderson PL, Rothbaum BO, Hodges L. Virtual reality: using the virtual world to improve quality of life in the real world. Bull Menninger Clin 2001;65(1):78-91. [doi: 10.1521/bumc.65.1.78.18713] [Medline: 11280960]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
VR: virtual reality

Edited by G Eysenbach; submitted 01.06.20; peer-reviewed by M Noll-Hussong, A Tsanas; comments to author 01.07.20; revised version received 12.07.20; accepted 14.07.20; published 19.08.20.

Please cite as:
Dellazizzo L, Potvin S, Luigi M, Dumais A
Evidence on Virtual Reality–Based Therapies for Psychiatric Disorders: Meta-Review of Meta-Analyses
J Med Internet Res 2020;22(8):e20889
URL: http://www.jmir.org/2020/8/e20889/
doi:10.2196/20889
PMID:32812889

Original Paper

Exploring eHealth Literacy and Patient-Reported Experiences With Outpatient Care in the Hungarian General Adult Population: Cross-Sectional Study

Zsombor Zrubka[1,2], MD, MBA, PhD; Óscar Brito Fernandes[1,3], MSc; Petra Baji[1], PhD, Habil; Ottó Hajdu[4], DSC; Levente Kovacs[5], PhD, Habil; Dionne Kringos[3], PhD; Niek Klazinga[3], MD, PhD; László Gulácsi[1,2], MD, DSC; Valentin Brodszky[1], MD, PhD, Habil; Fanni Rencz[1,6], MD, PhD; Márta Péntek[1], MD, PhD, Habil

[1]Department of Health Economics, Corvinus University of Budapest, Budapest, Hungary

[2]University Research and Innovation Center, Óbuda University, Budapest, Hungary

[3]Department of Public and Occupational Health, Amsterdam UMC, Amsterdam Public Health Research Institute, University of Amsterdam, Amsterdam, Netherlands

[4]Department of Comparative Economics, Eötvös Loránd University, Budapest, Hungary

[5]Physiological Controls Research Center, University Research and Innovation Center, Óbuda University, Budapest, Hungary

[6]Premium Postdoctoral Research Program, Hungarian Academy of Sciences, Budapest, Hungary

Corresponding Author:
Zsombor Zrubka, MD, MBA, PhD
Department of Health Economics
Corvinus University of Budapest
Fővám tér 8.
Budapest, 1093
Hungary
Phone: 36 30 202 9415
Email: zsombor.zrubka@uni-corvinus.hu

Abstract

Background: Digital health, which encompasses the use of information and communications technology in support of health, is a key driving force behind the cultural transformation of medicine toward people-centeredness. Thus, eHealth literacy, assisted by innovative digital health solutions, may support better experiences of care.

Objective: The purpose of this study is to explore the relationship between eHealth literacy and patient-reported experience measures (PREMs) among users of outpatient care in Hungary.

Methods: In early 2019, we conducted a cross-sectional survey on a large representative online sample recruited from the Hungarian general population. eHealth literacy was measured with the eHealth Literacy Scale (eHEALS). PREMs with outpatient care were measured with a set of questions recommended by the Organisation for Economic Co-operation and Development (OECD) for respondents who attended outpatient visit within 12 months preceding the survey. Bivariate relationships were explored via polychoric correlation, the Kruskal–Wallis test, and chi-square test. To capture nonlinear associations, after controlling covariates, we analyzed the relationship between eHEALS quartiles and PREMs using multivariate probit, ordinary least squares, ordered logit, and logistic regression models.

Results: From 1000 survey respondents, 666 individuals (364 females, 54.7%) were included in the study with mean age of 48.9 (SD 17.6) years and mean eHEALS score of 29.3 (SD 4.9). Respondents with higher eHEALS scores were more likely to understand the health care professionals' (HCPs') explanations (χ^2_9=24.2, *P*=.002) and to be involved in decision making about care and treatment (χ^2_9=18.2, *P*=.03). In multivariate regression, respondents with lowest (first quartile) and moderately high (third quartile) eHEALS scores differed significantly, where the latter were more likely to have an overall positive experience (*P*=.02) and experience fewer problems (*P*=.02). In addition, those respondents had better experiences in terms of how easy it was to understand the HCPs' explanations (*P*<.001) and being able to ask questions during their last consultation (*P*=.04). Patient-reported experiences of individuals with highest (fourth quartile) and lowest (first quartile) eHEALS levels did not differ significantly in any items of the PREM instrument, and neither did composite PREM scores generated from the PREM items (*P*>.05 in all models).

Conclusions: We demonstrated the association between eHealth literacy and PREMs. The potential patient-, physician-, and system-related factors explaining the negative experiences among people with highest levels of eHealth literacy warrant further investigation, which may contribute to the development of efficient eHealth literacy interventions. Further research is needed to establish causal relationship between eHealth literacy and patient-reported experiences.

(J Med Internet Res 2020;22(8):e19013) doi:10.2196/19013

KEYWORDS

health literacy; eHealth literacy, patient-reported experience measures; patient-reported outcome measures; ambulatory care; shared decision making; Hungary; survey

Introduction

People-centeredness has shaped the cultural transformation of medicine, where we transitioned from a traditional paternalistic model toward a new model of care, grounded in partnerships and putting patients' values and preferences in the forefront of medical decision making [1,2]. To cope with the ever-increasing pressure on health care budgets [3], improving the technical efficiency of health care systems remains a key imperative, in which eliminating waste and maximizing the value that matters to patients is a top priority for developed economies [4-7]. Therefore, in addition to its humanistic merits, people-centeredness has a strong economic rationale: enhancing the participation of patients in the health production process [4-8]. Hence, to involve people in health production and operationalize people-centered care, it is necessary that people have the information, education, and support they need to inform decision making about their own care [8].

Digital health is a key driving force behind the cultural transformation of medicine toward people-centeredness [9,10]. Digital health encompasses the use of information and communications technology in support of health and health-related fields (electronic health [eHealth]) and emerging areas such as the use of computing sciences in big data, artificial intelligence, and genomics [11,12]. Currently, the implementation of digital tools that facilitate people-centered care is a priority for policymakers [13]. With the digital transformation of health care, people have easier access to health information from online sources; simultaneously, people are called to assume responsibility for evaluating the accuracy and reliability of such information, and thus, rely on their eHealth literacy. eHealth literacy has been defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [14]. Higher levels of eHealth literacy have been associated with better subjective health status [15-17], healthier lifestyle [18-21], and lower risk of chronic disease [22]. Furthermore, with appropriate interventions eHealth literacy has been shown to be a modifiable factor even in later stages of life [23-25]. With the spread of people-centric values, patient-reported experience measures (PREMs)—quality indicators of health care from the patients' perspective—have gained international attention. Signaling this is the work developed by the Organisation for Economic Co-operation and Development (OECD), where PREMs are used to evaluate the performance of health systems through a patient's experience of care, including that of the patient–physician relationship [26].

Hungary has a tax-funded single-payer health system providing universal health coverage for the population. Most inpatient and specialist ambulatory care services are delivered by the public health system. Primary care—provided by private general practitioners (GPs)—acts as a gatekeeper. Per-capita spending on health care is among the lowest and the share of out-of-pocket contributions including informal payments is among the highest within the European Union. Life expectancy lags behind most European Union countries, mainly driven by lifestyle-related causes. Health inequalities are largely determined by sociodemographic variables [27-29]. Recent evidence suggested the need for improving patients' experience of care in Hungarian outpatient settings with regard to shared decision making [30]. This is aligned with findings of another study, which showed that patients' preferences and interests were less likely to be taken into consideration by GPs in Hungary, in comparison with other European countries [29].

By making inferences from traditional health literacy studies to the eHealth domain [31], one may assume that higher eHealth literacy enables patients to orientate better in the health care system, find better access to care, use online information efficiently to reduce waiting times, and make more out of the interaction with health care professionals (HCPs). Among others, these factors may ultimately lead to better overall experience. Online health information may enhance the patient–physician relationship and although the evidence is mixed [32,33], the importance of eHealth literacy in translating the benefits of innovative digital health solutions to better experiences of care has been recognized [34-36]. For example, positive relationship between eHealth literacy and shared decision making has been found [37] and its contribution to patients' decision-making styles has been demonstrated [32]. However, to our knowledge, no studies have focused on the relationship between eHealth literacy and PREMs to date.

This study aims to explore the association between eHealth literacy and OECD's set of recommend PREMs for users of outpatient care, who were recruited from a large representative online sample from the Hungarian general adult population.

Methods

Study Design and Sample

We considered the CHERRIES checklist when reporting this study [38]. In early 2019, we conducted a large cross-sectional interned-based survey among the general adult population of Hungary, which explored eHealth literacy [39]. In addition, shared decision making [40] and PREMs [30,41] were

investigated among those respondents, who used ambulatory care over the past 12 months due to health problems. We recruited 1000 online respondents. Quotas were used to ensure the representativeness of the sample according to the 2011 Population Census [42] data by gender, age, educational level, type of settlement, and NUTS 1 (Nomenclature of Territorial Units for Statistics) region of residence, including a fair representation of people aged 65 and over. Recruitment and data collection were carried out from a commercial online panel by a survey company (Big Data Scientist Kft); reports on dropout rates and the sampling frame were not available. All materials were in Hungarian and all participants spoke the same language (Hungarian). Participation was voluntary and anonymous. No incentives were offered for answering the survey. Participants gave their informed consent prior to the study. Ethical approval was obtained from the Hungarian Medical Research Council (ID: 47654-2/2018/EKU). The electronic questionnaire was piloted by the authors. Respondents could revise and change their answers for completed items, and full completion was required unless the "do not know/do not want to answer" option was offered.

Our sample included those respondents who had a face-to-face appointment with an HCP in the previous 12 months due to their own health problems and answered whether or not the visit had happened at their usual HCP.

eHealth Literacy Scale (eHEALS)

eHealth literacy was measured with the Hungarian version of the self-reported eHealth Literacy Scale (eHEALS) [39,43]. This instrument has been used internationally both as a descriptive tool [22,39] and as a patient-reported outcome measure of digital health interventions [19,23,43]. The eHEALS consists of 8 items, each scored on a 5-point Likert scale. Items 1 and 2 are related to the awareness of health resources ("I know what health resources are available on the Internet"; "I know where to find helpful health resources on the Internet"); items 3 and 4 are related to searching for health resources ("I know how to find helpful health resources on the Internet"; "I know how to use the Internet to answer my questions about health"); items 5 and 8 are related to the utilization of health resources ("I know how to use the health information I find on the Internet to help me"; "I feel confident in using information from the Internet to make health decisions"); and items 6 and 7 are related to the appraisal of health resources ("I have the skills I need to evaluate the health resources I find on the Internet"; "I can tell high quality health resources from low quality health resources on the Internet"). Item levels are added for a total score ranging from 8 to 40. Higher scores indicate greater eHealth literacy [43]. The psychometric properties of the Hungarian eHEALS as well as its association with health outcomes and behavioral health risk factors in the general population have been shown in the validation study [39]. Because eHEALS showed convex relationship with self-rated health in the validation study, we decided to group respondents into 4 quartiles to explore potential nonlinear associations with PREMs in an easily interpretable manner [39,44,45]. We also performed sensitivity analysis using alternative eHEALS category boundaries. The eHEALS questionnaire is included in Multimedia Appendix 1.

OECD-Proposed Set of Questions on Patients' Experiences with Ambulatory Care (PREM)

Respondents' experiences with ambulatory care were assessed by the set of questions recommended by the OECD's Health Care Quality Indicators Project [26]. The questionnaire consists of 2 sections: (1) access to care, including questions on unmet medical needs and waiting times; and (2) patient experiences. The *access to care* section consists of 4 binary questions about unmet medical needs over the past 12 months (missed medical visits, interventions, or medications due to travel difficulties, or cost burden) and 4 questions on waiting times concerning the last medical appointment: waiting time to get the appointment (appointment waiting time) and on the day of consultation (office waiting time), and whether waiting was a problem in either case. In the *patient experiences* section, respondents were asked if the HCP (1) spent enough time with them; (2) provided easy-to-understand explanations; (3) provided the opportunity to ask questions or raise concerns about recommended treatment; and (4) involved in decisions about care and treatment as much as the respondent wanted to be. Answers were recorded on a 4-point Likert scale, with higher scores indicating more perceived problems. In the final item, respondents rated their perception of the overall quality of the appointment on a 5-point Likert scale ranging from *poor* (0) to *excellent* (4). Additional items inquired the HCP type, setting and time of the last visit, and whether the respondent visited his/her usual HCP. Questions related to unmet medical needs were posed if respondents had had health problems over 12 months preceding the survey, while waiting times and patient experiences were inquired only if the respondent had participated in ambulatory consultation with an HCP. The full PREM questionnaire is included in Multimedia Appendix 2.

PREM Scores

Following the practice of countries using PREMs for monitoring health system performance, we created composite scores from PREM items [26]. The *Unmet Medical Needs Score* (range 0-4) reflected the number of areas where respondents experienced an unmet need (missed visit due to travel burden; missed visit due to cost burden; missed intervention due to cost burden; and missed medication due to cost burden). The binary *Any Unmet Medical Need* variable indicated if patients experienced unmet need in any of the 4 items. Waiting times were transformed to continuous variables by considering the midpoint of the respective waiting time answer option. Log-waiting times were used in regression analyses, assuming a 0.5-day waiting time for respondents with an appointment on the same day. We used the binary variable *Any Waiting Problem* to indicate if the office waiting time or appointment waiting time was a problem to the respondent. The 4 PREMs were used to create a composite *Problem Score*, which was the sum of the individual answer options of each PREM (1=yes, definitely; 2=yes, to some extent; 3=no, not really; and 4=definitely not). Hence, the composite score ranged from 4 to 16, where higher values represent more problems experienced during the visit. We also constructed a *Negative Experiences Score* (range 0-4) by counting the PREM items that did not receive a "yes, definitely" answer. Finally, we created the binary variable *Any Negative Experience*, which

indicated if the response to a PREM item was other than "yes, definitely."

Background Variables

We recorded respondents' sociodemographic variables, such as age, gender (female or male), education (primary, secondary, or tertiary), family status (married or not married), employment status (with a paid job or without a paid job including students, pensioners, unemployed, etc), and place of residence (capital, other cities, or village). Age groups were formed according to main Medical Subject Heading (MeSH) categories, adding 18-year olds to the young adult category (young adults: 18–24-year olds; adults: 25–44-year olds; middle aged: 45–64-year-olds; aged > 80: 65+ year-olds) [46]. Net monthly household income was queried in 11 range categories, and per-capita household income was calculated by dividing the category midrange values by the number of household members, without adjustment for the number of children. The midrange value of the upper open category was calculated by fitting the Pareto curve as proposed by Parker and Fenwick [47]. We generated income groups according to quintiles of per-capita monthly net household income, with lower limits of €203 (US $241), €285 (US $338), €365 (US $432), and €463 (US $549), respectively, for the second, third, fourth, and fifth quintiles calculated from the third to eighth national decile group means by linear interpolation [48]. We also recorded respondents' health status using the Minimal European Health Module (MEHM) [49,50]. The MEHM included an item on self-perceived health (very bad, bad, fair, good, or very good); an item on whether the respondent had long-standing health problems (1=chronic morbidity present); and the Global Activity Limitation Indicator, which assessed for activity limitations due to health problems (not limited at all, limited but not severely, or severely limited) [50].

Statistical Methods

Descriptive methods were used when analyzing the sociodemographic characteristics of the sample as well as the PREM items. To test the basic psychometric properties of the PREM scores constructed from multiple items, we assessed their distributional properties, performed exploratory factor analysis (EFA), and calculated internal consistency (Cronbach α) [51]. The normality assumption was tested via the Shapiro–Wilk test [52] and the Kaiser–Meyer–Olkin (KMO) test was used to check the suitability of data for EFA [53]. Pairwise biserial or polychoric correlations were calculated between PREM items and eHEALS scores. Polychoric correlation assumes bivariate normally distributed latent variables behind ordinal response items and provides the correlation coefficients for those latent variables using a maximum-likelihood estimation [54]. The bivariate associations between eHEALS quartiles and PREM scores as well as demographic variables were tested via ANOVA, the Kruskal–Wallis test, and the chi-square test of independence [55-57].

We performed multivariate regression analyses to explore the relationship between eHEALS quartiles and PREM items, as well as the composite PREM scores, after controlling for sociodemographic variables, respondents' health status (MEHM), the setting of the visit (GP, public specialist, or private specialist), and type of HCP (GP, specialist, or other allied health professional). The following models were conducted: (1) logistic regression for binary PREM items or constructed binary variables, (2) ordered logit models for polytomous PREM items, and (3) ordinary least squares (OLS) models for waiting times and composite PREM scores. We tested the joint significance of eHEALS quartiles as a single predictor variable using the Wald test. OLS models were tested for heteroskedasticity via the Breusch–Pagan test and for specification error via the Ramsey regression equation specification error test (RESET). We applied robust regression if heteroskedasticity was detected [58]. In case we detected model functional misspecification error, log-transformation or square-root transformation was performed on the dependent variable [59]. Goodness of fit of logistic and ordered logit models were tested, respectively, by the binary and ordinal versions of the Hosmer–Lemeshow test [60]. Unmet medical needs were also explored in an extended sample of those respondents who experienced health problems over the past 12 months, regardless of whether they had ambulatory consultation with an HCP. All calculations were performed using the Stata version 14.2 statistical software package (StataCorp) [61]. The level of significance was set at $P<.05$, and we applied no more than 15 observations per predictor variable when running multiple regression models [62]. Analyses were carried out without applying weights on the sample.

Results

Respondents' Characteristics

From the 1000 survey respondents, 736 had ambulatory HCP consultation within 12 months, out of which 5 happened over telephone. In 118 cases the respondent did not have a health problem, and 25 respondents could not tell if the visit happened at the regular HCP. After applying all criteria in sequence, 666 individuals were included in the sample (Table 1). Respondents with tertiary education and from the highest income quintile were slightly over-represented, whereas rural citizens were slightly under-represented compared with the general population. Mean age of our sample was 48.9 (SD 17.6) years. The demographic characteristics of the sample, all survey respondents, and the general population are summarized in Table 1. Responses on the income, chronic morbidity, and activity limitation items were provided by 86.5% (576/666), 89.0% (593/666), and 96.4% (642/666) of the respondents, respectively. The first, second, third, fourth, and fifth income quintiles of the sample corresponded to €115 (US $136), €247 (US $293), €332 (US $393), €397 (US $470), and €669 (US $786) mean per-capita household income levels, using the April 2020 12-month-average exchange rate of 330.73 HUF/€ (279.14 HUF/US $) [63]. The responses by PREM items are summarized in Table 2.

Table 1. Sample characteristics.

Characteristics	Sample (N=666), n (%)	Survey (N=1000), n (%)	General adult population [42], %
Sociodemographic			
Age group			
18-24	62 (9.3)	118 (11.8)	10.6
25-44	234 (35.1)	389 (38.9)	35.7
45-64	191 (28.7)	272 (27.2)	33.1
65+	179 (26.9)	221 (22.1)	20.6
Gender			
Female	364 (54.7)	550 (55.0)	53.4
Male	302 (45.4)	450 (45.0)	46.6
Education			
No primary school	__[a]	—	0.6
Primary	213 (31.9)	341 (34.1)	48.1
Secondary	244 (36.6)	363 (36.3)	33.5
Tertiary	209 (31.4)	296 (29.6)	17.8
Household income per capita			
First quintile	142 (21.3)	228 (22.8)	20.0
Second quintile	105 (15.8)	167 (16.7)	20.0
Third quintile	57 (8.6)	81 (8.1)	20.0
Fourth quintile	86 (12.9)	118 (11.8)	20.0
Fifth quintile	186 (27.9)	254 (25.4)	20.0
Missing[b]	90 (13.5)	152 (15.2)	—
Family status			
Married/domestic partnership	432 (64.9)	618 (61.8)	—
Single/divorced/widow	234 (35.1)	382 (38.2)	—
Employment status			
Paid job	319 (47.9)	500 (50.0)	48.3
Without paid job	347 (52.1)	500 (50.0)	51.7
Residence			
Budapest	146 (21.9)	213 (21.3)	17.4
City	371 (55.7)	557 (55.7)	52.1
Village	149 (22.4)	230 (23.0)	30.5
NUTS[c] 1 region			
Central Hungary	236 (35.4)	348 (34.8)	30.0
Transdanubia	237 (35.6)	299 (29.9)	30.4
Great Plain and North	193 (28.9)	353 (35.3)	39.6
MEHM[d]			
Self-perceived health			
Very bad	3 (0.5)	5 (0.5)	—
Bad	62 (9.3)	77 (7.7)	—
Fair	252 (37.8)	323 (32.3)	—
Good	293 (43.9)	471 (47.1)	—

Characteristics	Sample (N=666), n (%)	Survey (N=1000), n (%)	General adult population [42], %
Very good	56 (8.4)	124 (12.4)	—
Chronic morbidity			
No	200 (30.0)	390 (39.0)	—
Yes	393 (59.0)	489 (48.9)	—
Missing	73 (10.9)	121 (12.1)	—
Activity limitations			
Not limited at all	342 (51.4)	579 (57.9)	—
Limited but not severely	254 (38.1)	313 (31.3)	—
Severely limited	46 (6.9)	56 (5.6)	—
Missing	24 (3.6)	52 (5.2)	—
Inclusion criteria			
Ambulatory HCP[e] visit in past 12 months			
No/not face-to-face/missing	0 (0.0)	269 (26.9)	—
Yes, but not for own health problem	0 (0.0)	52 (5.2)	—
Yes, at regular HCP	546 (81.9)	546 (54.6)	—
Yes, but not at regular HCP	120 (18.0)	120 (12.0)	—
Yes, missing if regular HCP	0 (0.0)	13 (1.3)	—

[a]Not available.

[b]Missing: missing responses/do not know/do not want to answer.

[c]NUTS: Nomenclature of Territorial Units for Statistics.

[d]MEHM: Minimal European Health Module.

[e]HCP: health care professional.

Table 2. Patient responses by PREM[a] items (N=666).

Patient response	n (%)
Access to care: last visit	
Health care setting	
GP[b]	278 (41.7)
Public specialist	316 (47.4)
Private specialist	72 (10.8)
Type of HCP[c]	
GP	278 (41.7)
Specialist	360 (54.1)
Nurse/other HCP	28 (4.2)
Time of last visit	
In the last 30 days	277 (41.6)
Between 1 and 3 months ago	180 (27.0)
Between 3 and 6 months ago	95 (14.3)
Between 6 and 12 months ago	114 (17.1)
Access to care: unmet medical needs	
Missed visit due to travel burden	
No	506 (76.0)
Yes	147 (22.1)
Missing[d]	13 (2.0)
Missed visit due to cost burden	
No	534 (80.2)
Yes	120 (18.0)
Missing	12 (1.8)
Missed intervention due to cost burden	
No	559 (83.9)
Yes	99 (14.9)
Missing	8 (1.2)
Missed medication due to cost burden	
No	508 (76.3)
Yes	148 (22.2)
Missing	10 (1.5)
Access to care: waiting times	
Problem with waiting to be seen on the day of consultation	
No	487 (73.1)
Yes	179 (26.9)
Problem with waiting for appointment	
No	564 (84.7)
Yes	102 (15.3)
Patient experiences	
Doctor spending enough time with patient in consultation	
Yes, definitely	427 (64.1)

Patient response	n (%)
Yes, to some extent	160 (24.0)
No, not really	57 (8.6)
Definitely not	17 (2.6)
Missing	5 (0.8)
Doctor providing easy to understand explanations	
Yes, definitely	459 (68.9)
Yes, to some extent	166 (24.9)
No, not really	27 (4.1)
Definitely not	12 (1.8)
Missing	2 (0.3)
Doctor giving opportunity to ask questions or raise concerns	
Yes, definitely	414 (62.2)
Yes, to some extent	164 (24.6)
No, not really	63 (9.5)
Definitely not	15 (2.3)
Missing	10 (1.5)
Doctor involving patient in decisions about care and treatment	
Yes, definitely	338 (50.8)
Yes, to some extent	195 (29.3)
No, not really	77 (11.6)
Definitely not	19 (2.9)
Missing	37 (5.6)
Overall quality of the visit	
Poor	19 (2.9)
Fair	60 (9.0)
Good	186 (27.9)
Very good	205 (30.8)
Excellent	193 (29.0)
Missing	3 (0.5)

[a]PREM: OECD-proposed set of questions on Patients' Experiences with Ambulatory Care.

[b]GP: general practitioner.

[c]HCP: health care professional.

[d]Missing: missing responses/do not know/do not want to answer.

eHEALS

Mean eHEALS score was 29.3 (SD 4.9). eHEALS quartile mean scores were as follows: first quartile 23.5 (range 12-26; 191/666, 28.7%), second quartile 28.2 (range 27-29; 151/666, 22.7%), third quartile 31.2 (range 30-32; 182/666, 27.3%), and fourth quartile 36.0 (range 33-40; 142/666, 21.3%). Mean age of individuals in the fourth eHEALS quartile was 44.5 years (SD 17.1), which was lower than that of individuals in the first (49.9 years [SD 17.4]), second (51.2 [SD 17.5]), and third quartiles (49.3 [SD 17.7; $F_{3,662}$=4.07, P=.007). Mean eHEALS scores did not differ between male and female respondents (t_{664}=1.27, P=.21). However, while the percentage of female respondents decreased evenly from the first (101/364, 27.8%), second (92/356, 25.8%), third (86/364, 23.6%), and fourth eHEALS quartiles (85/364, 23.4%), male respondents were concentrated in the first (90/302, 29.8%) and third (96/302, 31.8%) quartiles (χ^2_3=8.2, P=.04). The difference in terms of education (χ^2_6=5.6, P=.47) and income (χ^2_{12}=7.9.6, P=.79) was not significant between the four eHEALS groups.

PREM Unmet Medical Needs

A majority of respondents (380/631, 60.2%) did not report unmet medical needs in any areas. One unmet need was reported by 18.5% (117/631), 2 unmet needs by 8.9% (56/631), 3 unmet needs by 7.4% (47/631), and 4 unmet needs by 4.9% (31/631)

of respondents. The Unmet Medical Needs Score had a single-factor structure with a KMO value of 0.73, suggesting moderately adequate sampling for EFA. The Cronbach α of .73 suggested acceptable internal consistency of this score constructed by adding the PREM items of the Unmet Medical Needs section.

PREM Waiting Times

Mean office waiting times were 63.3 (SD 71.0) minutes; 23.0% (152/661) of respondents waited less than 15 minutes, while waiting time was longer than 2 hours for 14.2% (94/661) of the sample. Long office waiting time was a problem for 26.9% (179/666) of all respondents, and for 34.8% (179/514) of those who waited longer than 15 minutes. Mean appointment waiting time was 16.8 (SD 27.8) days. Whereas 37.6% (242/643) of the sample could get an appointment on the same day, 18.2% (117/643) of respondents had to wait longer than 30 days. Long appointment waiting time was a problem for 15.3% (102/666) of all respondents, and for 24.1% (102/424) of those who did not get appointment on the same day. Any waiting problem either at the HCP office or before getting an appointment was reported by 33.5% (223/666) of the sample.

PREM Patient Experiences

The Problem Score showed strong right skew (mean 2.0, SD 2.5; median 1; kurtosis 4.5; skewness 1.4; Shapiro–Wilk test $P<.001$). EFA suggested a single-factor structure with adequate sampling (KMO statistic=0.82) and good internal consistency (Cronbach α=.87). Whereas 0.5% (3/623) of respondents indicated the worst experience in all domains (*definitely not* answers for all 4 items; score 16), the experience was flawless (*definitely yes* answers for all 4 items; score 4) for 40.9% (255/623). The Negative Experiences Score had bimodal distribution (mean 1.5, SD 1.6; kurtosis 1.7; skewness 0.5; Shapiro–Wilk test $P<.001$) and a single-factor structure with adequate sampling (KMO statistic=.80) and good internal consistency (Cronbach α=.83). Problems were reported in 0, 1, 2, 3, and 4 domains by 40.9% (255/623), 18.0% (112/623), 11.2% (70/623), 11.4% (71/623), and 18.5% (115/623) of respondents, respectively. The strong correlation between the Problem Score and Negative Experiences Score (polyserial ρ=0.95) suggested that counting the domains with answers other than *definitely yes* accounted for most of the information within the patient experiences section.

Correlation Between PREM and eHEALS Scores

The polychoric correlation matrix of PREM items and eHEALS score is shown in Table 3. The patient experience measures were strongly intercorrelated, whereas the correlation between those PREMs and waiting times and unmet medical need measures were moderate or weak. The corresponding waiting times and waiting problems were strongly correlated. The overall quality of the visit showed a strong negative correlation with items of the patient experiences section of the survey; the correlation was moderate or weak with remaining items. The eHEALS score showed a weak negative correlation with any of the PREMs. The correlation between PREM items and the time of the last visit was minimal.

Table 3. Correlation matrix of PREM[a] items.[b]

Variable	Patient experiences					Access to care							
						Waiting times				Unmet medical needs			
	Time	Understand	Questions	Decisions	Overall quality	oWT	oWP	aWT	aWP	Travel	Visit	Intervention	Medication
Time[c]	1.00												
Understand[d]	0.75	1.00											
Questions[e]	0.77	0.77	1.00										
Decisions[f]	0.71	0.76	0.83	1.00									
Overall quality[g]	–0.79	–0.75	–0.78	–0.74	1.00								
oWT[h]	0.35	0.30	0.34	0.24	–0.36	1.00							
oWP[i]	0.42	0.43	0.38	0.38	–0.45	0.67	1.00						
aWT[j]	0.17	0.16	0.11	0.14	–0.12	0.11	0.15	1.00					
aWP[k]	0.42	0.32	0.38	0.35	–0.37	0.32	0.50	0.67	1.00				
Travel[l]	0.20	0.21	0.21	0.24	–0.16	0.17	0.33	0.05	0.38	1.00			
Visit[m]	0.36	0.30	0.36	0.27	–0.32	0.14	0.40	0.10	0.36	0.61	1.00		
Intervention[n]	0.17	0.16	0.20	0.19	–0.18	0.12	0.37	0.18	0.44	0.61	0.89	1.00	
Medication[o]	0.22	0.23	0.26	0.16	–0.17	0.16	0.33	0.15	0.43	0.47	0.66	0.67	1.00
eHEALS[p]	–0.03	–0.13	–0.04	–0.04	0.11	0.01	0.14	–0.04	–0.02	–0.02	0.00	0.02	–0.06
Last visit[q]	0.03	-0.01	0.00	0.01	–0.06	0.00	0.06	–0.10	–0.08	0.00	–0.07	–0.10	–0.04

[a]PREM: OECD (Organisation for Economic Co-operation and Development)-proposed set of questions on Patients' Experiences with Ambulatory Care.

[b]Pairwise tetrachoric correlations for binary item pairs, polychoric correlations for polytomous items, polyserial and biserial correlations between eHEALS scores and polytomous and binary items, respectively.

[c]Doctor spending enough time with patient in consultation (4-point Likert scale; higher points indicate more problems).

[d]Doctor providing easy-to-understand explanations (4-point Likert scale; higher points indicate more problems).

[e]Doctor giving opportunity to ask questions or raise concerns (4-point Likert scale; higher points indicate more problems).

[f]Doctor involving patient in decisions about care and treatment (4-point Likert scale; higher points indicate more problems).

[g]Overall quality of last appointment (5-point Likert scale; higher points indicate better experience).

[h]Waiting time to be seen on the day of consultation (office waiting time [oWT]).

[i]Problem with waiting to be seen on the day of consultation (office waiting was a problem [oWP]).

[j]Waiting time to get the appointment (appointment waiting time [aWT]).

[k]Problem with waiting for appointment: yes (appointment waiting time was a problem [aWP]).

[l]Missed visit due to travel burden.

[m]Missed visit due to cost burden.

[n]Missed intervention due to cost burden.

[o]Missed medication due to cost burden.

[p]eHEALS: eHealth Literacy Scale.

[q]Time of last visit: 4 categories; higher points indicate more time elapsed since last visit.

Bivariate Association Between PREM and eHEALS Quartiles

The association between eHEALS quartiles and the problem score was not significant (Kruskal–Wallis test with ties, χ^2_3=4.9, *P*=.18; Figure 1); conversely, the association was significant (χ^2_{12}=24.4, *P*=.01) when the negative experiences score was considered (Figure 2). While we found no relationship between eHEALS quartiles and the time spent with the patient (χ^2_9=14.5, *P*=.11) and opportunity to ask questions (χ^2_9=9.2, *P*=.42), the association was significant (χ^2_9=24.2, *P*=.002) between how easy it was to understand the HCP's explanations (Figure 3) and the extent to which the HCP involved the respondent in decisions (χ^2_9=18.2, *P*=.03; Figure 4). The association between

eHEALS quartiles and the overall quality score was significant (Kruskal–Wallis test with ties χ^2_3=10.1, *P*=.02; Figure 5). By contrast, the association between eHEALS quartiles and the share of respondents by overall quality categories was not significant (χ^2_{12}=20.6, *P*=.07; Figure 6). Although the differences were small, the results suggest that respondents in the lowest eHEALS quartile had the least positive experience with HCP communication. Positive experiences were most frequently reported in the third and fourth eHEALS quartiles, whereas the subgroup with the highest eHEALS scores reported somewhat more negative experiences, when compared with respondents in the third eHEALS quartile. We found no association between eHEALS quartiles and either unmet medical needs or waiting times.

Figure 1. Problem score by eHEALS (eHealth Literacy Scale) quartiles.

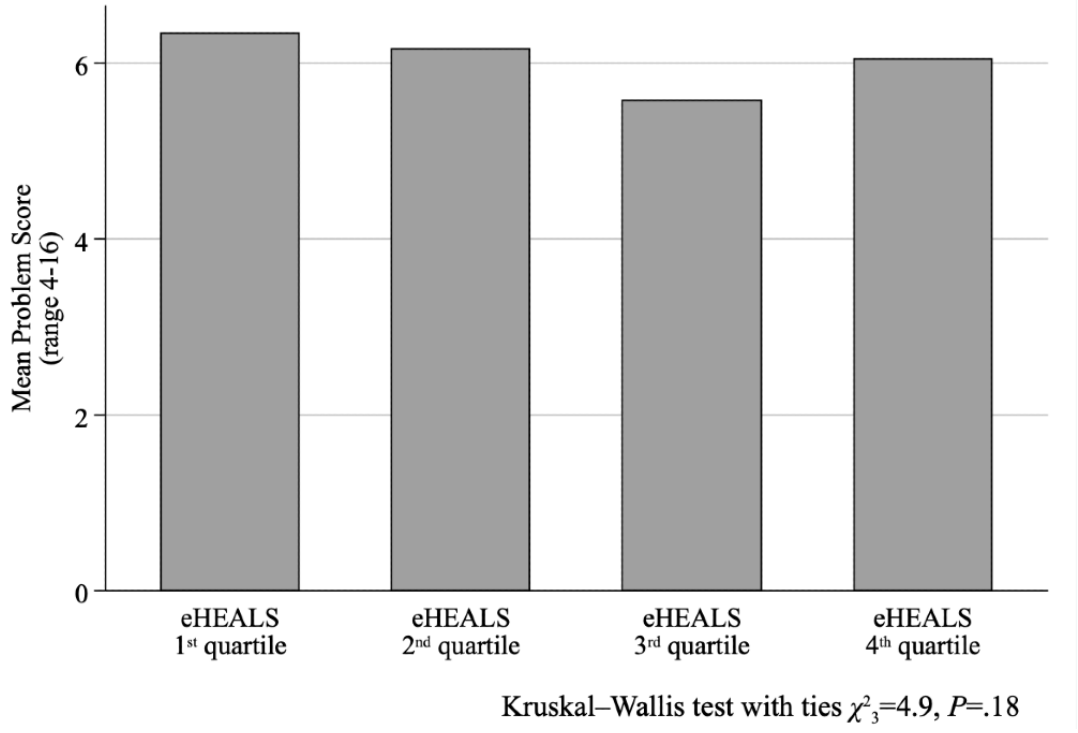


Figure 2. Negative Experiences Score by eHEALS (eHealth Literacy Scale) quartiles.

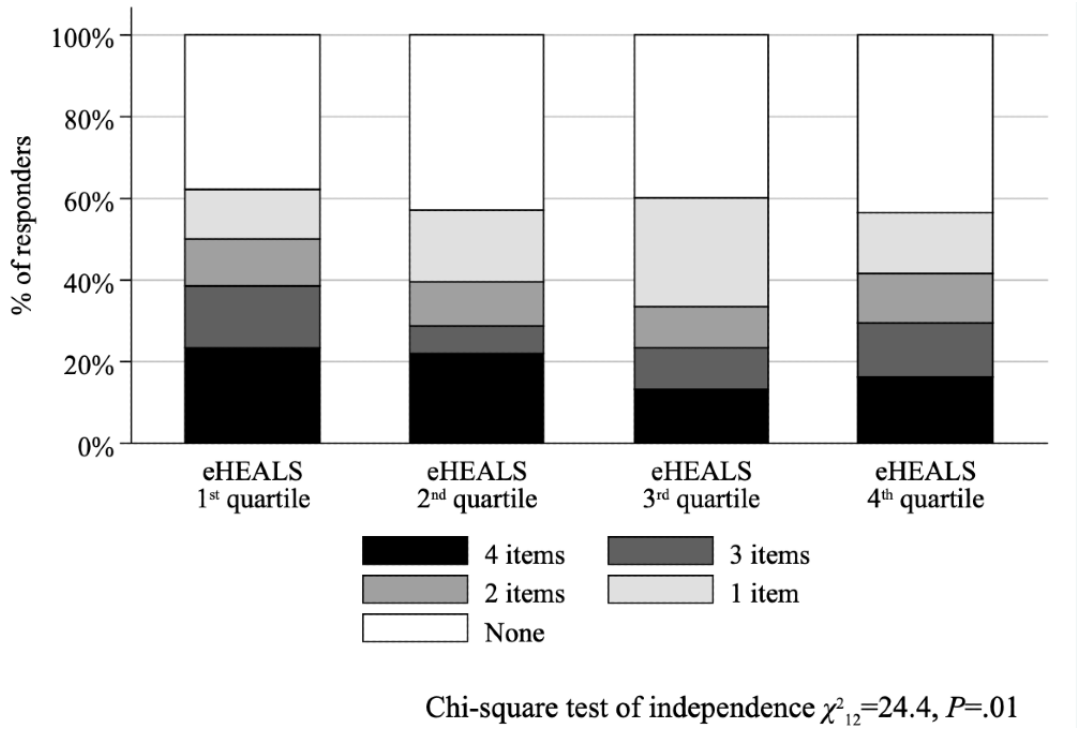


Figure 3. Perceived easiness of understanding the explanations of the health care professional by eHEALS (eHealth Literacy Scale) quartiles.

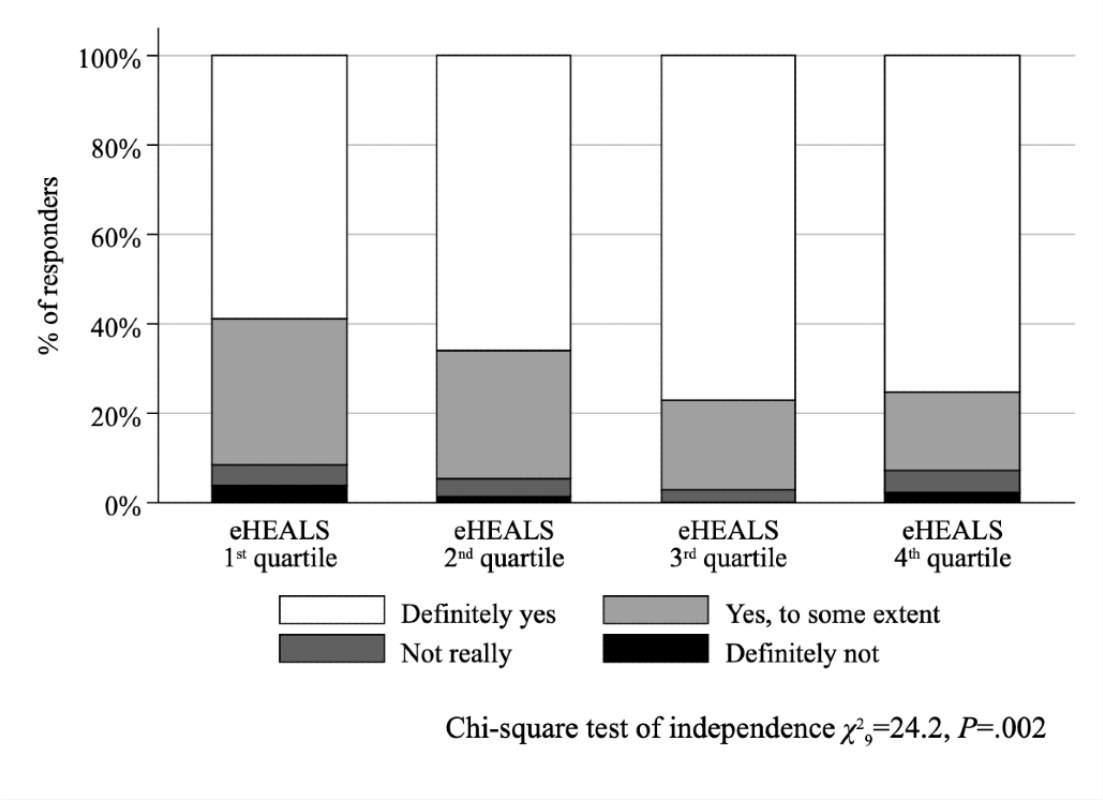


Figure 4. Perceived involvement of the respondent by the health care professional in decisions about care and treatment by eHEALS (eHealth Literacy Scale) quartiles.

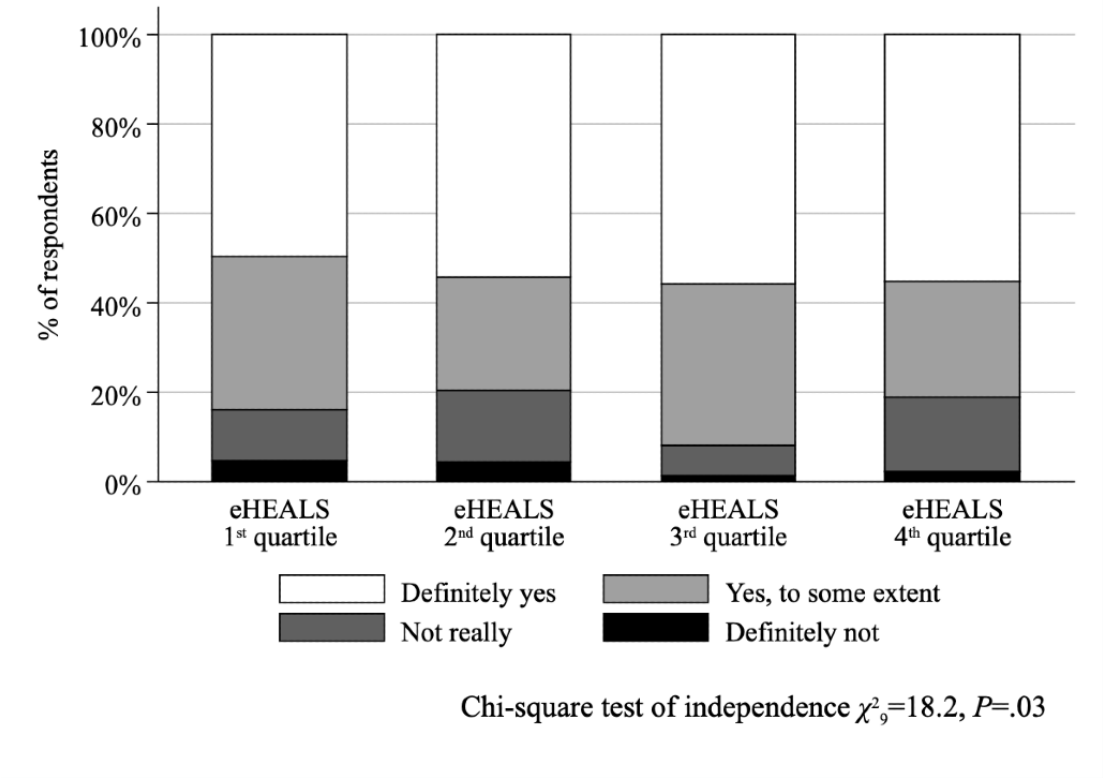


Figure 5. Mean overall quality of the last visit by eHEALS (eHealth Literacy Scale) quartiles.

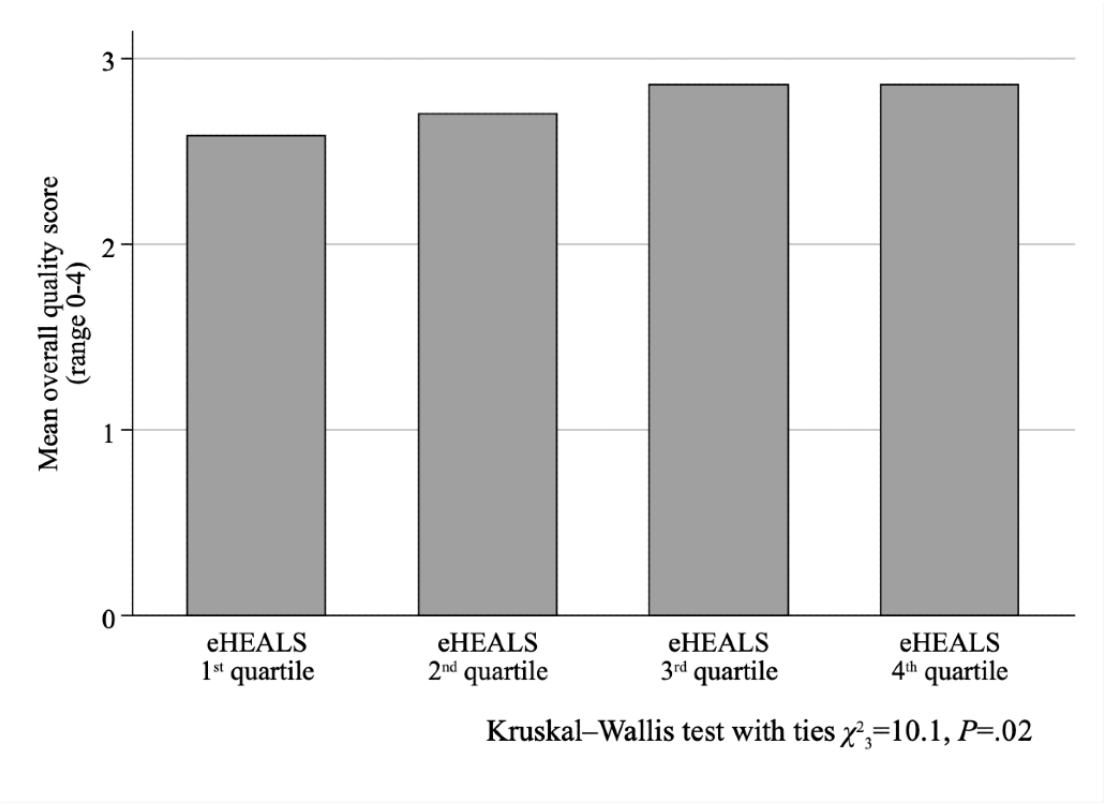
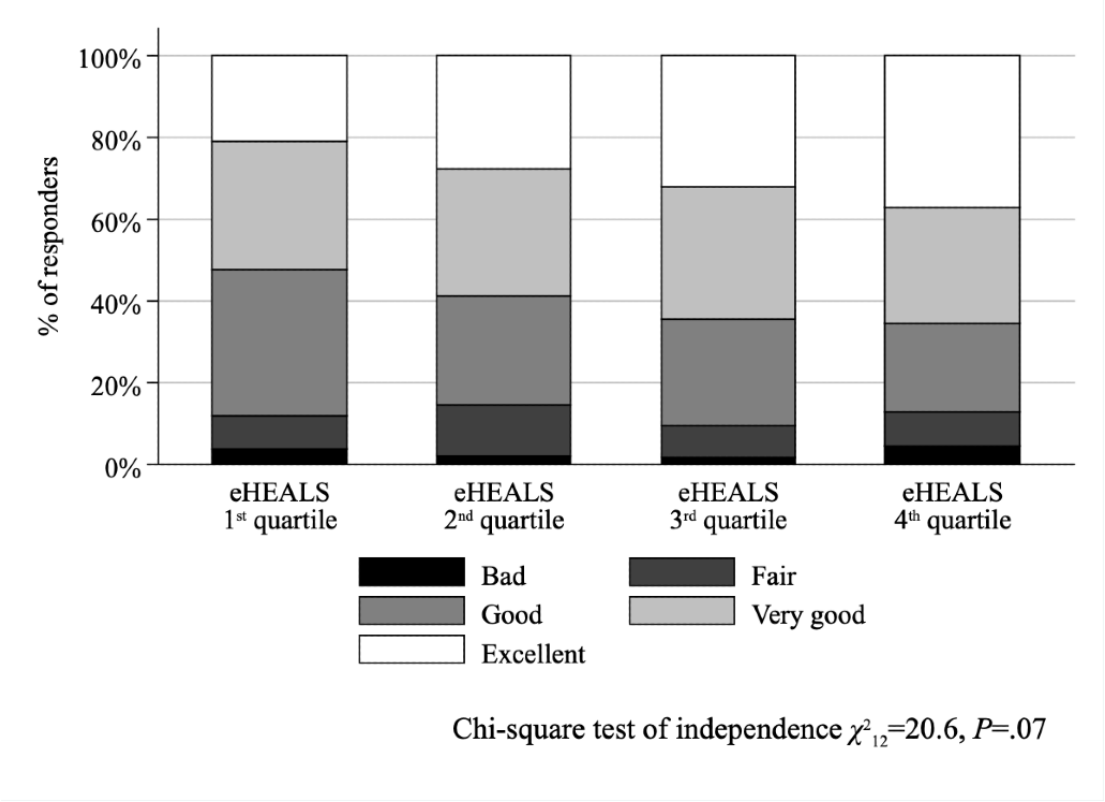


Figure 6. Overall quality categories of the last visit by eHEALS (eHealth Literacy Scale) quartiles.



Regression Analyses of Individual PREM Items

After controlling for sociodemographic variables, respondents' health status, the setting of the visit, and type of HCP in ordered logit models (Table 4), the association was significant between eHEALS quartiles and the "Easy to understand explanations" item (Wald test χ^2_3=11.8, *P*=.008). Although eHEALS quartiles jointly were not significant predictors of the "Opportunity to ask questions" item (Wald test χ^2_3=4.7, *P*=.19), the differences

between the first and third eHEALS quartiles were significant in both of those items ($P<.001$ and $P=.04$, respectively). The association was not significant between eHEALS quartiles and either the time spent with the patient (Wald test $\chi^2_3=1.8$, $P=.61$) or the involvement of the patient in decisions about care and treatment (Wald test $\chi^2_3=3.4$, $P=.33$).

After controlling for covariates, the overall quality also differed between respondents in the first and third eHEALS quartiles (Table 5), although the joint effect of eHEALS quartiles on the overall quality was not significant (Wald test $\chi^2_3=6.0$, $P=.11$). Unmet medical needs (Multimedia Appendix 3) and waiting times (Multimedia Appendix 4) were not associated with the eHEALS quartiles. The sensitivity analysis has shown similar findings in the majority of models using alternative eHEALS group boundaries (Multimedia Appendix 6). In half of the alternative scenarios the probability of missed interventions was also significantly lower ($P=.05$ to $P=.02$) in the third (moderately high) than in the first (lowest) and eHEALS score group, but not in the predefined quartiles ($P=.05$). We did not find significant ($P=.16$ to $P=.88$) association between eHEALS quartiles and any unmet needs variables in the extended sample involving those respondents who had a health problem over the past 12 months, regardless their participation in ambulatory HCP consultation (Multimedia Appendix 7). The detailed analysis of the effect of covariates on the PREM modules was out of scope for this paper, and has been provided elsewhere [30,41,64].

Table 4. Ordered logit regression of the patient experience PREM[a] items.

Variables	Time[b]		Understand[c]		Questions[d]		Decisions[e]	
	β	*P* value	β	*P* value	β	*P* value	β	*P* value
eHealth Literacy Scale[f]								
Second quartile	–0.30	.28	–0.29	.31	–0.42	.12	–0.20	.47
Third quartile	–0.31	.24	–0.98	<.001	–0.54	.04	–0.30	.23
Fourth quartile	–0.14	.61	–0.51	.09	–0.26	.34	0.15	.57
Age group[g]								
25-44 years old	–0.87	.02	–0.78	.04	–0.81	.03	–0.73	.05
45-64 years old	–1.06	.009	–1.45	<.001	–1.38	<.001	–0.88	.03
65+ years old	–1.38	.002	–1.81	<.001	–1.58	<.001	–1.42	<.001
Education[h]								
Secondary	–0.27	.31	0.10	.71	–0.03	.90	–0.01	.98
Tertiary	0.25	.37	0.22	.47	0.14	.62	0.08	.78
Gender								
Male	–0.37	.09	0.12	.59	–0.20	.36	–0.08	.68
Income[i]								
Second quintile	–0.01	.97	–0.31	.36	0.32	.31	0.40	.19
Third quintile	0.22	.57	–0.17	.67	0.50	.17	0.53	.14
Fourth quintile	0.25	.47	–0.22	.54	0.24	.49	0.38	.25
Fifth quintile	0.08	.8	0.07	.83	0.30	.32	0.49	.10
Paid employment								
Yes	–0.34	.19	–0.14	.60	–0.13	.59	–0.14	.58
Family status								
Married/domestic partnership	–0.27	.21	–0.22	.32	0.01	.97	–0.12	.55
Residence[j]								
City	–0.07	.76	–0.20	.44	0.10	.67	0.21	.37
Village	–0.70	.03	–0.34	.31	–0.29	.34	–0.59	.06
Self-perceived health[k]								
Very bad	–13.30	.99	–12.77	.99	0.13	.92	0.03	.98
Bad	1.20	.03	1.67	.02	0.72	.19	1.46	.01
Fair	0.52	.26	1.80	.003	0.41	.35	0.96	.04
Good	0.35	.41	1.35	.02	0.15	.71	0.63	.14
Global Activity Limitation Indicator[l]								
Limited but not severely	0.26	.27	0.15	.54	0.11	.62	0.22	.34
Severely limited	0.19	.66	0.12	.80	0.33	.44	0.23	.56
Chronic morbidity								
Yes	0.18	.49	0.20	.48	0.29	.27	0.39	.12
Setting[m]								
Public specialist	1.23	.009	1.23	.01	__[n]	.99	0.37	.46
Private specialist	0.90	.11	1.00	.08	–0.37	.53	0.34	.54
HCP type[o,p]								

Variables		Time[b]		Understand[c]		Questions[d]		Decisions[e]	
		β	*P* value	β	*P* value	β	*P* value	β	*P* value
	Specialist	–1.39	.003	–1.32	.005	–0.41	.42	–0.83	.09
	Nurse/other HCP	—	—	—	—	—	—	—	—
Regular HCP									
	Yes	–0.29	.28	0.05	.87	–0.35	.18	–0.27	.31
N		502		504		500		477	
LR[q] test χ^2_{28}		52.7	.003	60.6	<.001	43.7	.03	50.7	.005
GOF[r] test χ^2_{26}		18.5	.86	13.5	.98	21.7	.71	24.9	.52

[a]PREM: OECD (Organisation for Economic Co-operation and Development)-proposed set of questions on Patients' Experiences with Ambulatory Care.

[b]Doctor spending enough time with patient in consultation (4-point Likert scale).

[c]Doctor providing easy to understand explanations (4-point Likert scale).

[d]Doctor giving opportunity to ask questions or raise concerns (4-point Likert scale).

[e]Doctor involving patient in decisions about care and treatment (4-point Likert scale).

[f]Base: first quartile.

[g]Base: 18-24 years old.

[h]Base: primary.

[i]Base: first quintile.

[j]Base: capital.

[k]Base: very good.

[l]Base: not limited.

[m]Base: general practitioner.

[n]Not available.

[o]Base: general practitioner.

[p]HCP: health care professional.

[q]Likelihood ratio; omnibus test for independence, current model versus null model.

[r]Goodness of fit; ordinal version of the Hosmer–Lemeshow test.

Table 5. Multivariate regression of PREM[a] scores.

Model	Overall quality		Log-problem score		Negative experience score		Any negative experience	
	Ordered logit		Robust[b]		Robust		Logistic	
	β	*P* value	β	*P* value	β	*P* value	β	*P* value
eHEALS[c,d]								
Second quartile	0.24	.31	–0.06	.23	–0.37	.08	–0.25	.38
Third quartile	0.55	.02	–0.10	.02	–0.46	.02	–0.16	.54
Fourth quartile	0.34	.16	–0.02	.74	–0.17	.40	–0.17	.56
Age group[e]								
25–44 years old	0.56	.09	–0.15	.03	–0.46	.08	–0.64	.14
45–64 years old	0.71	.04	–0.22	.002	–0.83	.003	–1.15	.01
65+ years old	1.12	.003	–0.29	<.001	–1.16	<.001	–1.60	.001
Education[f]								
Secondary	–0.12	.60	__[g]	.97	–0.01	.96	–0.04	.89
Tertiary	–0.39	.10	0.04	.36	0.18	.37	–0.03	.92
Gender								
Male	0.07	.69	–0.03	.42	–0.03	.86	0.21	.32
Income[h]								
Second quintile	–0.10	.70	0.04	.35	0.22	.29	0.71	.03
Third quintile	0.17	.59	0.05	.40	0.24	.39	0.78	.04
Fourth quintile	–0.01	.97	0.02	.69	0.20	.41	0.44	.20
Fifth quintile	0.06	.81	0.05	.33	0.27	.21	0.61	.047
Paid employment								
Yes	0.12	.59	–0.04	.31	–0.08	.64	0.02	.93
Family status								
Married/domestic partnership	0.25	.17	–0.03	.45	–0.09	.55	0.12	.59
Residence[i]								
City	–0.04	.85	—	.99	0.03	.85	0.10	.70
Village	0.26	.34	–0.11	.03	–0.49	.02	–0.66	.03
Self-perceived health[j]								
Very bad	0.21	.88	–0.05	.72	0.04	.95	0.01	.99
Bad	–0.97	.047	0.24	.007	1.24	<.001	1.32	.02
Fair	–1.00	.01	0.15	.02	0.80	.003	0.69	.11
Good	–0.85	.02	0.10	.09	0.59	.01	0.49	.21
Activity limitations[k]								
Limited but not severely	–0.27	.18	0.04	.28	0.17	.34	0.34	.14
Severely limited	–0.23	.55	0.05	.48	0.20	.52	–0.08	.85
Chronic morbidity								
Yes	0.04	.85	0.04	.32	0.15	.41	0.03	.91
Setting[l]								
Public specialist	–0.60	.16	0.18	.09	0.49	.22	0.95	.12
Private specialist	–0.05	.91	0.13	.24	0.22	.62	0.29	.66

Model	Overall quality		Log-problem score		Negative experience score		Any negative experience	
	Ordered logit		Robust[b]		Robust		Logistic	
	β	*P* value	β	*P* value	β	*P* value	β	*P* value
HCP type[m,n]								
Specialist	0.72	.09	–0.24	.02	–0.75	.05	–1.19	.046
Nurse/other HCP	—	—	—	—	—	—	—	—
Regular HCP								
Yes	0.03	.91	–0.05	.24	–0.25	.22	–0.36	.21
Constant	—	—	1.87	<.001	1.82	<.001	0.81	.25
N	503		473		473		505	
LR[o] test χ^2_{28}	42.1	.04					49.6	.007
LR test $F_{28,444}$			2.63	<.001	2.27	<.001		
R^2			0.13		0.13			
GOF[p] test χ^2_{35}	30.5	.68						
GOF test χ^2_{470}							503.3	.14
Ramsey RESET[q] $F_{3,434}$			2.37	.07	0.07	.98		

[a]PREM: OECD-proposed set of questions on Patients' Experiences with Ambulatory Care.

[b]Ordinary least squares (OLS) regression with robust standard errors.

[c]Base: first quartile.

[d]eHEALS: eHealth Literacy Scale.

[e]Base: 18-24 years old.

[f]Base: Primary.

[g]Not available.

[h]Base: first quintile.

[i]Base: Capital.

[j]Base: Very good.

[k]Base: Not limited.

[l]Base: General practitioner.

[m]Base: General practitioner.

[n]HCP: health care professional.

[o]Likelihood ratio; omnibus test for independence, current model versus null model.

[p]Goodness of fit; Hosmer–Lemeshow test.

[q]Regression equation specification error test.

Regression Analyses of Composite PREM Scores

The specification of robust linear regression models was acceptable for the log-problem score and the negative experience score (Table 5). Findings show that the difference was significant between the first and third eHEALS quartiles in the log-problem score (*P*=.02) and negative experience score models (*P*=.02). The joint Wald test of eHEALS quartiles was not significant in either model (log-problem score $F_{3,430}$=2.28, *P*=.08; negative experience score $F_{3,430}$=2.17, *P*=.09; any negative experience χ^2_3=0.8, *P*=.84). In addition, logistic regression models for *any unmet medical needs* and *any waiting problems* had an acceptable fit (Multimedia Appendix 5); eHEALS was not a significant predictor in any of these models (*P*=.05 to *P*=.42). In several scenarios of the sensitivity analysis, the *unmet medical needs score* and *any unmet medical needs* suggested less unmet needs in the third (moderately high) than in the first (lowest) eHEALS score groups (Multimedia Appendix 6).

Discussion

To our knowledge, this is the first study that explores the relationship between eHealth literacy and PREMs with outpatient care. Our findings show a weak concave relationship between eHEALS scores and PREMs. We observed significant differences between respondents with lowest self-reported eHealth literacy levels (first eHEALS quartile) and the ones with moderately high levels (third eHEALS quartile) in terms of how easy it was to understand the explanations of the HCP, having the opportunity to ask questions, the number of items

where respondents experienced problems, and the overall quality of the last visit. Sensitivity analysis using alternative boundaries between eHEALS groups confirmed these findings in multiple alternative scenarios. Although the bivariate association between eHealth literacy and the involvement of respondents in decision making was significant, after controlling for covariates in multiple regression analyses, respondents' perception of spending enough time in the consultation and involvement in decision making did not show a statistically significant relationship with the eHEALS scores. Besides, our findings show no significant association between eHealth literacy and unmet medical needs and waiting times.

Although our literature search did not reveal papers reporting the association between PREM and eHealth literacy, several studies explored the effect of eHealth literacy (measured with the eHEALS instrument) on aspects of people-centered care such as the patient–physician relationship. A study among Iranian patients with multiple myeloma found a positive relationship between eHealth literacy and shared decision making, where eHealth literacy had a direct positive influence on shared decision making and an indirect positive effect mediated by collaborative patient communication patterns and trust in the health care system [37]. In a large survey among the Israeli general population, higher eHealth literacy score was associated with a more extensive interaction and a more balanced power position vis-à-vis with the treating physician [22]. While eHealth literacy had a direct positive influence on productive functional behaviors in all domains of patient empowerment among members of a Slovenian online health community, it also had a moderating effect on both dysfunctional and functional behaviors [65]. Using the Health Literacy Questionnaire, a survey on a large Dutch online panel of health care users demonstrated positive relationship between information appraisal, a higher-order health literacy skill, and shared decision making [66,67].

Our results show a strong negative correlation between the overall quality of the visit and the perceived problems with HCP communication including the involvement of the respondents in decision making. However, the relationship between overall patient-reported experience and eHealth literacy was not linear. The slightly increased probability of negative patient experiences among respondents with highest eHEALS scores is in line with the findings of a large international qualitative study among online health information users, where participants frequently reported reluctance to discuss the online content due to the expected negative reception from their HCPs [68]. On the same note, a systematic review on the impact of online health information on patient–physician relationship identified a positive effect on the majority of the cases, although several studies reported negative feelings concerning the discussion of online information with HCPs [33]. In the 2007 Health Information National Trends Survey, patients' concerns about the quality of online health information increased the likelihood of discussing it with their HCPs, while they were also more likely to experience negative reactions from the HCPs concerning the shared information [36].

Recognizing the multidimensional determinants of the patient–physician interaction, a recent line of research aimed to establish patient profiles characterized by various skill levels and attitudes, including eHealth literacy [32,65,69]. In a large multicountry survey, 4 distinct patient decision styles were described. While patients with a passive decision-making style had the lowest eHealth literacy skills, the autonomous-collaborator group showed somewhat higher eHealth literacy and worse patient–physician communication, compared with that of the collaborators, who were most likely to engage in shared decision making [32].

Among several potential contributing factors, the emergence of negative experiences among patients with greater eHealth literacy levels may partly be explained by the properties of the eHEALS instrument. Showing low correlation with objective measures of eHealth literacy, eHEALS has been described as a tool measuring rather self-efficacy related to eHealth literacy than actual skills [70]. Patients with low functional health literacy presenting high eHEALS scores tended to rely on non-established criteria when evaluating online health information compared to ones with high functional health literacy, who relied on more established criteria [71]. Overconfident use of low-quality health information due to ignorance about the actual low skill level [72] combined with high psychological empowerment may lead to dangerous self-management [73], evoking negative reactions from HCPs. Therefore, since the original definition coined by Norman and Skinner in 2006 [14], efficient communication skills or a supportive patient–HCP relationship has been included in several updated concepts of eHealth literacy [74] and eHealth readiness [69].

We also assume that access to high-quality online information including international best-in-class services may raise the expectations of people that may contrast their real-world experiences with the Hungarian health system, which operates at a lower efficiency and expenditure levels compared with other high-income societies [7]. Besides, GP gatekeeping systems, such as the Hungarian one, have been designed to restrict the demand side of health care, and are perceived as being less patient centric than non-gatekeeping systems [29]. It has been shown that patients that face barriers to access to care are usually more prone to health information–seeking behaviors [75]. Furthermore, dissatisfaction with the patient-centeredness of physicians and high eHealth literacy were among the key reasons of postvisit online information seeking in a US online health community [76].

Although the relationship between eHealth literacy and unmet medical needs or waiting times was not significant in our sample, we found the highest eHEALS scores among respondents with worst self-reported health [39]. It has been demonstrated that chronic patients develop health literacy skills over time [77], and higher health literacy was associated with better outcomes even in difficult-to-treat patients [78]. By contrast, multimorbid patients often experience issues such as insufficient coordination of care, access barriers, poor professional communication, and the lack of involvement in decisions [79]. Our results suggest that in addition to being a resource for positive experiences, high eHealth literacy may develop as a response to mitigate negative experiences with

care or unfavorable health outcomes. However, these links are yet to be elucidated.

Our study was conducted in the general population without focusing on any particular disease area. Most of the respondents reported on the last ambulatory visit at their usual HCP, therefore our results reflect the general experiences of individuals with outpatient care, regardless of the nature, number, or severity of their health conditions. We applied Hungarian versions of validated instruments that have been used widely in multiple countries, such as the eHEALS or the OECD's PREM questionnaire. We demonstrated that the composite PREM scores used in our analyses had adequate psychometric properties. However, caution is needed when generalizing our findings beyond the Hungarian setting, due to the differences of health systems, communication culture, or economic status of countries.

Furthermore, a number of limitations of our study have to be highlighted. First, only a small part of the variance of PREM items was explained by our OLS models, suggesting that potentially important determinants of patient experiences remained unexplored in our study. Moreover, eHEALS quartiles were jointly significant predictors only in case of a single PREM item, whereas—despite significant differences between the first and third quartiles—the joint test of eHEALS was not significant in 2 items. Applying refined analytical methods on a larger sample may explain patient-, physician-, and system-related factors that shape patients' experiences of care and also clarify the relationship between unmet medical needs and eHealth literacy, which yielded mixed results in our study. A further limitation of our study is the wide recall period spanning up to 1 year between the survey and the last patient visit. Recall bias has been reported in connection with patient-experience surveys, raising concerns about the comparability between data collected with different recall periods [80]. Although recall bias of responses cannot be excluded in our study, the correlation between the time of last visit and PREM responses was minimal, suggesting negligible influence of recall bias on our results.

The potential of eHealth to improve the efficiency of health systems has been recognized by policymakers. Low health literacy is a barrier to efficient implementation of eHealth interventions [81]. eHealth literacy is viewed as resource for patients to achieve better health outcomes and participate efficiently in health production [15-21], and it can be modified with appropriate interventions [19,23]. However, in accordance with recent systematic reviews, we emphasize that the causal link between eHealth literacy and favorable patient outcomes related to health status, risk behaviors, or experiences with the health care system has not been established yet [31]. Until robust methods and clear causal links are in place, we suggest caution when implementing large-scale public health interventions based on overoptimistic expectations. Although low eHealth literacy was associated with the presence of chronic conditions, its association with a number of health-related outcomes and health behaviors has been mixed [22,31]. Our study draws the attention on another potential risk group: individuals who rate their eHealth literacy in the highest range. We found that high eHealth literacy levels were associated with both positive and negative patient experiences, a relationship which requires further exploration. Understanding the drivers of inferior experiences may help to design efficient eHealth literacy interventions, which provide individuals with resilience to navigate the health system efficiently, enable them to engage in productive partnership with HCPs, and ultimately turn online information into better health outcomes and satisfactory patient experience.

As a conclusion, our results suggest that eHealth literacy, a modifiable patient-related factor [19,23], is associated with PREMs. It is tempting to develop interventions that develop eHealth literacy along with the eHealth infrastructure and eHealth interventions aiming for better patient experiences. However, further studies are needed to establish the causal relationship between eHealth literacy and patient-reported experiences, with special focus on vulnerable individuals with low eHealth literacy levels.

Acknowledgments

This publication was supported by the Higher Education Institutional Excellence Program of the Ministry of Ministry for Innovation and Technology in the framework of the "Financial and Public Services" research project (NKFIH-1163-10/2019) at the Corvinus University of Budapest. The contribution of OBF occurred within a Marie Skłodowska-Curie Innovative Training Network (HealthPros—Healthcare Performance Intelligence Professionals) that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 765141 (https://healthpros-h2020.eu).

Conflicts of Interest

None declared.

Multimedia Appendix 1
eHealth Literacy Scale (eHEALS).
[PDF File (Adobe PDF File), 244 KB - jmir_v22i8e19013_app1.pdf]

Multimedia Appendix 2
OECD-proposed Set of Questions on Patient Experiences with Ambulatory Care (PREM).
[PDF File (Adobe PDF File), 141 KB - jmir_v22i8e19013_app2.pdf]

Multimedia Appendix 3
Logistic regression analysis of unmet medical needs.
[PDF File (Adobe PDF File), 167 KB - jmir_v22i8e19013_app3.pdf]

Multimedia Appendix 4
Regression analysis of waiting times.
[PDF File (Adobe PDF File), 172 KB - jmir_v22i8e19013_app4.pdf]

Multimedia Appendix 5
Regression analyses of unmet medical needs and waiting times PREM scores.
[PDF File (Adobe PDF File), 175 KB - jmir_v22i8e19013_app5.pdf]

Multimedia Appendix 6
Sensitivity analysis.
[PDF File (Adobe PDF File), 280 KB - jmir_v22i8e19013_app6.pdf]

Multimedia Appendix 7
Analysis of unmet medical needs on extended sample.
[PDF File (Adobe PDF File), 186 KB - jmir_v22i8e19013_app7.pdf]

References

1. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21 Century. Washington, DC: Institute of Medicine; 2005.
2. Jo Delaney L. Patient-centred care as an approach to improving health care in Australia. Collegian 2018 Feb;25(1):119-123. [doi: 10.1016/j.colegn.2017.02.005]
3. Boncz I, Sebestyén A. Financial deficits in the health services of the UK and Hungary. Lancet 2006 Sep 09;368(9539):917-918. [doi: 10.1016/S0140-6736(06)69369-0] [Medline: 16962878]
4. Tsevat J, Moriates C. Value-Based Health Care Meets Cost-Effectiveness Analysis. Ann Intern Med 2018 Sep 04;169(5):329-332. [doi: 10.7326/M18-0342] [Medline: 30083766]
5. Expert Panel on Effective Ways of Investing in Health (EXPH). Opinion on defining value in “value-based healthcare”. 2019. URL: https://ec.europa.eu/health/expert_panel/sites/expertpanel/files/024_valuebasedhealthcare_en.pdf [accessed 2020-07-28]
6. OECD. Ministerial Statement: the Next Generation of Health Reforms, OECD Health Ministerial Meeting, 17 January 2017. 2017. URL: https://www.oecd.org/newsroom/oecd-health-ministerial-statement-the-next-generation-of-health-reforms.htm [accessed 2020-07-28]
7. Medeiros J, Schwierz C. Efficiency estimates of health care systems. 2015. URL: https://ec.europa.eu/economy_finance/publications/economic_paper/2015/pdf/ecp549_en.pdf [accessed 2020-07-28]
8. World Health Organization. WHO global strategy on people-centred and integrated health services - Interim report. Geneva: WHO; 2015. URL: https://www.who.int/servicedeliverysafety/areas/people-centred-care/global-strategy/en/ [accessed 2020-07-28]
9. Anderson JG, Rainey MR, Eysenbach G. The impact of CyberHealthcare on the physician-patient relationship. J Med Syst 2003 Feb;27(1):67-84. [Medline: 12617199]
10. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
11. Executive Board, 142 (World Health Organization). mHealth: Use of appropriate digital technologies for public health: Report by the director-general. Geneva. 2017. URL: https://apps.who.int/iris/handle/10665/274134 [accessed 2020-07-28]
12. World Health Organization. Agenda item 4 in Seventy-first World Health Assembly. Digital health resolution. 2018. URL: https://apps.who.int/gb/ebwha/pdf_files/WHA71/A71_R7-en.pdf?ua=1 [accessed 2020-07-28]
13. European Commission. Communication on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society. 2018. URL: https://ec.europa.eu/digital-single-market/en/news/communication-enabling-digital-transformation-health-and-care-digital-single-market-empowering [accessed 2020-07-28]
14. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
15. Aponte J, Nokes KM. Electronic health literacy of older Hispanics with diabetes. Health Promot Int 2017 Jun 01;32(3):482-489. [doi: 10.1093/heapro/dav112] [Medline: 26681770]

16. Hogan TP, Hill JN, Locatelli SM, Weaver FM, Thomas FP, Nazi KM, et al. Health Information Seeking and Technology Use Among Veterans With Spinal Cord Injuries and Disorders. PM R 2016 Feb;8(2):123-130. [doi: 10.1016/j.pmrj.2015.06.443] [Medline: 26164352]
17. Vicente MR, Madden G. Assessing eHealth skills across Europeans. Health Policy and Technology 2017 Jun;6(2):161-168. [doi: 10.1016/j.hlpt.2017.04.001]
18. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
19. Kim S, Son Y. Relationships Between eHealth Literacy and Health Behaviors in Korean Adults. Comput Inform Nurs 2017 Mar;35(2):84-90. [doi: 10.1097/CIN.0000000000000255] [Medline: 27258808]
20. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth Literacy With Health Behavior Among Adult Internet Users. J Med Internet Res 2016 Jul;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
21. Neuhauser L, Kreps GL. eHealth communication and behavior change: promise and performance. Social Semiotics 2010 Feb 17;20(1):9-27. [doi: 10.1080/10350330903438386]
22. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
23. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
24. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011 Nov;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]
25. Mitsuhashi T. Effects of two-week e-learning on eHealth literacy: a randomized controlled trial of Japanese Internet users. PeerJ 2018;6:e5251 [FREE Full text] [doi: 10.7717/peerj.5251] [Medline: 30013857]
26. Fujisawa R, Klzainga S. Measuring patient experiences (PREMS): Progress made by the OECD and its member countries between 2006 and 2016. 2017. URL: https://doi.org/10.1787/893a07d2-en [accessed 2020-07-28]
27. OECD, European Observatory on Health Systems and Policies. Hungary: Country Health Profile 2019. Brussels: State of Health in the EU, OECD Publishing, Paris/European Observatory on Health Systems and Policies; 2019. URL: https://doi.org/10.1787/4b7ba48c-en [accessed 2020-07-28]
28. Baji P, Pavlova M, Gulácsi L, Zsófia HC, Groot W. Informal payments for healthcare services and short-term effects of the introduction of visit fee on these payments in Hungary. Int J Health Plann Manage 2012;27(1):63-79. [doi: 10.1002/hpm.1106] [Medline: 22009836]
29. Rotar AM, Van Den Berg MJ, Schäfer W, Kringos DS, Klazinga NS. Shared decision making between patient and GP about referrals from primary care: Does gatekeeping make a difference? PLoS One 2018;13(6):e0198729 [FREE Full text] [doi: 10.1371/journal.pone.0198729] [Medline: 29889861]
30. Brito Fernandes Ó, Baji P, Kringos D, Klazinga N, Gulácsi L, Lucevic A, et al. Patient experiences with outpatient care in Hungary: results of an online population survey. Eur J Health Econ 2019 Jun;20(Suppl 1):79-90 [FREE Full text] [doi: 10.1007/s10198-019-01064-z] [Medline: 31098884]
31. Neter E, Brainin E. Association between health literacy, eHealth literacy, and health outcomes among patients with long-term conditions. European Psychologist 2019 Jan;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
32. FitzPatrick MA, Hess AC, Sudbury-Riley L, Schulz PJ. A Typology of Patients Based on Decision-Making Styles: Cross-Sectional Survey Study. J Med Internet Res 2019 Dec 20;21(11):e15332 [FREE Full text] [doi: 10.2196/15332] [Medline: 31746770]
33. Tan SS, Goonawardene N. Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
34. Dunn P, Hazzard E. Technology approaches to digital health literacy. Int J Cardiol 2019 Oct 15;293:294-296. [doi: 10.1016/j.ijcard.2019.06.039] [Medline: 31350037]
35. Conard S. Best practices in digital health literacy. Int J Cardiol 2019 Oct 01;292:277-279. [doi: 10.1016/j.ijcard.2019.05.070] [Medline: 31230937]
36. Chung JE. Patient-provider discussion of online health information: results from the 2007 Health Information National Trends Survey (HINTS). J Health Commun 2013;18(6):627-648. [doi: 10.1080/10810730.2012.743628] [Medline: 23590202]
37. Nejati B, Lin C, Aaronson NK, Cheng ASK, Browall M, Lin C, et al. Determinants of satisfactory patient communication and shared decision making in patients with multiple myeloma. Psychooncology 2019 Jul;28(7):1490-1497. [doi: 10.1002/pon.5105] [Medline: 31087365]
38. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
39. Zrubka Z, Hajdu O, Rencz F, Baji P, Gulácsi L, Péntek M. Psychometric properties of the Hungarian version of the eHealth Literacy Scale. Eur J Health Econ 2019 Jul;20(Suppl 1):57-69 [FREE Full text] [doi: 10.1007/s10198-019-01062-1] [Medline: 31098883]

40. Rencz F, Tamási B, Brodszky V, Gulácsi L, Weszl M, Péntek M. Validity and reliability of the 9-item Shared Decision Making Questionnaire (SDM-Q-9) in a national survey in Hungary. Eur J Health Econ 2019 Jul;20(Suppl 1):43-55. [doi: 10.1007/s10198-019-01061-2] [Medline: 31111402]
41. Lucevic A, Péntek M, Kringos D, Klazinga N, Gulácsi L, Brito Fernandes Ó, et al. Unmet medical needs in ambulatory care in Hungary: forgone visits and medications from a representative population survey. Eur J Health Econ 2019 Jul;20(Suppl 1):71-78 [FREE Full text] [doi: 10.1007/s10198-019-01063-0] [Medline: 31102157]
42. Hungarian Central Statistical Office. Population Census. 2018. URL: http://www.ksh.hu/nepszamlalas/tables_regional_00?lang=en [accessed 2011-11-01]
43. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
44. Morton S, Rolph J. Public Policy and Statistics - Case Studies from RAND. New York: Springer Verlag; 2000.
45. Xesfingi S, Vozikis A. eHealth literacy: in the quest of the contributing factors. Interact J Med Res 2016 May 25;5(2):e16 [FREE Full text] [doi: 10.2196/ijmr.4749] [Medline: 27226146]
46. National Library of Medicine. Medical Subject Headings (MeSH): Age Groups. 1998. URL: https://www.ncbi.nlm.nih.gov/mesh/68009273 [accessed 2020-07-28]
47. Parker RN, Fenwick R. The Pareto Curve and Its Utility for Open-Ended Income Distributions in Survey Research. Social Forces 1983 Mar 01;61(3):872-885. [doi: 10.1093/sf/61.3.872]
48. Hungarian Central Statistical Office. Income deciles of all households by region and type of settlement. 2019. URL: https://www.ksh.hu/docs/hun/xstadat/xstadat_eves/i_zhc014a.html?down=2650 [accessed 2020-05-30]
49. Cox B, van Oyen H, Cambois E, Jagger C, le Roy S, Robine J, et al. The reliability of the Minimum European Health Module. Int J Public Health 2009;54(2):55-60. [doi: 10.1007/s00038-009-7104-y] [Medline: 19183846]
50. Eurostat. Glossary: Minimum European Health Module (MEHM). 2017. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Minimum_European_Health_Module_(MEHM) [accessed 2020-07-28]
51. Bartolucci F, Bacci S, Gnaldi M. Statistical Analysis of Questionnaires - A Unified Approach Based on R and Stata. Boca Raton, FL: CRC Press; 2016.
52. Royston JP. An Extension of Shapiro and Wilk's W Test for Normality to Large Samples. Applied Statistics 1982;31(2):115. [doi: 10.2307/2347973]
53. Kaiser HF, Rice J. Little Jiffy, Mark Iv. Educational and Psychological Measurement 2016 Jul 02;34(1):111-117. [doi: 10.1177/001316447403400115]
54. Kolenikov S, Angeles G. The Use of Discrete Data in PCA: Theory, Simulations and Applications to Socioeconomic Indices: a Working Paper. Chapel Hill, NC: Carolina Population Center; 2004. URL: https://www.measureevaluation.org/resources/publications/wp-04-85 [accessed 2020-07-28]
55. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952 Dec;47(260):583-621. [doi: 10.1080/01621459.1952.10483441]
56. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine 2009 Apr 21;50(302):157-175. [doi: 10.1080/14786440009463897]
57. Fischer R. Statistical Methods for Research Workers. Edinburgh: Oliver & Boyd; 1925.
58. Wooldridge J. Introductory Econometrics: A Modern Approach (5th edition). Mason, OH: South-Western, Cengage Learning; 2013.
59. Verbeek M. A Guide to Modern Econometrics (2nd edition). Chicester, England: John Wiley & Sons; 2008.
60. Fagerland MW, Hosmer DW. A goodness-of-fit test for the proportional odds regression model. Stat Med 2013 Jul 15;32(13):2235-2249. [doi: 10.1002/sim.5645] [Medline: 23037691]
61. StataCorp LP. Stata Statistical Software: Release 14 [Software]. College Station, TX: StataCorp LP; 2015.
62. Pituch K, Stevens J. Applied Multivariate Statistics for the Social Sciences - Analysis With SAS and IBM's SPSS (6th edition). New York: Routledge; 2006.
63. Hungarian Forint. European Central Bank. 2020. URL: https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/eurofxref-graph-huf.en.html [accessed 2020-07-28]
64. Lucevic A, Péntek M, Kringos D, Klazinga N, Gulácsi L, Zrubka Z, et al. NS197: Waiting times for health care services in Hungary: results of a representative population survey. Value in Health 2019 Nov;22:S794. [doi: 10.1016/j.jval.2019.09.2097]
65. Petrič G, Atanasova S, Kamin T. Impact of Social Processes in Online Health Communities on Patient Empowerment in Relationship With the Physician: Emergence of Functional and Dysfunctional Empowerment. J Med Internet Res 2017 Dec 13;19(3):e74. [doi: 10.2196/jmir.7002] [Medline: 28288953]
66. Brabers AEM, Rademakers JJDJM, Groenewegen PP, van Dijk L, de Jong JD. What role does health literacy play in patients' involvement in medical decision-making? PLoS One 2017;12(3):e0173316 [FREE Full text] [doi: 10.1371/journal.pone.0173316] [Medline: 28257472]

67. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
68. Diviani N, Fredriksen EH, Meppelink CS, Mullan J, Rich W, Sudmann TT. Where else would I look for it? A five-country qualitative study on purposes, strategies, and consequences of online health information seeking. J Public Health Res 2019 Mar 11;8(1):1518 [FREE Full text] [doi: 10.4081/jphr.2019.1518] [Medline: 31044138]
69. Kayser L, Rossen S, Karnoe A, Elsworth G, Vibe-Petersen J, Christensen JF, et al. Development of the Multidimensional Readiness and Enablement Index for Health Technology (READHY) Tool to Measure Individuals' Health Technology Readiness: Initial Testing in a Cancer Rehabilitation Setting. J Med Internet Res 2019 Feb 12;21(2):e10377 [FREE Full text] [doi: 10.2196/10377] [Medline: 30747717]
70. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
71. Diviani N, van den Putte B, Meppelink CS, van Weert JCM. Exploring the role of health literacy in the evaluation of online health information: Insights from a mixed-methods study. Patient Educ Couns 2016 Jun;99(6):1017-1025. [doi: 10.1016/j.pec.2016.01.007] [Medline: 26817407]
72. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol 1999 Dec;77(6):1121-1134. [Medline: 10626367]
73. Schulz PJ, Nakamoto K. Patient behavior and the benefits of artificial intelligence: the perils of. Patient Educ Couns 2013 Aug;92(2):223-228. [doi: 10.1016/j.pec.2013.05.002] [Medline: 23743214]
74. Paige SR, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C. Proposing a Transactional Model of eHealth Literacy: Concept Analysis. J Med Internet Res 2018 Oct 02;20(10):e10175 [FREE Full text] [doi: 10.2196/10175] [Medline: 30279155]
75. Bhandari N, Shi Y, Jung K. Seeking health information online: does limited healthcare access matter? J Am Med Inform Assoc 2014;21(6):1113-1117 [FREE Full text] [doi: 10.1136/amiajnl-2013-002350] [Medline: 24948558]
76. Li N, Orrange S, Kravitz RL, Bell RA. Reasons for and predictors of patients' online health information seeking following a medical appointment. Fam Pract 2014 Oct;31(5):550-556. [doi: 10.1093/fampra/cmu034] [Medline: 24963151]
77. Edwards M, Wood F, Davies M, Edwards A. The development of health literacy in patients with a long-term health condition: the health literacy pathway model. BMC Public Health 2012;12:130 [FREE Full text] [doi: 10.1186/1471-2458-12-130] [Medline: 22332990]
78. Scrivner B, Szaflarski M, Baker EH, Szaflarski JP. Health literacy and quality of life in patients with treatment-resistant epilepsy. Epilepsy Behav 2019 Oct;99:106480. [doi: 10.1016/j.yebeh.2019.106480] [Medline: 31465909]
79. van der Aa MJ, van den Broeke JR, Stronks K, Plochg T. Patients with multimorbidity and their experiences with the healthcare process: a scoping review. J Comorb 2017;7(1):11-21 [FREE Full text] [doi: 10.15256/joc.2017.7.97] [Medline: 29090185]
80. Marshall-Aiyelawo K, Bannick R, Beamer S. Effect of change in the CG CAHPS survey instrument recall period on patient experience scores on healthcare utilization. Patient Experience Journal 2019;6(3):114-123. [doi: 10.35680/2372-0247.1359]
81. Mackert M, Champlin SE, Holton A, Muñoz II, Damásio MJ. eHealth and Health Literacy: A Research Methodology Review. J Comput-Mediat Comm 2014 Apr 12;19(3):516-528. [doi: 10.1111/jcc4.12044]

Edited by G Eysenbach; submitted 31.03.20; peer-reviewed by MA Bahrami, K Alipasandi, CY Lin; comments to author 28.04.20; revised version received 10.06.20; accepted 15.06.20; published 11.08.20.

Please cite as:
Zrubka Z, Brito Fernandes Ó, Baji P, Hajdu O, Kovacs L, Kringos D, Klazinga N, Gulácsi L, Brodszky V, Rencz F, Péntek M
Exploring eHealth Literacy and Patient-Reported Experiences With Outpatient Care in the Hungarian General Adult Population: Cross-Sectional Study
J Med Internet Res 2020;22(8):e19013
URL: https://www.jmir.org/2020/8/e19013
doi:10.2196/19013
PMID:32667891

Review

Applying the Electronic Health Literacy Lens: Systematic Review of Electronic Health Interventions Targeted at Socially Disadvantaged Groups

Christina Cheng[1,2], MHlthProm; Alison Beauchamp[3,4,5], PhD; Gerald R Elsworth[1], PhD; Richard H Osborne[1], PhD

[1]Centre for Global Health and Equity, Faculty of Health, Arts and Design, Swinburne University of Technology, Hawthorn, Australia

[2]Deakin University, School of Health and Social Development, Faculty of Health, Burwood, Australia

[3]Department of Rural Health, Monash University, Melbourne, Australia

[4]Department of Medicine - Western Health, The University of Melbourne, Melbourne, Australia

[5]Australian Institute for Musculoskeletal Science, Sunshine Hospital, St Albans, Australia

Corresponding Author:
Christina Cheng, MHlthProm
Centre for Global Health and Equity
Faculty of Health, Arts and Design
Swinburne University of Technology
Room 907, Level 9, AMDC Building
453/469-477 Burwood Road
Hawthorn, VIC 3122
Australia
Phone: 61 392145470
Email: cccheng@swin.edu.au

Abstract

Background: Electronic health (eHealth) has the potential to improve health outcomes. However, eHealth systems need to match the eHealth literacy needs of users to be equitably adopted. Socially disadvantaged groups have lower access and skills to use technologies and are at risk of being digitally marginalized, leading to the potential widening of health disparities.

Objective: This systematic review aims to explore the role of eHealth literacy and user involvement in developing eHealth interventions targeted at socially disadvantaged groups.

Methods: A systematic search was conducted across 10 databases for eHealth interventions targeted at older adults, ethnic minority groups, low-income groups, low-literacy groups, and rural communities. The eHealth Literacy Framework was used to examine the eHealth literacy components of reviewed interventions. The results were analyzed using narrative synthesis.

Results: A total of 51 studies reporting on the results of 48 interventions were evaluated. Most studies were targeted at older adults and ethnic minorities, with only 2 studies focusing on low-literacy groups. eHealth literacy was not considered in the development of any of the studies, and no eHealth literacy assessment was conducted. User involvement in designing interventions was limited, and eHealth intervention developmental frameworks were rarely used. Strategies to assist users in engaging with technical systems were seldom included in the interventions, and accessibility features were limited. The results of the included studies also provided inconclusive evidence on the effectiveness of eHealth interventions.

Conclusions: The findings highlight that eHealth literacy is generally overlooked in developing eHealth interventions targeted at socially disadvantaged groups, whereas evidence about the effectiveness of such interventions is limited. To ensure equal access and inclusiveness in the age of eHealth, eHealth literacy of disadvantaged groups needs to be addressed to help avoid a digital divide. This will assist the realization of recent technological advancements and, importantly, improve health equity.

(J Med Internet Res 2020;22(8):e18476) doi:10.2196/18476

KEYWORDS

eHealth; health literacy; internet; health care; telecommunications; digital divide; health equity

Introduction

Background

Electronic health (eHealth), "the use of information and communications technology (ICT) in support of health and health-related fields" [1], is increasingly being integrated into the delivery of health resources and services. The World Health Organization (WHO) [2] also recognizes that digital technologies have the potential to accelerate toward achieving Sustainable Development Goals by improving health services. However, not everyone has substantive ICT access or skills to take advantage of the benefits of eHealth.

The issue of inequitable access, usage or skills, and outcomes relating to ICT by subgroups of society, described as the *digital divide* [3-6], is a recognized public health concern [7]. The sociodemographic factors associated with health disparities, such as age, income, education, and ethnicity, are similar to the characteristics of people who have limited ICT access or skills [8-10]. Older age, less education, lower income, being from an ethnic minority group, or living in a remote area are all associated with decreased access or less use of the internet for activities such as health information seeking, communicating with health care providers, monitoring health, or using personal health records [11-15]. As such, these socially disadvantaged groups are usually overlooked in eHealth design [15] and are at risk of becoming digitally marginalized [7,16], leading to a potential widening of health disparities.

In recognition of the different sets of skills required for using eHealth, the concept of eHealth literacy, defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem," was introduced in 2006 [17]. This concept is grounded in health literacy [17,18], which is recognized as a critical determinant of health [19]. The concept of eHealth literacy has since been considered amid the everchanging landscape of ICT, and there is also a growing recognition that eHealth strategies will be ineffective and inequitable if the eHealth literacy needs of users are not addressed [20-22]. In 2015, Norgaard et al [23] developed the eHealth Literacy Framework (eHLF) by integrating the perspectives and experiences of a wide range of eHealth stakeholders, and 7 domains of eHealth literacy were identified. On the basis of this framework and applying a validity-driven approach to scale development [24], the eHealth Literacy Questionnaire (eHLQ) was also developed and tested [25]. The 7 domains of eHealth literacy are as follows: (1) using technology to process health information, (2) understanding of health concepts and language, (3) ability to actively engage with digital services, (4) feel safe and in control, (5) motivated to engage with digital services, (6) access to digital services that work and (7) digital services that suit individual needs.

According to the eHLF, eHealth literacy is not only the ability of an individual user but also relates to the system and how the two interact. For an eHealth intervention to be adopted, the system needs to align with the eHealth literacy needs of target users [23], which may differ across settings and contexts [25]. By assessing the eHealth literacy of target users, weaknesses in certain domains of eHealth literacy can be identified, and interventions can be designed to respond to the relevant weaknesses [23].

In reviewing the evaluation of the now defunct UK web-based personal health record HealthSpace, Monkman and Kushniruk [26] commented that the system did not match the eHealth literacy or information needs of users. Apart from the consideration of literacy, the evaluation also recommended that user-centered principles, such as involving users in design and development [27-29], be applied in any future endeavors [30].

Objectives

eHealth literacy plays an important role in improving health outcomes across the socioeconomic spectrum. This systematic review aimed to apply an eHealth literacy lens to explore current practices in the development of eHealth interventions targeting socially disadvantaged groups, who are at risk of being digitally marginalized. Guided by the eHLF, this review examined not only the usability of eHealth interventions but also how interventions motivate users or address privacy concerns as part of the effort to respond to eHealth literacy needs. With the WHO recognizing health literacy as having the potential to empower and drive equity [19], insights into how interventions meet the needs of disadvantaged groups will highlight gaps in research and advance the role of eHealth literacy in making eHealth more accessible. The purpose of this review was to answer the following research questions:

1. Was eHealth literacy considered during the development of eHealth interventions targeted at socially disadvantaged groups? If yes, what approaches were used to determine the eHealth literacy needs of the target group?
2. What frameworks or theories were used to guide the development of eHealth interventions besides theories on eHealth literacy?
3. Were users involved in the development of eHealth interventions?
4. What eHealth literacy domains, as described in the eHLF, were likely addressed in the identified eHealth interventions?
5. Were eHealth interventions targeted at socially disadvantaged groups effective when eHealth literacy was considered?

Methods

Review Design

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols 2015 checklist [31]. This was a review with no patient or public involvement.

Eligibility Criteria

The development of the inclusion criteria was based on the *PICO* (population, intervention, control, and outcomes) model [32]. The population referred to socially disadvantaged groups with any health condition, who were disadvantaged because of age, education, migrant status, living in a rural or remote area, or socioeconomic status [33]. For age, older adults were defined

as people aged 60 years or older [34]. An intervention referred to eHealth interventions, systems, or applications mainly delivered through the internet via ICT such as computers, tablets, or mobile phones, targeted and operated by individual participants through platforms such as websites, apps, social media, email, or text messaging [35]. The interventions were those aimed at improving health or preventing or reducing the risk of illness. Study design included randomized controlled trials (RCTs) and non-RCT studies. Outcomes included clinical health outcomes or health knowledge and behavior. Only studies published in English peer-reviewed journals with full text available were included. Publication dates of studies were from January 2007 to July 2019. January 2007 was chosen because the concept of eHealth literacy was first introduced in late 2006 [17].

Studies were excluded if they were protocols, literature or systematic reviews, and studies of nonhealth outcomes, such as feasibility studies, usability evaluations, or economic evaluations. Studies of telehealth or telemedicine for monitoring physical conditions or medications that required no active participation from participants or only for communication with carers and health professionals were excluded. In addition, studies of consultations via videoconferencing, eHealth initiatives for risk assessment of physical conditions or motor- or cognitive skills training, or computer skills training and eHealth programs targeted at health care providers or carers were excluded. In cases where studies based on the same intervention with similar outcome measures were identified, any pilot studies of that intervention were excluded.

Search Strategy and Study Selection

Studies were identified from 10 databases, including Academic Search Complete, AgeLine, Cumulative Index to Nursing and Allied Health Literature (CINAHL) Complete, Communication & Mass Media Complete, Excerpta Medica dataBASE (EMBASE), Education Resources Information Center (ERIC), Global Health, Medical Literature Analysis and Retrieval System Online (MEDLINE) Complete, American Psychological Association PsycInfo database (PsycINFO), and Sociology Research Database (SocINDEX), with searches conducted in November 2018 and updated in July 2019. Search terms were based on keywords from the inclusion criteria (Multimedia Appendix 1). The reference lists of relevant studies were also scanned for potential studies. The search and screening of titles and abstracts were conducted by one author (CC), who also reviewed the full text of potential studies with reasons for exclusion documented.

Data Extraction and Quality Assessment

Following study selection, data were extracted based on the research questions, and study quality was appraised using the Effective Public Health Practice Project Quality Assessment Tool (Multimedia Appendix 2) [36-87]. The tool is considered a valid and reliable instrument, adaptable to most public health systematic reviews for evaluating a range of study designs [88,89]. Data extraction and quality assessment of 10% (6/51) of the included studies were independently reviewed by 2 authors (AB and CC). Discrepancies were resolved through discussion and consensus. Decisions from the discussion were used to guide the data extraction and the quality assessment of the remaining studies undertaken by one author (CC).

Data Analysis

Owing to the heterogeneity of study designs and outcome measures among the included studies, a narrative synthesis was used to answer the research questions. For the research question relating to whether eHealth literacy domains were likely addressed in interventions, a directed content analysis approach was adopted. This approach allows researchers to use an existing theory or framework as coding categories, to develop operational definitions for each category as determined by the theory or framework, and to analyze the content accordingly [90]. For this review, the eHLF was used to code the eHealth literacy domains. The intervention components likely addressing each eHealth literacy domain were based on components derived from the concept mapping workshops used to develop the framework [23,25] and matched with the description of the intervention in the included studies. For example, the use of passwords to access the system or intervention is expected to promote a sense of security. Hence, the feature is coded as meeting the needs of *Domain 4 Feel safe and in control*. Providing information in users' preferred language for interventions that target ethnic minorities will be a component that matches *Domain 7 Digital services that suit individual needs*. The classification of intervention components was initially undertaken by one author (CC), followed by discussion and review with one of the eHLF developers (RO) and among the other authors. The details of the intervention components relating to the eHealth literacy domains are presented in Table 1 [23]. The coding of 10% (6/51) of the studies was independently conducted by 2 authors (AB and CC). Discrepancies were resolved by discussion and consensus. Decisions following discussion were used to guide the coding of the remaining studies undertaken by 1 author (CC).

For the research question regarding the effectiveness of interventions, the overall effect size could not be determined because of the diversity of outcome measures and data analysis methods. Therefore, effectiveness was estimated by reporting statistically significant improvement between intervention and control groups or between before and after intervention for one-group pretest-posttest for the outcome measures stipulated. If more than 1 primary outcome measure was stated, only the first 3 were included.

Table 1. Examples of intervention components that likely address electronic health literacy domains derived from the eHealth Literacy Framework.

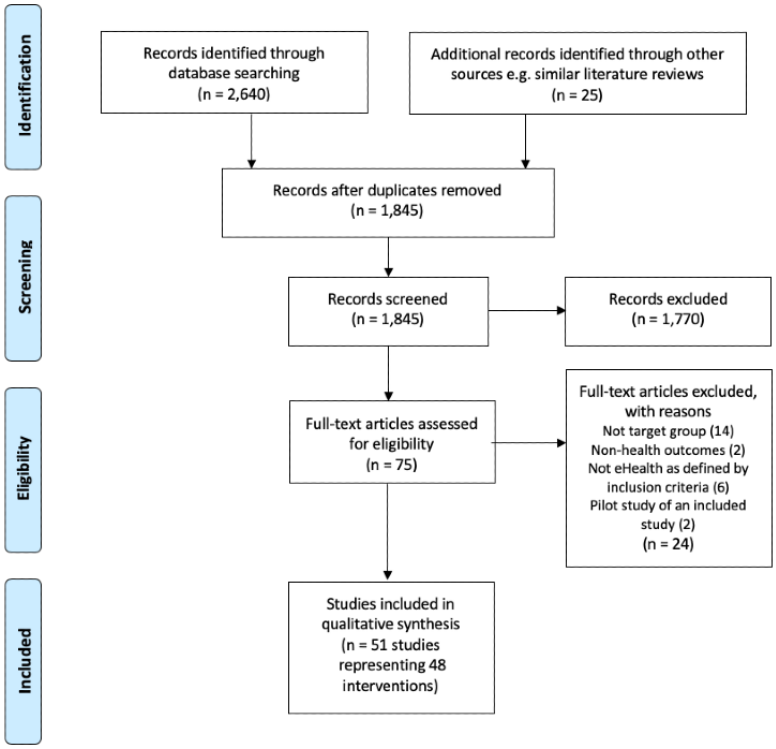
Descriptions	Examples of intervention components
1. Using technology to process health information	
Able to read, write, and remember; apply basic numerical concepts; and understand context-specific language (such as health, technology, and English) as well as critically appraise information. Know when, how, and what information to use	• Contains information about health conditions • Contains health information in a format that can easily be understood (such as text in low reading grade, video, graphics, animations, graphs, stories, examples, culturally or locally relevant materials) • Contains information that can help make decisions • Can use the system to share information with family, friends, and health professionals • Can use the system to organize or record personal health information (such as recording or monitoring activities, journal, diary, worksheets) • Provides access to other information resources
2. Understanding of health concepts and language	
Know about basic physiological functions and own current health status. Aware of risk factors and how to avoid them or reduce their influence on own health	• Contains information that one can take responsibility for one's own health (such as setting personal goals or plans, monitoring health, practical skills or tips, practical and usable information such as recipes, activities or opportunities to join events, and download information) • Tailored information, instructions or personal guidance, and chat sessions • Homework assignments or tests of knowledge or evaluation • Provide easy-to-use tools for measurements or assessment or monitoring
3. Ability to actively engage with digital services	
Being comfortable using digital services for handling information	• Easy navigation around the system • Detailed and easy-to-understand instructions • Provide training or a manual to use the system
4. Feel safe and in control	
Feel that they have the ownership of personal data stored in the system and that their data are safe and can be accessed only by people to whom the data are relevant (such as own doctor and nurse)	• Unique username and password protected • Secure website or database or communication • Provide means to ensure privacy • *Closed* system to which only authorized personnel have access • Can maintain anonymity if needed
5. Motivated to engage with digital services	
Feel that engaging in the use of digital services will be useful for them in managing their health	• Incentives to return to use the systems • Encouragement to continue to use the systems • Alerts and notifications • Quick response to queries • Provides tailored feedback, progress reports, or support • Provides new content regularly • Regular meetup sessions or discussion forums • Provides peer or professional support • Quick and easy communication (such as sending or receiving emails, asking questions, and inquiries)
6. Access to digital services that work	
Have access to digital services that the users trust to be working when they need it and as they expect it to work	• Provides access to the hardware or system • Provides technical support • Can be accessed anytime anywhere • Access to tools or devices that can be integrated into the system
7. Digital services that suit individual needs	
Have access to digital services that suit the specific needs and preferences of the users. This includes responsive features of both the information technology and health care system as well as adaptation of devices and interfaces to be used by people with physical and mental disabilities	• Consists of accessibility features such as change of font size or audio function • Easy to use, efficient, and user-friendly interface (such as large buttons and large icon) • Available in users' preferred language

Results

Selection of Studies

The search resulted in 2640 studies; after removing 820 duplicates and an additional 25 records identified through other sources, a total of 1845 records were screened. Following a review of titles and abstracts, 75 studies were retrieved for full text review. A total of 24 studies were excluded for reasons including non–target groups, non–health outcomes, and not eHealth as defined by the inclusion criteria and 2 pilot studies of included studies, resulting in 51 studies reporting on the findings of 48 interventions (Figure 1).

Figure 1. Selection process using the PRISMA Flow Diagram [31].



Characteristics of the Studies

Among the 51 studies, 11 were pilot studies, 43 were RCTs (40 were two-armed RCTs, 2 were three-armed RCTs, and 1 was a five-armed RCT). A total of 7 studies used the one-group pretest-posttest design, and 1 was a case study. Among the 43 RCTs, 42 control groups received no intervention, waitlist, standard care, usual care, or in-person education, whereas 1 study did not describe this. The sample size ranged from 1 to 755 (Table 2).

Apart from the postintervention assessment, 13 studies also conducted follow-up assessments, ranging from 1 week to 12 months. Clinical health outcomes were reported in 22 studies, whereas 28 studies measured health-related outcomes such as attitude, behaviors, or knowledge, and 1 study measured both behavior and clinical outcomes. For the quality rating, 2 studies were rated as strong, 24 studies were rated as moderate, and 25 studies were of weak quality. Among the quality rating criteria, 46 studies received a weak rating for selection bias because of their recruitment strategy or fewer than 60% of eligible participants taking part. Only 1 study received a strong rating for blinding, whereas the remaining studies either indicated blinding was not possible or did not report on blinding (Multimedia Appendix 2).

Table 2. Characteristics of studies.

Authors (year)	Study designs	Sample sizes	Outcome measures	Quality ratings
Agyapong et al (2017) [37]	Two-armed RCT[a]	73	• BDI-II[b]	Strong
Anand et al (2016) [38]	Two-armed RCT	343	• Myocardial Infarction Risk Score	Moderate
Arora et al (2014) [39]	Two-armed RCT	128	• HbA_{1c}[c]	Moderate
Bennett et al (2018) [40]	Two-armed RCT	351	• Body weight	Weak
Bond et al (2010) [41]	Two-armed RCT	62	• CES-D[d] • The Problem Areas in Diabetes Scale • Diabetes Support Scale	Weak
Broekhuizen et al (2016) [42] and Wijsman et al (2013) [43]	Two-armed RCT	236	• Research and Development 36-item health survey • Ankle and wrist accelerometer	Moderate
Buller et al (2008) [44]	Two-armed RCT	755	• Adapted all-day screener and self-report of servings	Moderate
Carroll et al (2019) [45]	Two-armed RCT	360	• Patient Activation Measure	Weak
Caster et al (2017) [46]	One-group pretest-posttest	243	• Knowledge scores	Weak
Chen et al (2016) [47]	Case study	1	• Sleep satisfaction rating	Weak
Chen et al (2018) [48]	Two-armed RCT	233	• Attendance rate • Diabetic retinopathy • Knowledge scores	Strong
Choi et al (2012) [49][e]	Two-armed RCT	63	• Chinese versions of Beck Depression Inventory • Chinese bilingual version of PHQ-9[f]	Weak
Dang et al (2017) [50]	Two-armed (2:1) RCT	61	• Self-Efficacy for Managing Chronic Disease	Moderate
Dear et al (2015) [51][e]	Two-armed RCT	72	• GAD-7[g] • PHQ-9	Weak
Dugas et al (2018) [52]	Five-armed RCT	27	• HbA_{1c}	Weak
Fortmann et al (2017) [53]	Two-armed RCT	126	• HbA_{1c}	Weak
Gilmore et al (2017) [54]	Two-armed RCT	40	• Body weight	Moderate
Griffin et al (2018) [55]	One-group pretest-posttest	109	• Body weight • BMI	Weak
Hacking et al (2016) [56]	Two-armed RCT	223	• Knowledge scores	Weak
Hageman et al (2014) [57]	Three-armed RCT	289	• Blood pressure • BMI • Waist circumference	Moderate
Herring et al (2017) [58]	Two-armed RCT	66	• Body weight	Moderate
Hill et al (2006) [59][h]	Two-armed RCT	120	• The Personal Resource Questionnaire • Rosenberg Self-Esteem Scale • Chronic Illness Empowerment Scale	Weak

Authors (year)	Study designs	Sample sizes	Outcome measures	Quality ratings
Hong et al (2015) [60]	One-group pretest-posttest	30	• Quality of life (self-reported seven-item questionnaire) • Level of PA[i]	Weak
Ingersoll et al (2015) [61]	Two-armed RCT	63	• Medication adherence (pharmacy refill data) • Proportion of missed visits	Moderate
Jarvis et al (2019) [62]	Two-armed RCT	32	• Disconnection and Rejection domains of the Young Schema Questionnaire • de Jong Gierveld Loneliness Scale • World Health Organization-Five Well-Being Index	Weak
Joseph et al (2015) [63]	Two-armed RCT	29	• Sedentary behavior • PA	Moderate
Kamal et al (2015) [64]	Two-armed RCT	200	• Morisky Medication Adherence Scale	Moderate
King et al (2013) [65]	Two-armed RCT	40	• Community Health Activities Model Program for Seniors questionnaire • Daily steps	Moderate
Kiropoulos et al (2011) [66]	Two-armed RCT	202	• Depression literacy scores • Depression Stigma Scale • BDI-II	Weak
Lee et al (2014) [67] and Lee et al (2016) [68]	One-group pretest-posttest	30	• Adapted 15-item scale of Taylor et al [92] • Intent (investigator developed questionnaire) • Actual vaccination or test	Weak
Lee et al (2017) [69]	Two-armed RCT	131	• Completed mammograms	Weak
MacDonell et al (2016) [70]	Two-armed RCT	49	• Medication adherence • Asthma control	Moderate
Marcus et al (2016) [71]	Two-armed RCT	205	• Increased minutes/week of moderate to vigorous PA • PA by accelerometer	Moderate
Mauriello et al (2016) [72]	Two-armed RCT	335	• Self-reported behavior risks • Daily fruit and vegetable consumption • Daily minutes of stress management activity	Weak
Miller et al (2018) [73]	Two-armed RCT	450	• Completed screening	Moderate
Moussa et al (2013) [74]	Two-armed RCT	45	• Literacy Assessment for Diabetes	Moderate
Neafsey et al (2011) [75]	Two-armed RCT	160	• The Adverse Self-Medication Behavior Risk Score	Moderate
Nelson et al (2016) [76]	One-group pretest-posttest	80	• Diabetes Self-Care Activities Medication subscale	Weak
Neuenschwander et al (2013) [77]	Two-armed RCT	123	• 16-item questionnaire for low-income population for nutrition-related behavior outcomes	Moderate
Phelan et al (2017) [78]	Two-armed RCT	371	• Body weight	Moderate
Rubinstein et al (2016) [79]	Two-armed RCT	637	• Blood pressure	Moderate

Authors (year)	Study designs	Sample sizes	Outcome measures	Quality ratings
Ryan et al (2013) [80]	One-group pretest-posttest	24	• HbA_{1c} • Cholesterol	Weak
Steinberg et al (2013) [81]	Two-armed RCT	50	• Body weight	Moderate
Tessaro et al (2007) [82]	Two-armed RCT	395	• 34-item food frequency checklist • Dietary knowledge	Moderate
Titov et al (2015) [83][e]	Two-armed RCT	54	• PHQ-9 • GAD-7	Weak
Ünlü Ince et al (2013) [84]	Two-armed RCT	96	• CES-D	Weak
Wahbeh et al (2016) [85]	Two-armed RCT	20	• CES-D • Five-Facet Mindfulness Questionnaire • Positive and Negative Affect Schedule	Moderate
Wayne et al (2015) [86]	Two-armed RCT	97	• HbA_{1c}	Weak
Weinert et al (2008) [87][h]	Three-armed RCT	176	• Health knowledge score (investigator developed questionnaire)	Weak

[a]RCT: randomized controlled trial.

[b]BDI-II: Beck Depression Inventory II.

[c]HbA_{1c}: hemoglobin A1c.

[d]CES-D: Center for Epidemiological Studies Depression.

[e]Adaptations of a similar intervention.

[f]PHQ-9: Patient Health Questionnaire nine-item.

[g]GAD-7: Generalized Anxiety Disorder seven-item scale.

[h]Same intervention but different cohorts.

[i]PA: physical activity.

General Characteristics of the Interventions

Among the 48 interventions, 32 were from the United States, 4 from Australia, 3 from Canada, 2 from the Netherlands and South Africa, and 1 each from China, Malawi, Pakistan, and Taiwan, whereas 1 intervention was undertaken across 3 South American countries, namely, Argentina, Guatemala, and Peru. Low-income groups were the most common target group (n=20), followed by ethnic minorities (n=18), older adults (n=10), and rural communities (n=8). Low-literacy groups were targeted in 2 interventions [45,73]. A wide range of health issues were addressed among the 48 interventions, with diabetes being the most common (n=8), followed by 6 targeting physical inactivity and 5 targeting depression (Multimedia Appendix 3 [37-87]).

Websites were the most commonly used platforms, with 10 interventions using websites only and 12 interventions combining websites with other platforms such as email or text messaging. A total of 11 studies used text messaging alone, and 4 combined this with other platforms. A total of 10 interventions employed mobile apps. Facebook was used in 2 interventions, and WhatsApp was used in 1 intervention. Mobile phones were the most popular device, being used in 26 interventions, followed by the computer in 22 interventions. Tablets were used in 6 interventions.

Among the 48 interventions, 37 were interactive, providing information, tailored content, and/or health-engaging activities, and 11 were noninteractive, providing information or reminder text messages only. The duration of interventions ranged from one 30-min session to a 13-month program, with 3 months being the most common duration.

Use of eHealth Literacy

No interventions explicitly reported that eHealth literacy needs were considered during the development, and no assessment of eHealth literacy was undertaken. In fact, eHealth literacy was only mentioned in a study by Carroll et al [45], which included eHealth literacy as one of the secondary outcome measures and used the eHealth Literacy Scale [93] for assessment. Apart from eHealth literacy, 4 interventions undertook other literacy assessments. Ingersoll et al assessed functional English literacy by using the Wide Range Achievement Test 4 [61], and health literacy was assessed in 3 interventions using different measures, including the Short Test of Functional Health Literacy in Adults [71], the Rapid Estimate of Adults Literacy in Medicine [75], or a single question [73]. All such assessments were conducted at baseline with no discussion as to whether baseline assessment played any role in intervention development (Table 3).

Table 3. The role of electronic health literacy and users in intervention development.

Authors (year)	Developmental frameworks	eHealth literacy or other literacy assessment or application of user-centered principles or user involvement
Agyapong et al (2017) [37]	• Cognitive behavioral therapy principles	• Content written in partnership with patients
Anand et al (2016) [38]	• Integrative behavioral modification strategy • Social cognitive social learning theories • Goal setting theory • Transtheoretical model	• Pilot study
Arora et al (2014) [39]	• Not reported	• Not reported
Bennett et al (2018) [40]	• Social cognitive theory • Interactive obesity treatment approach	• Not reported
Bond et al (2010) [41]	• Not reported	• Not reported
Broekhuizen et al (2016) [42] and Wijsman et al (2013) [43]	• Transtheoretical model • I-Change model	• Not reported
Buller et al (2008) [44]	• Social cognitive theory • Diffusion of innovations model	• Focus groups • Usability testing
Carroll et al (2019) [45]	• Capability, opportunity, motivation, and behavior model for behavior change • Community-based participatory research	• eHealth Literacy Scale used to measure eHealth literacy as one of the secondary outcomes • Participatory research involving users
Caster et al (2017) [46]	• Not reported	• Focus groups
Chen et al (2016) [47]	• Not reported	• Not reported
Chen et al (2018) [48]	• Not reported	• Not reported
Choi et al (2012) [49][a]	• Adaptation of the sadness internet-delivered cognitive behavioral therapy program	• Not reported
Dang et al (2017) [50]	• Not reported	• Not reported
Dear et al (2015) [51][a]	• Previous studies	• Not reported
Dugas et al (2018) [52]	• Not reported	• Not reported
Fortmann et al (2017) [53]	• Not reported	• Not reported
Gilmore et al (2017) [54]	• Learning theory • Theory of planned behavior • Theory of reasoned actions • Social cognitive theory	• Not reported
Griffin et al (2018) [55]	• Social cognitive theory	• Not reported
Hacking et al (2016) [56]	• Social cognitive theory	• Not reported
Hageman et al (2014) [57]	• Pender's Health Promotion Model based on social cognitive theory	• Not reported
Herring et al (2017) [58]	• Social cognitive theory • Social ecological model	• Focus groups • Semistructured interviews
Hill et al (2006) [59] and Weinert et al (2008) [87]	• Not reported	• Pilot study
Hong et al (2015) [60]	• Theory of goal setting	• Usability testing

Authors (year)	Developmental frameworks	eHealth literacy or other literacy assessment or application of user-centered principles or user involvement
Ingersoll et al (2015) [61]	• IMB[b] model of adherence • Social action theory	• Functional English literacy assessed by Wide Range Achievement Test 4 • Focus groups • Interviews • Usability testing
Jarvis et al (2019) [62]	• Theoretical framework of loneliness • Literature review • Developed by a cognitive behavioral therapy specialist psychologist, a mental health nurse, and a mobile health expert	• Not reported
Joseph et al (2015) [63]	• Social cognitive theory	• Not reported
Kamal et al (2015) [64]	• The health belief model • Social cognitive theory • Michie's taxonomy of behavioral change	• Not reported
King et al (2013) [65]	• Social cognitive theory • Transtheoretical model	• Participatory formative research
Kiropoulos et al (2011) [66]	• Not reported	• Not reported
Lee et al (2014) [67] and Lee et al (2016) [68]	• Fogg behavioral model	• Community advisory group • Focus groups • Usability testing
Lee et al (2017) [69]	• Fogg behavioral model • Health belief model • Concept of persuasive technology	• Community advisory group • Focus groups • Usability testing
MacDonell et al (2016) [70]	• Principles of motivational interviewing • IMB skills model	• Pilot testing
Marcus et al (2016) [71]	• Social cognitive theory • Transtheoretical model	• Health literacy assessed by the Short Test of Functional Health Literacy • Focus groups
Mauriello et al (2016) [72]	• Transtheoretical model of behavior change	• Usability testing
Miller et al (2018) [73]	• Not reported	• Health literacy assessed by asking a single question, "how confident are you filling out medical forms by yourself?" • Pilot testing
Moussa et al [74]	• Not reported	• Not reported
Neafsey et al (2011) [75]	• Social cognitive theory	• Health literacy assessed by Rapid Estimate of Adult Literacy in Medicine • Usability testing • Pilot testing
Nelson et al (2016) [76]	• Adapted from the SuperEgo mobile communications platform	• Usability testing
Neuenschwander et al (2013) [77]	• Kolb's learning styles and experiential learning model • Use of the US Department of Health and Human Services' Research-based Web Design and Usability Guidelines • Previous users' needs and requests	• Pilot testing
Phelan et al (2017) [78]		• Not reported

Authors (year)	Developmental frameworks	eHealth literacy or other literacy assessment or application of user-centered principles or user involvement
	• Social cognitive theory • Based on the diabetes prevention program and Look Ahead lifestyle interventions	
Rubinstein et al (2016) [79]	• Transtheoretical model • Health belief model	• Focus groups • Pilot study
Ryan et al (2013) [80]	• Social cognitive theory	• Not reported
Steinberg et al (2013) [81]	• Interactive obesity treatment approach	• Not reported
Tessaro et al (2007) [82]	• Health belief model • Social learning theory • Social support theory	• Focus groups
Titov et al (2015) [83][a]	• Psychological principles	• Not reported
Ünlü Ince et al (2013) [84]	• Not reported	• Not reported
Wahbeh et al [85]	• Modification of the mindfulness-based cognitive therapy and mindfulness-based stress reduction	• Not reported
Wayne et al (2015) [86]	• Motivational interviewing • Cognitive behavioral therapy	• Pilot study

[a]Adaptations of similar programs.

[b]IMB: Information, Motivation and Behavior Skills

Use of Developmental Framework

Theoretical frameworks were the most used guidelines for developing interventions, with social cognitive theories (n=15) and the transtheoretical model (n=6) most commonly used. A total of 7 interventions were adaptations or modifications of previous programs, whereas 13 interventions did not provide any details about their theoretical frameworks or developmental frameworks. Only 1 intervention was reported on using the *Research-based Web Design and Usability Guidelines* developed by the US Department of Health and Human Services (UDHHS) [94] to inform the creation of their intervention website (Table 3) [77].

User Involvement

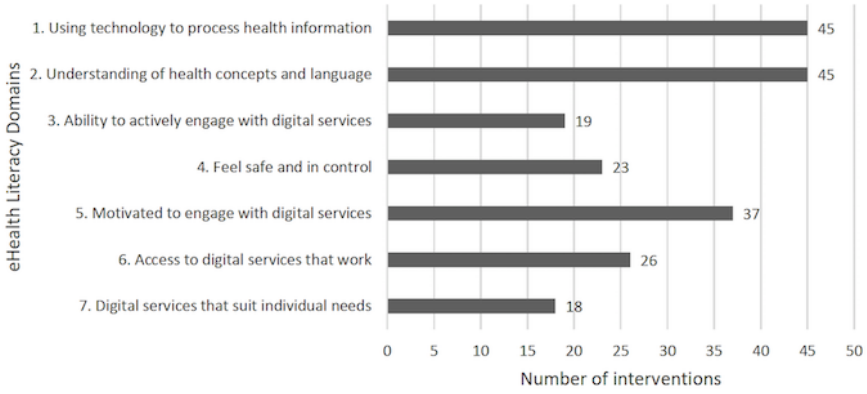
User-centered principles were not discussed in any of the papers. The development of interventions in the included studies was mostly expert driven. A total of 22 interventions reported on involving users during development, with needs assessments using focus groups or interviews in 8 interventions, usability or pilot testing in 15 interventions, and 2 interventions mentioning participatory formative research with no details provided. Only Agyapong et al [37] reported that patients were involved in content writing (Table 3).

Addressing eHealth Literacy Domains

Most reviewed interventions did not address all eHealth literacy domains. The number of domains addressed ranged from 1 to 7, with only 1 study containing components addressing all 7 eHealth literacy domains [65]. A total of 7 studies representing 6 interventions contained components of 6 domains [57,59,64,71,78,80,87], whereas 20 studies contained components addressing 5 domains (Multimedia Appendix 4 [37-87]).

The 2 most addressed eHealth literacy domains, *1. Using technology to process health information* and *2.Understanding of health concepts and language*, were both identified in 45 interventions. The domain 5. *Motivated to engage with digital services* was addressed in 37 interventions through different strategies to encourage users to engage with interventions. A total of 26 interventions provided access to hardware, data plans, or technical support to address the domain *6. Access to digital services that work*, whereas 23 interventions supported the domain *4. Feel safe and in control* by requiring personal log-in or other forms of privacy measures. The 2 most overlooked domains were *3. Ability to actively engage with digital services* and *7. Digital services that suit individual needs*; both were identified in less than half of the 48 interventions. Of the 19 interventions containing components of *3. Ability to actively engage with digital services*, 15 provided training or instructions on using the intervention, whereas only 6 featured an easy-to-use navigation interface. Among the 18 programs addressing the domain *7. Digital services that suit individual needs*, the main strategy was to provide the preferred language of users. Accessibility features catering to individual capability or providing a user interface that suited individual needs such as large fonts or icons or audio options were only identified in 6 interventions [64,65,73-75,82]. Figure 2 shows the number of interventions addressing each of the 7 domains.

Figure 2. Number of interventions addressing the seven domains of the eHealth Literacy Framework [23, 25].



Effectiveness of eHealth Interventions

Although no study explicitly considered or assessed eHealth literacy, the effectiveness of eHealth interventions was nevertheless examined, and the results were mixed. Approximately one-fourth of the reviewed interventions (n=13) did not find statistically significant improvements in their primary outcomes. Although 19 studies reported significant improvements in their primary outcomes, another 12 studies found significant improvements in some primary outcomes, but not all. In addition, 4 studies found improvements but did not report whether such differences were significant. The results of long-term effectiveness also produced mixed evidence. Among the 13 studies that conducted follow-up assessments, 8 found the effects sustainable up to a period of 12 months. A total of 3 studies found effects were not sustainable, whereas 2 did not report on the significance (Multimedia Appendix 3).

Among the reported effective 19 studies, there were no consistent patterns of intervention characteristics or eHealth literacy domains likely addressed. These interventions could be interactive or noninteractive, although platforms and devices also varied. The number of eHealth literacy domains likely addressed ranged from 1 to 7. Although a study by King et al [65] likely addressed all 7 domains and found their intervention effective with a large effect size (0.8-1.2), a study by Chen et al [48] also reported their study as effective, although only the domain *1. Using technology to process health information* could be identified within the intervention components.

Discussion

Principal Findings

Although the concept of eHealth literacy was introduced more than a decade ago [17], this review finds that utilization of the concept for enhancing eHealth use and engagement is rarely recognized. The eHealth literacy needs of users were not explicitly considered during intervention development in any of the included studies, and no eHealth literacy assessment was conducted to ensure that such needs were met. This result is echoed in an earlier systematic review of eHealth and telehealth tools for vulnerable populations, which reported that eHealth literacy was not assessed in any of the 18 included studies [95]. In fact, eHealth literacy is only mentioned in 1 of the 51 papers of this review. Although 3 studies conducted health literacy assessments at baseline [71,73,75] and 1 assessed eHealth literacy as a secondary outcome [45], the results were not used for intervention development. The fact that eHealth literacy is overlooked may be because of the lack of comprehensive measures before 2018 and an associated knowledge gap in using such assessment to inform eHealth design [96]. To move forward, in-depth research on eHealth literacy is required, such as the application of the eHLF and the recently developed comprehensive eHLQ, designed to support eHealth intervention development and evaluation. Developed on the basis of eHLF, the eHLQ is a 35-item questionnaire that produces 7 scores representing 7 eHealth literacy domains of users. The resulting scores provide insights into users' strengths and weaknesses in using eHealth such that interventions can be tailored accordingly. For example, if target users reported good ability to use technology (higher scores in *3. Ability to actively engage with digital services*) but lack motivation (lower scores in *5. Motivated to engage with digital services*), features to address motivation should be a prominent feature of the intervention. However, if target users demonstrated limited ability to use technologies, interventions such as simple unidirectional text messages, rather than interactive mobile apps, are likely to be more suitable for the target users. Hence, the eHLF and eHLQ will have the potential to advance the field of eHealth literacy and strengthen the reach and impact of digital health interventions [25].

Addressing eHealth Literacy Needs

There is growing concern that frameworks or guidelines informing the development of eHealth interventions so that they meet users' needs are lacking [97], and this concern is reflected in the findings of this review. Only 1 study [77] used the *Research-based Web Design and Usability Guidelines* by the UDHHS to inform intervention development. However, the guideline authors specifically indicate that they may not be applicable to all audiences, such as people with low literacy who may have different reading and layout needs [94]. In addition, only 22 studies in this review discussed user involvement whereas needs assessments were usually in the form of focus groups or interviews involving a limited number of users. Only 1 study reported the inclusion of patients in content writing [37]. Such practice means that interventions are expert driven instead of user driven, echoing the concern that users and patients are the most underused resources in developing eHealth interventions [98]. Although eHealth literacy

is only one of the factors in developing effective eHealth strategies, it has been advocated that it is a primary and critical factor that affects usability and adoption [17,21,26]. Even if an intervention is grounded in theory, it will not be usable if it does not align with the literacy needs and abilities of end users and may lead to nonadoption [30,99,100]. Hence, research efforts into eHealth developmental frameworks incorporating eHealth literacy need assessment, and user-centered principles are required such that equal access and usage can be achieved for all users.

Although eHealth literacy needs may not be explicitly considered when developing eHealth programs, this review still finds that interventions generally have features that may meet eHealth literacy needs based on the eHLF. However, the common focus is on providing information or features that address the domains of *1. Using technology to process health information*, *2. Understanding of health concepts and language*, and *5. Motivated to engage with digital services*. Strategies to assist users in using or engaging with technology and accessibility features of systems that are tailored and responsive to an individual's ability and capability are generally overlooked. These findings resonate with those from a systematic review of diabetes apps targeted at older adults that there is a limited variety of accessibility features [101]. This is of special importance when an intervention is designed for older people or people with disabilities who may require specialized tools because of functional and cognitive impairments [102,103] or people with low literacy skills who may have different reading and design needs [94]. In addition, applying the eHLF to determine whether certain eHealth literacy domains were addressed may not necessarily mean that the eHealth literacy needs of users were met as the actual eHealth literacy needs of target users were not assessed and, therefore, not known. The results highlight that in developing interventions using technologies, designers are mainly responsible for ensuring that users' needs and capabilities are met in the hope that users will adopt the intervention to improve or change their health behavior. However, Chang et al [104] noted that eHealth intervention designers were typically not trained to meet the communication needs of underserved communities. Showell et al [105] also pointed out that eHealth systems tended to be designed for users who were similar to the designers, who were usually middle-class professionals. As such, the needs of disadvantaged patients were generally overlooked in the design process [15].

Effectiveness of eHealth Interventions

In addition to exploring the role of eHealth literacy and eHealth intervention development, this review also examined the effectiveness of eHealth interventions targeted at socially disadvantaged groups and found inconclusive evidence. Although significant improvements were found in 19 studies, these findings should be interpreted with caution, as 10 studies are of weak quality and 7 studies are of moderate quality. Although 3 studies reported a large effect size, they had smaller sample sizes and were of moderate or weak quality [51,65,83]. The sustainability of effects is also mixed and cannot be ascertained, as most studies have short follow-up times. These findings are similar to reviews of eHealth interventions, which also report inconclusive evidence on effectiveness [35,103,106,107]. The lack of comprehensive eHealth literacy assessments also prevents this review from exploring the link between eHealth literacy and the effectiveness of eHealth programs. Further robust empirical studies need to be undertaken to better understand the role of eHealth literacy in eHealth interventions to help address the digital divide and improve health disparities.

Limitations

Several limitations of this review need to be acknowledged. Only peer-reviewed journals were included for this review, and there may be other studies that were not accounted for. The search was conducted by one researcher, which may have led to potential bias. The findings of this review may not reflect all details of the actual intervention, as authors generally only briefly describe their intervention development processes [108], and few studies report how users are involved such that interventions are aligned with their needs [27]. However, not reporting certain features suggests that the authors may not consider such features as relevant. Furthermore, the included studies do not represent all socially disadvantaged groups. This review only focused on certain categories of disadvantaged groups and did not include other underserved populations, such as people with disabilities or indigenous people who may also have limited access or skills to use ICT [109]. Future reviews should consider inclusion of these groups to advance eHealth research among vulnerable populations.

Conclusions

The WHO recognizes health literacy as a critical determinant of health that has the potential to empower individuals and bring about health equity [19]. However, this systematic review finds that the role of eHealth literacy in designing eHealth interventions targeted at socially disadvantaged groups is generally overlooked. eHealth literacy was not explicitly considered or assessed during intervention development. There was also a lack of frameworks or theories informing eHealth designers on how to meet users' needs. Although users were involved in some of the reviewed studies, intervention development was mainly expert driven rather than user driven. By using the eHLF to examine the eHealth literacy components of eHealth interventions, it was found that the design of features such that they suited individual capability was not common. Furthermore, whether the eHealth literacy needs of users were actually addressed in the reviewed interventions cannot be ascertained because of the lack of comprehensive eHealth literacy assessment. The link between eHealth literacy and effectiveness of eHealth interventions cannot be explored. Moreover, the paucity of robust studies also delivers limited empirical evidence on how to effectively reach these vulnerable populations and bridge the digital divide.

Despite the concept of eHealth literacy being introduced in 2006, its potential role in empowering individuals has not been realized. Without meeting the eHealth literacy needs of disadvantaged groups, adoption of eHealth interventions is likely to be low, resulting in ineffective interventions [17,21,30,99,100,110]. To ensure that no one is left behind as determined in the Shanghai Declaration on Promoting Health

[19], eHealth literacy must be acknowledged and included in the development of eHealth interventions to assist the realization of technological advancement and improve health equity.

Acknowledgments

The authors would like to thank Associate Professor Lars Kayser, University of Copenhagen; Assistant Professor Sarah Fredsted Villadsen, University of Copenhagen; and Dr Nancy Huang, Chief Medical Advisor, Australian Unity for their contribution during the conceptualization stage of this systematic review. The authors would also like to thank Dr Polina Putrik, Visiting Fellow at Deakin University, for her input in search strategy. RO was funded in part through the National Health and Medical Research Council of Australia Principal Research Fellowship #APP1155125.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File , 20 KB - jmir_v22i8e18476_app1.docx]

Multimedia Appendix 2
Quality assessment.
[DOCX File , 27 KB - jmir_v22i8e18476_app2.docx]

Multimedia Appendix 3
Intervention characteristics and key findings.
[DOCX File , 50 KB - jmir_v22i8e18476_app3.docx]

Multimedia Appendix 4
Summary of eHealth literacy domains likely addressed.
[DOCX File , 54 KB - jmir_v22i8e18476_app4.docx]

References

1. World Health Organization. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. Geneva, Switzerland: World Health Organization; 2019.
2. Digital Health. World Health Organization. 2018. URL: http://apps.who.int/gb/ebwha/pdf_files/WHA71/A71_R7-en.pdf?ua=1 [accessed 2020-03-19]
3. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
4. Litt E. Measuring users' internet skills: a review of past assessments and a look toward the future. New Media Soc 2013 May 24;15(4):612-630. [doi: 10.1177/1461444813475424] [Medline: 87775130]
5. Falling Through the Net: Defining the Digital Divide, A Report on the Telecommunications and Information Technology Gap in America. Washington, DC: US Department of Commerce; 1999.
6. Wei K, Teo H, Chan H, Tan B. Conceptualizing and testing a social cognitive model of the digital divide. Inf Syst Res 2011 Mar 22;22(1):170-187 [FREE Full text] [doi: 10.1287/isre.1090.0273]
7. Ginossar T. Predictors of online cancer prevention information seeking among patients and caregivers across the digital divide: a cross-sectional, correlational study. JMIR Cancer 2016 Mar 9;2(1):e2 [FREE Full text] [doi: 10.2196/cancer.5108] [Medline: 28410177]
8. Lustria ML, Smith SA, Hinnant CC. Exploring digital divides: an examination of eHealth technology use in health information seeking, communication and personal health information management in the USA. Health Informatics J 2011 Sep;17(3):224-243. [doi: 10.1177/1460458211414843] [Medline: 21937464]
9. Asoh DA, Rivers PA. The empowerment and quality health value propositions of e-health. Health Serv Manage Res 2010 Nov;23(4):181-184. [doi: 10.1258/hsmr.2010.010007] [Medline: 21097729]
10. Viswanath K, Kreuter MW. Health disparities, communication inequalities, and eHealth. Am J Prev Med 2007 May;32(5 Suppl):S131-S133 [FREE Full text] [doi: 10.1016/j.amepre.2007.02.012] [Medline: 17466818]
11. Baur C. An analysis of factors underlying e-health disparities. Camb Q Healthc Ethics 2008;17(4):417-428. [doi: 10.1017/S0963180108080547] [Medline: 18724881]

12. Kontos E, Blake KD, Chou WY, Prestin A. Predictors of eHealth usage: insights on the digital divide from the health information national trends survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
13. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/internet use. J Med Internet Res 2013 May 2;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
14. Kim E, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Challenges to using an electronic personal health record by a low-income elderly population. J Med Internet Res 2009 Oct 27;11(4):e44 [FREE Full text] [doi: 10.2196/jmir.1256] [Medline: 19861298]
15. Showell C, Cummings E, Turner P. The invisibility of disadvantage: why do we not notice? Stud Health Technol Inform 2017;235:388-392. [Medline: 28423820]
16. Hasnain-Wynia R, Wolf MS. Promoting health care equity: is health literacy a missing link? Health Serv Res 2010 Aug;45(4):897-903 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01134.x] [Medline: 20646073]
17. Norman CD, Skinner HA. Ehealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
18. Austin R. EHealth Literacy for Older Adults - Part II. EBSCOhost Connection. 2012. URL: http://connection.ebscohost.com/c/articles/79632384/ehealth-literacy-older-adults-part-ii [accessed 2020-07-28]
19. World Health Organization. Shanghai declaration on promoting health in the 2030 agenda for sustainable development. Health Promot Int 2017 Feb 1;32(1):7-8. [doi: 10.1093/heapro/daw103] [Medline: 28180270]
20. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
21. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the effectiveness of consumer-focused health information technology systems through ehealth literacy: a framework for understanding users' needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
22. Monkman H, Griffith J, Kushniruk A. Evidence-based heuristics for evaluating demands on ehealth literacy and usability in a mobile consumer health application. Stud Health Technol Inform 2015;216:358-362. [Medline: 26262071]
23. Norgaard O, Furstand D, Klokker L, Karnoe A, Batterham R, Kayser L. The e-health literacy framework: a conceptual framework for characterizing e-health users and their interaction with e-health systems. Know Manag E-Learn 2015;7(4):522-540. [doi: 10.34105/j.kmel.2015.07.035]
24. Buchbinder R, Batterham B, Elsworth G, Dionne CE, Irvin E, Osborne RH. A validity-driven approach to the understanding of the personal and societal burden of low back pain: development of a conceptual and measurement model. Arthritis Res Ther 2011;13(5):R152 [FREE Full text] [doi: 10.1186/ar3468] [Medline: 21933393]
25. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the ehealth literacy framework: development and initial validity testing of the ehealth literacy questionnaire (EHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
26. Monkman H, Kushniruk A. eHealth literacy issues, constructs, models, and methods for health information technology design and evaluation. Know Manag E-Learn 2015;7(4):541-549. [doi: 10.34105/j.kmel.2015.07.036]
27. de vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
28. LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. Int J Med Inform 2013 Nov;82(11):e251-e268. [doi: 10.1016/j.ijmedinf.2011.03.006] [Medline: 21481635]
29. Roehrer E, Cummings E, Ellis L, Turner P. The role of user-centred design within online community development. Stud Health Technol Inform 2011;164:256-260. [Medline: 21335720]
30. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. Br Med J 2010 Nov 16;341:c5814 [FREE Full text] [doi: 10.1136/bmj.c5814] [Medline: 21081595]
31. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 1;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
32. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007 Jun 15;7:16 [FREE Full text] [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]
33. Physical Activity Promotion In Socially Disadvantaged Groups: Principles for Action. WHO/Europe. 2013. URL: http://www.euro.who.int/__data/assets/pdf_file/0006/193092/PHAN-brochure_ENG.pdf [accessed 2020-11-03]
34. World Population Prospects: The 2017 Revision: Key Findings and Advance Tables. United Nations Population Division. 2017. URL: https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_KeyFindings.pdf [accessed 2020-11-03]

35. Bennett GG, Steinberg DM, Stoute C, Lanpher M, Lane I, Askew S, et al. Electronic health (eHealth) interventions for weight management among racial/ethnic minority adults: a systematic review. Obes Rev 2014 Oct;15(Suppl 4):146-158. [doi: 10.1111/obr.12218] [Medline: 25196411]
36. Quality Assessment Tool for Quantitative Studies. Canadian Health Care. 2009. URL: https://www.ephpp.ca/quality-assessment-tool-for-quantitative-studies/ [accessed 2017-12-03]
37. Agyapong VI, Juhás M, Ohinmaa A, Omeje J, Mrklas K, Suen VY, et al. Randomized controlled pilot trial of supportive text messages for patients with depression. BMC Psychiatry 2017 Aug 2;17(1):286 [FREE Full text] [doi: 10.1186/s12888-017-1448-2] [Medline: 28768493]
38. Anand SS, Samaan Z, Middleton C, Irvine J, Desai D, Schulze KM, South Asian Heart Risk Assessment Investigators. A digital health intervention to lower cardiovascular risk: a randomized clinical trial. JAMA Cardiol 2016 Aug 1;1(5):601-606. [doi: 10.1001/jamacardio.2016.1035] [Medline: 27438754]
39. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014 Jun;63(6):745-54.e6. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
40. Bennett GG, Steinberg D, Askew S, Levine E, Foley P, Batch BC, et al. Effectiveness of an app and provider counseling for obesity treatment in primary care. Am J Prev Med 2018 Dec;55(6):777-786 [FREE Full text] [doi: 10.1016/j.amepre.2018.07.005] [Medline: 30361140]
41. Bond GE, Burr RL, Wolf FM, Feldt K. The effects of a web-based intervention on psychosocial well-being among adults aged 60 and older with diabetes: a randomized trial. Diabetes Educ 2010;36(3):446-456. [doi: 10.1177/0145721710366758] [Medline: 20375351]
42. Broekhuizen K, de Gelder J, Wijsman CA, Wijsman LW, Westendorp RG, Verhagen E, et al. An internet-based physical activity intervention to improve quality of life of inactive older adults: a randomized controlled trial. J Med Internet Res 2016 Apr 27;18(4):e74 [FREE Full text] [doi: 10.2196/jmir.4335] [Medline: 27122359]
43. Wijsman CA, Westendorp RG, Verhagen EA, Catt M, Slagboom PE, de Craen AJ, et al. Effects of a web-based intervention on physical activity and metabolism in older adults: randomized controlled trial. J Med Internet Res 2013 Nov 6;15(11):e233 [FREE Full text] [doi: 10.2196/jmir.2843] [Medline: 24195965]
44. Buller DB, Woodall WG, Zimmerman DE, Slater MD, Heimendinger J, Waters E, et al. Randomized trial on the 5 a day, the Rio Grande Way Website, a web-based program to improve fruit and vegetable consumption in rural communities. J Health Commun 2008;13(3):230-249 [FREE Full text] [doi: 10.1080/10810730801985285] [Medline: 18569356]
45. Carroll JK, Tobin JN, Luque A, Farah S, Sanders M, Cassells A, et al. 'Get ready and empowered about rreatment' (GREAT) study: a pragmatic randomized controlled trial of activation in persons living with HIV. J Gen Intern Med 2019 Sep;34(9):1782-1789. [doi: 10.1007/s11606-019-05102-7] [Medline: 31240605]
46. Caster MM, Norris AH, Butao C, Carr Reese P, Chemey E, Phuka J, et al. Assessing the acceptability, feasibility, and effectiveness of a tablet-based cervical cancer educational intervention. J Cancer Educ 2017 Mar;32(1):35-42 [FREE Full text] [doi: 10.1007/s13187-015-0953-6] [Medline: 26637473]
47. Chen YX, Hung YP, Chen HC. Mobile application-assisted cognitive behavioral therapy for insomnia in an older adult. Telemed J E Health 2016 Apr;22(4):332-334. [doi: 10.1089/tmj.2015.0064] [Medline: 26393902]
48. Chen T, Zhu W, Tang B, Jin L, Fu H, Chen Y, et al. A mobile phone informational reminder to improve eye care adherence among diabetic patients in rural China: a randomized controlled trial. Am J Ophthalmol 2018 Oct;194:54-62. [doi: 10.1016/j.ajo.2018.07.006] [Medline: 30053472]
49. Choi I, Zou J, Titov N, Dear BF, Li S, Johnston L, et al. Culturally attuned Internet treatment for depression amongst Chinese Australians: a randomised controlled trial. J Affect Disord 2012 Feb;136(3):459-468. [doi: 10.1016/j.jad.2011.11.003] [Medline: 22177742]
50. Dang S, Karanam C, Gómez-Marín O. Outcomes of a mobile phone intervention for heart failure in a minority county hospital population. Telemed J E Health 2017 Jun;23(6):473-484. [doi: 10.1089/tmj.2016.0211] [Medline: 28051357]
51. Dear BF, Zou JB, Ali S, Lorian CN, Johnston L, Sheehan J, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of anxiety: a randomized controlled trial. Behav Ther 2015 Mar;46(2):206-217. [doi: 10.1016/j.beth.2014.09.007] [Medline: 25645169]
52. Dugas M, Crowley K, Gao GG, Xu T, Agarwal R, Kruglanski AW, et al. Individual differences in regulatory mode moderate the effectiveness of a pilot mHealth trial for diabetes management among older veterans. PLoS One 2018;13(3):e0192807 [FREE Full text] [doi: 10.1371/journal.pone.0192807] [Medline: 29513683]
53. Fortmann AL, Gallo LC, Garcia MI, Taleb M, Euyoque JA, Clark T, et al. Dulce digital: an mhealth SMS-based intervention improves glycemic control in Hispanics with type 2 diabetes. Diabetes Care 2017 Oct;40(10):1349-1355 [FREE Full text] [doi: 10.2337/dc17-0230] [Medline: 28600309]
54. Gilmore LA, Klempel MC, Martin CK, Myers CA, Burton JH, Sutton EF, et al. Personalized mobile health intervention for health and weight loss in postpartum women receiving women, infants, and children benefit: a randomized controlled pilot study. J Womens Health (Larchmt) 2017 Jul;26(7):719-727 [FREE Full text] [doi: 10.1089/jwh.2016.5947] [Medline: 28338403]

55. Griffin JB, Struempler B, Funderburk K, Parmer SM, Tran C, Wadsworth DD. My quest, an intervention using text messaging to improve dietary and physical activity behaviors and promote weight loss in low-income women. J Nutr Educ Behav 2018 Jan;50(1):11-8.e1. [doi: 10.1016/j.jneb.2017.09.007] [Medline: 29325657]
56. Hacking D, Haricharan HJ, Brittain K, Lau YK, Cassidy T, Heap M. Hypertension health promotion via text messaging at a community health center in South Africa: a mixed methods study. JMIR Mhealth Uhealth 2016 Mar 10;4(1):e22 [FREE Full text] [doi: 10.2196/mhealth.4569] [Medline: 26964505]
57. Hageman PA, Pullen CH, Hertzog M, Boeckner LS. Effectiveness of tailored lifestyle interventions, using web-based and print-mail, for reducing blood pressure among rural women with prehypertension: main results of the wellness for women: dashing towards health clinical trial. Int J Behav Nutr Phys Act 2014 Dec 6;11:148 [FREE Full text] [doi: 10.1186/s12966-014-0148-2] [Medline: 25480461]
58. Herring SJ, Cruice JF, Bennett GG, Darden N, Wallen JJ, Rose MZ, et al. Intervening during and after pregnancy to prevent weight retention among African American women. Prev Med Rep 2017 Sep;7:119-123 [FREE Full text] [doi: 10.1016/j.pmedr.2017.05.015] [Medline: 28660118]
59. Hill W, Weinert C, Cudney S. Influence of a computer intervention on the psychological status of chronically ill rural women: preliminary results. Nurs Res 2006;55(1):34-42 [FREE Full text] [doi: 10.1097/00006199-200601000-00005] [Medline: 16439927]
60. Hong YA, Goldberg D, Ory MG, Towne SD, Forjuoh SN, Kellstedt D, et al. Efficacy of a mobile-enabled web app (iCanFit) in promoting physical activity among older cancer survivors: a pilot study. JMIR Cancer 2015 Jun 26;1(1):e7 [FREE Full text] [doi: 10.2196/cancer.4389] [Medline: 28410158]
61. Ingersoll KS, Dillingham RA, Hettema JE, Conaway M, Freeman J, Reynolds G, et al. Pilot RCT of bidirectional text messaging for ART adherence among nonurban substance users with HIV. Health Psychol 2015 Dec;34S:1305-1315 [FREE Full text] [doi: 10.1037/hea0000295] [Medline: 26651472]
62. Jarvis MA, Padmanabhanunni A, Chipps J. An evaluation of a low-intensity cognitive behavioral therapy mhealth-supported intervention to reduce loneliness in older people. Int J Environ Res Public Health 2019 Apr 11;16(7):1-14 [FREE Full text] [doi: 10.3390/ijerph16071305] [Medline: 30979042]
63. Joseph RP, Keller C, Adams MA, Ainsworth BE. Print versus a culturally-relevant Facebook and text message delivered intervention to promote physical activity in African American women: a randomized pilot trial. BMC Womens Health 2015 Mar 27;15:30 [FREE Full text] [doi: 10.1186/s12905-015-0186-1] [Medline: 25886945]
64. Kamal AK, Shaikh Q, Pasha O, Azam I, Islam M, Memon AA, et al. A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored short messaging service (SMS)-SMS4Stroke study. BMC Neurol 2015 Oct 21;15:212 [FREE Full text] [doi: 10.1186/s12883-015-0471-5] [Medline: 26486857]
65. King AC, Bickmore TW, Campero MI, Pruitt LA, Yin JL. Employing virtual advisors in preventive care for underserved communities: results from the COMPASS study. J Health Commun 2013;18(12):1449-1464 [FREE Full text] [doi: 10.1080/10810730.2013.798374] [Medline: 23941610]
66. Kiropoulos LA, Griffiths KM, Blashki G. Effects of a multilingual information website intervention on the levels of depression literacy and depression-related stigma in Greek-born and Italian-born immigrants living in Australia: a randomized controlled trial. J Med Internet Res 2011 Apr 19;13(2):e34 [FREE Full text] [doi: 10.2196/jmir.1527] [Medline: 21504872]
67. Lee HY, Koopmeiners JS, Rhee TG, Raveis VH, Ahluwalia JS. Mobile phone text messaging intervention for cervical cancer screening: changes in knowledge and behavior pre-post intervention. J Med Internet Res 2014 Aug 27;16(8):e196 [FREE Full text] [doi: 10.2196/jmir.3576] [Medline: 25164545]
68. Lee HY, Koopmeiners JS, McHugh J, Raveis VH, Ahluwalia JS. Mhealth pilot study: text messaging intervention to promote HPV vaccination. Am J Health Behav 2016 Jan;40(1):67-76 [FREE Full text] [doi: 10.5993/AJHB.40.1.8] [Medline: 26685815]
69. Lee H, Ghebre R, Le C, Jang YJ, Sharratt M, Yee D. Mobile phone multilevel and multimedia messaging intervention for breast cancer screening: pilot randomized controlled trial. JMIR Mhealth Uhealth 2017 Nov 7;5(11):e154 [FREE Full text] [doi: 10.2196/mhealth.7091] [Medline: 29113961]
70. MacDonell KK, Naar S, Gibson-Scipio W, Lam P, Secord E. The Detroit young adult asthma project: pilot of a technology-based medication adherence intervention for African-American emerging adults. J Adolesc Health 2016 Oct;59(4):465-471 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.05.016] [Medline: 27475032]
71. Marcus BH, Hartman SJ, Larsen BA, Pekmezi D, Dunsiger SI, Linke S, et al. Pasos Hacia La Salud: a randomized controlled trial of an internet-delivered physical activity intervention for Latinas. Int J Behav Nutr Phys Act 2016 May 28;13:62 [FREE Full text] [doi: 10.1186/s12966-016-0385-7] [Medline: 27234302]
72. Mauriello LM, van Marter DF, Umanzor CD, Castle PH, de Aguiar EL. Using mhealth to deliver behavior change interventions within prenatal care at community health centers. Am J Health Promot 2016 Sep;30(7):554-562. [doi: 10.4278/ajhp.140530-QUAN-248] [Medline: 26305603]
73. Miller DP, Denizard-Thompson N, Weaver KE, Case LD, Troyer JL, Spangler JG, et al. Effect of a digital health intervention on receipt of colorectal cancer screening in vulnerable patients: a randomized controlled trial. Ann Intern Med 2018 Apr 17;168(8):550-557 [FREE Full text] [doi: 10.7326/M17-2315] [Medline: 29532054]

74. Moussa M, Sherrod D, Choi J. An e-health intervention for increasing diabetes knowledge in African Americans. Int J Nurs Pract 2013 Sep;19(Suppl 3):36-43. [doi: 10.1111/ijn.12167] [Medline: 24090296]
75. Neafsey PJ, M'lan CE, Ge M, Walsh SJ, Lin CA, Anderson E. Reducing adverse self-medication behaviors in older adults with hypertension: results of an e-health clinical efficacy trial. Ageing Int 2011 Jun;36(2):159-191 [FREE Full text] [doi: 10.1007/s12126-010-9085-9] [Medline: 21654869]
76. Nelson LA, Mulvaney SA, Gebretsadik T, Johnson KB, Osborn CY. The messaging for diabetes (MED) intervention improves short-term medication adherence among low-income adults with type 2 diabetes. J Behav Med 2016 Dec;39(6):995-1000 [FREE Full text] [doi: 10.1007/s10865-016-9774-2] [Medline: 27488604]
77. Neuenschwander LM, Abbott A, Mobley AR. Comparison of a web-based vs in-person nutrition education program for low-income adults. J Acad Nutr Diet 2013 Jan;113(1):120-126. [doi: 10.1016/j.jand.2012.07.034] [Medline: 23092741]
78. Phelan S, Hagobian T, Brannen A, Hatley KE, Schaffner A, Muñoz-Christian K, et al. Effect of an internet-based program on weight loss for low-income postpartum women: a randomized clinical trial. J Am Med Assoc 2017 Jun 20;317(23):2381-2391 [FREE Full text] [doi: 10.1001/jama.2017.7119] [Medline: 28632867]
79. Rubinstein A, Miranda JJ, Beratarrechea A, Diez-Canseco F, Kanter R, Gutierrez L, GISMAL group. Effectiveness of an mHealth intervention to improve the cardiometabolic profile of people with prehypertension in low-resource urban settings in Latin America: a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Jan;4(1):52-63. [doi: 10.1016/S2213-8587(15)00381-2] [Medline: 26653067]
80. Ryan JG, Schwartz R, Jennings T, Fedders M, Vittoria I. Feasibility of an internet-based intervention for improving diabetes outcomes among low-income patients with a high risk for poor diabetes outcomes followed in a community clinic. Diabetes Educ 2013;39(3):365-375. [doi: 10.1177/0145721713484594] [Medline: 23610181]
81. Steinberg DM, Levine EL, Askew S, Foley P, Bennett GG. Daily text messaging for weight control among racial and ethnic minority women: randomized controlled pilot study. J Med Internet Res 2013 Nov 18;15(11):e244 [FREE Full text] [doi: 10.2196/jmir.2844] [Medline: 24246427]
82. Tessaro I, Rye S, Parker L, Mangone C, McCrone S. Effectiveness of a nutrition intervention with rural low-income women. Am J Health Behav 2007;31(1):35-43. [doi: 10.5555/ajhb.2007.31.1.35] [Medline: 17181460]
83. Titov N, Dear BF, Ali S, Zou JB, Lorian CN, Johnston L, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of depression: a randomized controlled trial. Behav Ther 2015 Mar;46(2):193-205. [doi: 10.1016/j.beth.2014.09.008] [Medline: 25645168]
84. Ünlü Ince B, Cuijpers P, van 't Hof E, van Ballegooijen W, Christensen H, Riper H. Internet-based, culturally sensitive, problem-solving therapy for Turkish migrants with depression: randomized controlled trial. J Med Internet Res 2013 Oct 11;15(10):e227 [FREE Full text] [doi: 10.2196/jmir.2853] [Medline: 24121307]
85. Wahbeh H, Goodrich E, Oken B. Internet-based mindfulness meditation for cognition and mood in older adults: a pilot study. Altern Ther Health Med 2016;22(2):44-53 [FREE Full text] [Medline: 27036056]
86. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health coaching reduces HbA1c in type 2 diabetic patients from a lower-socioeconomic status community: a randomized controlled trial. J Med Internet Res 2015 Oct 5;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]
87. Weinert C, Cudney S, Hill W. Health knowledge acquisition by rural women with chronic health conditions: a tale of two web approaches. Aust J Rural Health 2008 Oct;16(5):302-307 [FREE Full text] [doi: 10.1111/j.1440-1584.2008.01004.x] [Medline: 18808489]
88. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the cochrane collaboration risk of bias tool and the effective public health practice project quality assessment tool: methodological research. J Eval Clin Pract 2012 Feb;18(1):12-18. [doi: 10.1111/j.1365-2753.2010.01516.x] [Medline: 20698919]
89. Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. Worldviews Evid Based Nurs 2004;1(3):176-184. [doi: 10.1111/j.1524-475X.2004.04006.x] [Medline: 17163895]
90. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
91. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Know Manag E-Learn 2015;7(4):576-600. [doi: 10.34105/j.kmel.2015.07.038]
92. Taylor VM, Jackson JC, Tu S, Yasui Y, Schwartz SM, Kuniyuki A, et al. Cervical cancer screening among Chinese Americans. Cancer Detect Prev 2002;26(2):139-145 [FREE Full text] [doi: 10.1016/s0361-090x(02)00037-5] [Medline: 12102148]
93. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
94. Research-based Web Design and Usability Guidelines. The US Department of Health and Human Services (HHS). 2004. URL: https://webstandards.hhs.gov/guidelines/ [accessed 2019-03-18]

95. Parker S, Prince A, Thomas L, Song H, Milosevic D, Harris MF, IMPACT Study Group. Electronic, mobile and telehealth tools for vulnerable patients with chronic disease: a systematic review and realist synthesis. BMJ Open 2018 Aug 29;8(8):e019192 [FREE Full text] [doi: 10.1136/bmjopen-2017-019192] [Medline: 30158214]
96. Griebel L, Enwald H, Gilstad H, Pohl AL, Moreland J, Sedlmayr M. eHealth literacy research-Quo vadis? Inform Health Soc Care 2018 Dec;43(4):427-442. [doi: 10.1080/17538157.2017.1364247] [Medline: 29045164]
97. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
98. Cunningham S, Wake D, Waller A, Morris A. Definitions of eHealth. In: Gaddi A, Capella F, Manca M, editors. eHealth, Care and Quality of Life: Forewords by Sergio Bertolucci and Gianfranco Gensini. New York and Heidelberg: Springer; 2014:15-30.
99. Gaddi A, Capello F. The debate over eHealth. In: Gaddi A, Capello F, Manca M, editors. eHealth, Care and Quality of Life. New York and Heidelberg: Springer; 2014:1-13.
100. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
101. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014 Apr 9;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
102. Parker SJ, Jessel S, Richardson JE, Reid MC. Older adults are mobile too!Identifying the barriers and facilitators to older adults' use of mHealth for pain management. BMC Geriatr 2013 May 6;13:43 [FREE Full text] [doi: 10.1186/1471-2318-13-43] [Medline: 23647949]
103. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013 Feb 14;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
104. Chang BL, Bakken S, Brown SS, Houston TK, Kreps GL, Kukafka R, et al. Bridging the digital divide: reaching vulnerable populations. J Am Med Inform Assoc 2004;11(6):448-457 [FREE Full text] [doi: 10.1197/jamia.M1535] [Medline: 15299002]
105. Showell C, Turner P. The PLU problem: are we designing personal ehealth for people like us? Stud Health Technol Inform 2013;183:276-280. [Medline: 23388298]
106. Jonkman NH, van Schooten KS, Maier AB, Pijnappels M. eHealth interventions to promote objectively measured physical activity in community-dwelling older people. Maturitas 2018 Jul;113:32-39. [doi: 10.1016/j.maturitas.2018.04.010] [Medline: 29903646]
107. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
108. Barretto AI, Bingham CR, Goh KN, Shope JT. Developing a web-based health promotion intervention: a case study from a brief motivational alcohol program. Health Promot Pract 2011 Mar;12(2):193-201. [doi: 10.1177/1524839909353740] [Medline: 21051327]
109. Thomas J, Barraket J, Wilson C, Cook K, Louie Y. Measuring Australia's Digital Divide: The Australian Digital Inclusion Index 2018. Australian Digital Inclusion Index. 2018. URL: https://digitalinclusionindex.org.au/wp-content/uploads/2018/08/Australian-digital-inclusion-index-2018.pdf [accessed 2020-06-30]
110. Monkman H, Kushniruk AW. The consumer health information system adoption model. Stud Health Technol Inform 2015;218:26-31. [Medline: 26262522]

Abbreviations

eHLF: eHealth Literacy Framework
eHLQ: eHealth Literacy Questionnaire
ICT: information and communications technology
RCT: randomized controlled trial
UDHHS: US Department of Health and Human Services
WHO: World Health Organization

Edited by G Eysenbach; submitted 28.02.20; peer-reviewed by L Griebel, H Monkman, E Brainin; comments to author 18.04.20; revised version received 30.04.20; accepted 03.06.20; published 13.08.20.

Please cite as:
Cheng C, Beauchamp A, Elsworth GR, Osborne RH
Applying the Electronic Health Literacy Lens: Systematic Review of Electronic Health Interventions Targeted at Socially Disadvantaged Groups
J Med Internet Res 2020;22(8):e18476
URL: http://www.jmir.org/2020/8/e18476/
doi:10.2196/18476
PMID:32788144

Original Paper

The Associations Among Individual Factors, Media Literacy, and Dietary Supplement Use Among College Students: Cross-Sectional Study

Shu Ching Yang[1*], PhD; Wan-Chen Hsu[2*], PhD; Chia-Hsun Chiang[1*], PhD

[1]National Sun Yat-Sen University, Kaohsiung, Taiwan

[2]National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

[*]all authors contributed equally

Corresponding Author:
Chia-Hsun Chiang, PhD
National Sun Yat-Sen University
70 Lienhai Rd
Kaohsiung, 80424
Taiwan
Phone: 886 75251521
Fax: 886 75255892
Email: d996050002@student.nsysu.edu.tw

Abstract

Background: The mass media have been condemned for encouraging young people to take dietary supplements (DS). Media literacy, which includes authors and audiences (AA), messages and meanings (MM), and representation and reality (RR) domains, is a new approach to teaching young adults to make better informed health decisions. However, it is not clear which domains are the most important for media literacy education.

Objective: The purpose of this study is to investigate the associations among individual factors, media literacy, and DS use.

Methods: The survey instrument included demographic items, the DS Media Literacy Scale (DSMLS), and DS use items (users or nonusers, types of DS, current use of DS, and intention to use DS in the future). The DSMLS is an 11-item instrument designed to assess college students' AA, MM, and RR media literacy in relation to DS. A total of 467 Taiwanese college students participated in the study. Descriptive statistical analysis, logistic regression analysis, and multiple regression analysis were conducted.

Results: A total of 338/467 (72.4%) participants reported using DS, and 176/467 (37.7%) consumed 3 or more supplements. Moreover, the MM media literacy domain was associated with having been a DS user (odds ratio 0.63, *P*=.002), current DS use (β=–.10, *P*=.02), and intention to use DS in the future (β=–.12, *P*=.011). Finally, perceived importance of health was positively related to current DS use (β=.18, *P*=.001) and intention to use DS in the future (β=.18, *P*=.001).

Conclusions: This study showed that the majority of Taiwanese college students were DS users and used multiple types of supplements. Moreover, students with lower MM media literacy were more likely to be DS users, to take DS more frequently, and to have higher intentions for future frequent DS use. Finally, those who placed extreme importance on health were more likely to take DS frequently and have higher intentions for future frequent DS use.

(J Med Internet Res 2020;22(8):e19056) doi:10.2196/19056

KEYWORDS

college student; dietary supplement; media literacy; ehealth literacy

Introduction

In the food industry, the driving force behind the sale of dietary supplements (DSs) is the creation of a market niche to commercialize innovative products claiming to support good health and supplement the diet [1]. DSs are popular in many countries, and their use has been increasing [2]. In the United States, studies have shown that the majority of DS users (79%) reported using DSs every day within the last 30 days [3] and that multivitamin–mineral products were the most common type of DSs [2,4].

Individuals obtain information on DSs most frequently from the media [5]. DSs are accompanied by media messages and advertisements on the need to optimize nutrition and improve health [1]. A proactive, convenient approach to maintaining health is a powerful appeal to consumers [1]. Studies have found that, among DS users and nonusers, the decision to use DSs is influenced by the media [6,7].

Mass media have been condemned for encouraging young people to take DSs [8]. These specific DS products and advertising for DSs are frequently formulated for the young market [2]. A study has shown that compared with the general population, college students are more likely to take DSs and that over 60% of them take 1 or more supplements weekly [2]. Notably, advertising is more likely to increase adolescents' willingness to use DSs compared with other groups [9]. Individuals are exposed to a large number of diverse media messages, such as health sciences research reports that make sensational claims, in which some specific information is also manipulated for commercial interests [10]. However, adolescents usually do not have the ability to critically analyze the health information received from the media [11]. Scholars have argued that adolescents may take DSs inappropriately due to the questionable information about DSs from magazines [12].

A number of studies have reported that a vast majority of DSs that young people use are not only non-helpful but also often harmful [13]. For example, Or et al. [14] examined the DSs' harmful effects on children, adolescents, and young adults. Using the Food and Drug Administration adverse event data between 2004 and 2015, they found that, compared with vitamins, teenagers using DSs to lose weight, gain weight, or build muscle are nearly three times more likely to experience serious health problem and even death. Thus, they suggest efforts aimed at reducing access and consumption and actively implementing regulations as well as providing a clear warning and labeling to prevent serious medical consequences for children, adolescents, young people, adults, and general consumers.

Media literacy is a new approach to teaching youth to make smarter health decisions [15]. Media literacy is defined as accessing, understanding, and evaluating messages using different forms of media [16,17]. According to the theory of reasoned action, media literacy programs can potentially reduce the negative influence of media on individuals' attitudes and normative beliefs [18]. When applied to health, media literacy can help youth understand how products are promoted by the media and increase their awareness about what they see and hear [15]. Media literacy education guides and enables young people to more actively evaluate, analyze, and process mass media messages rather than passively remain targets of the media [19,20]. Such education is designed to empower young people to develop critical thinking abilities to ameliorate the negative effect of mass media messages and make good decisions about their health [21]. Media literacy education has been an effective approach in health behavior areas, such as nutrition behaviors, body image, or alcohol and drugs [22]. Recently, studies have shown that media literacy interventions are effective for decreasing the use of DSs such as creatine, carnitine, and amino acids among adolescents [23] and can reduce young athletes' use of new sport supplements at the 1-year follow-up [24].

According to Primack et al. [20], media literacy includes 3 domains: "authors and audiences (AA), messages and meanings (MM), and representation and reality (RR)". Most practitioners and researchers agree that the 3 domains are the foundation of media literacy education [15]. However, it is not clear which domains are the most important for media literacy education [15]. Knowledge of which domains are most important will inform research to design more effective media literacy programs to reduce college students' DS use. Therefore, this study investigates whether the AA, MM, and RR domains of media literacy were associated with DS use and examines which domain of media literacy exhibits the highest correlation with DS use. We thus propose the following hypotheses:

- H1: College students with lower AA media literacy are more likely to take DSs.
- H2: College students with lower MM media literacy are more likely to take DSs.
- H3: College students with lower RR media literacy are more likely to take DSs.

The study also investigated the influence of demographic characteristics (ie, perceived importance of health, gender, and subjective health status) on DS use. Studies have shown that DS users tend to take DSs to improve or maintain health [2,4,10]. Studies have also found gender differences in DS use. Compared with males, females were more likely to take DSs [1,25,26]. However, male college students were significantly more likely to take at least five types of DSs weekly [2]. In addition, some studies showed that DS users tend to report very good or excellent health [4,25]. However, other studies found an association between poor subjective health and frequent consumption of calcium tablets [1] and DSs [26]. Given the above studies, we want to investigate whether college students' demographic characteristics (ie, perceived importance of health, gender, and subjective health status) are related to their use of DS.

Methods

Study Design and Participants

The participants in this study were college students in Taiwan. After requesting consent from teachers at the selected schools, we distributed the pen-and-paper survey in their classes. We determined that a minimum of 110 students were needed for the pretest (10 times the number of items on the survey instrument) [27], and 350 students were needed for the formal study [28]. The pen-and-paper questionnaires were administered in class during the regularly scheduled class periods. Students took 10-15 minutes to complete the questionnaire and received a small gift as a reward, regardless of whether they completed the survey.

For the pretest, exploratory factor analysis (EFA) was used to assess the reliability of our survey instrument. A sample of 200 college students from 2 schools was recruited to participate in the survey. Ultimately, 73% of the sample returned questionnaires (N=146) and fully completed all the questions

in the survey, whereas the other 27% (N=54) did not return the survey. The low rate of returned questionnaires was due to the fact that it was the end of the semester, and thus, some teachers and students were not available.

In the formal study, a sample of 500 college students was drawn from 6 colleges and each member of the sample received a questionnaire. Of the 475 surveys received, 8 surveys were invalid (ie, respondents did not complete the entire survey), whereas 467 were valid. Of the 467 participants, 220/467 (47.1%) were female, and 247/467 (52.9%) were male. The participants' mean (SD) age was 20.39 (1.88) years.

Survey Instrument

The questionnaire was composed of items on participants' demographic information, the DS Media Literacy Scale (DSMLS), and items on DS use.

The demographic information included items on age, gender (male or female), perceived importance of health (1 item with a 5-point Likert scale), and subjective health status (1 item with a 5-point Likert scale).

We developed the DSMLS based on a thorough review of the literature [20,29,30], and we asked 3 specialist professors to examine the content validity of the DSMLS. The DSMLS contains 3 domains (11 items) to measure DS media literacy (Table 1). The AA (4 items) domain evaluates individuals' skills used to understand how media target specific consumers. The MM domain (3 items) evaluates individuals' skills used to analyze how messages are created with specific production techniques designed to influence consumers' attitudes and behaviors. The RR (4 items) domain assesses individuals' skills used to analyze how media omit and filter information. The items include a 5-point Likert response scale ranging from *strongly disagree* (coded as 1) to *strongly agree* (coded as 5). Higher scores on the respective levels indicate higher media literacy.

The results of the EFA (principal axis factors method) indicate that the Kaiser–Meyer–Olkin measure was 0.84, the Bartlett test for sphericity was significant (P<.001), the factor loadings ranged from 0.78 to 0.94 (Table 1), and the explained variance was 78.42%. The DSMLS also exhibited good internal consistency reliability (AA Cronbach α=.94, MM Cronbach α=.79, RR Cronbach α=.89).

Table 1. Factor loadings for the DSMLS[a].

Items	AA[b] domain	MM[c] domain	RR[d] domain
1. Media choose stories of dietary supplements based on what will attract the largest audience	0.94	–0.08	0.25
2. The owners of dietary supplement companies influence the advertising content that is produced	0.85	–0.15	0.31
3. Dietary supplement companies only care about making money	0.78	0.04	0.38
4. Some dietary supplement advertisements are designed to appeal to certain people	0.93	–0.08	0.25
5. Dietary supplement ads convince me that dietary supplements make me healthier and more energetic	–0.06	0.85	–0.03
6. Plots about dietary supplements in TV and movies affect my view of dietary supplements	–0.03	0.88	–0.07
7. I am influenced by celebrity endorsements of dietary supplements	–0.09	0.79	0.02
8. There are often hidden harmful messages in dietary supplement advertisements	0.21	–0.03	0.83
9. Advertisements usually highlight and emphasize the benefits of dietary supplements	0.26	0.06	0.83
10. Advertisements usually leave out some important information regarding dietary supplements	0.37	–0.10	0.80
11. When I see a dietary supplement advertisement, it is very important to think about what was left out of the advertisement.	0.25	–0.06	0.82
Explained variance, %	30.88	19.66	27.88

[a]DSMLS: DS Media Literacy Scale.

[b]AA: authors and audiences.

[c]MM: messages and meanings.

[d]RR: representation and reality.

The survey also included detailed questions on the types of DSs participants used and 2 questions about the frequency of DS use. Participants were instructed to list DSs that they used. The 2 questions about the frequency of DS use (rated on a 5-point Likert scale) included current DS use, defined as the frequency of DS use in the past year, and intention for future DS use, defined as the expected frequency of DS use in the next year.

Data Analysis

First, we used EFA to assess the reliability of the DSMLS in the pretest study. In the formal study, we used descriptive statistical analysis, logistic regression analysis, and multiple regression analysis. In the regression analyses, gender was a dummy variable (male=0, female=1), and perceived importance of health and subjective health status were continuous variables. The Omnibus test (P<.05) was used in the logistic regression to estimate the goodness-of-fit of the model. The variance

inflation factor (VIF) and tolerance assessment were used for diagnosing collinearity in multiple regressions. Values of VIF exceeding 4 or values of tolerance less than 0.2 indicate multicollinearity.

Results

Descriptive Statistical Analysis

Table 2 presents the descriptive statistics for the categorical variables. Among all participants, the mean (SD) score was 3.35 (0.81) for subjective health status, 3.51 (0.77) for perceived importance of health, 3.56 (0.63) for the AA domain, 3.43 (0.75) for the MM domain, and 3.87 (0.72) for the RR domain, indicating that the college students reported moderate to good health, considered health to be moderately important, and had moderate levels of AA, MM, and RR media literacy, respectively.

Regarding the frequency of DS use, the mean (SD) score for current DS use was 2.27 (1.05), whereas that of intention for future DS use was 2.28 (1.01). This result indicates that college students had occasionally used DSs in the past year and intended to occasionally use DSs in the next year.

Table 3 shows that 338/467 (72.4%) participants reported taking DSs for 1 year before the survey. Overall, 123/467 (26.3%) participants reported taking at least four types of DSs. The most popular DSs were vitamin B (252/467, 54.0%), multivitamins (191/467, 40.9%), and calcium (119/467, 25.5%).

Table 2. Descriptive statistics for the categorical variables.

Variables	Mean (SD)
Subjective health status	3.35 (0.81)
Perceived importance of health	3.51 (0.77)
AA[a] domain of media literacy	3.56 (0.63)
MM[b] domain of media literacy	3.43 (0.75)
RR[c] domain of media literacy	3.87 (0.72)
Current DS[d] use	2.27 (1.05)
Intention for future DS use	2.28 (1.01)

[a]AA: authors and audiences.
[b]MM: messages and meanings.
[c]RR: representation and reality.
[d]DS: dietary supplement.

Table 3. Participants' reported use of DSs (N=467)[a].

Variable and group	n (%)
DS	
Nonuser	129 (27.6)
User	338 (72.4)
Number of DSs	
1	103 (22.1)
2	59 (12.6)
3	53 (11.3)
4+	123 (26.3)
Top 5 DSs	
Vitamin B (B complex)	252 (54.0)
Multivitamin	191 (40.9)
Calcium	119 (25.5)
Iron or zinc	116 (24.8)
Lutein	43 (9.2)

[a]DS: dietary supplement.

Analysis of Demographic Characteristics, Media Literacy, and DS Use

In the logistic regression analysis, the Omnibus test (chi-square=17.88, *P*=.007) reveals the goodness-of-fit of the model. In the multiple regression analysis, the values of the VIF range from 0.59 to 0.99, and the values of tolerance range from 1.01 to 1.70, indicating that the data are free from multicollinearity.

The results of the logistic regression analysis and multiple regression analysis are presented in Tables 4 and 5 and show that the MM domain of media literacy was associated with being a DS user (odds ratio 0.63; 95% CI 0.47-0.84; *P*=.002). The MM domain of media literacy was negatively related to current DS use (β=–.10, *P*=.02) and intention for future DS use (β=–.12, *P*=.011). However, the AA and RR domains of media literacy were not related to being a DS user, current DS use, or intention for future DS use. Thus, Hypothesis 2 was supported; however, Hypothesis 1 and 3 were not supported.

Table 5 shows that perceived importance of health was positively related to current DS use (β=.18, *P*=.001) and intention for future DS use (β=.18, *P*=.001). However, Table 4 shows that the association between the perceived importance of health and being a DS user was not statistically significant. Gender and subjective health status were not related to being a DS user, current DS use, or intention for future DS use.

Table 4. Logistic regression analysis of DS[a] use.

Characteristic	DS user	
	Adjusted odds ratio (95% CI)	*P* value
Gender		
Male	1.00[e]	
Female	0.84 (0.55-1.30)	.47
Subjective health status	1.02 (0.76-1.38)	.90
Perceived importance of health	1.28 (0.94-1.74)	.12
AA[b] domain of media literacy	0.85 (0.56-1.31)	.46
MM[c] domain of media literacy	0.63 (0.47-0.84)	.002
RR[d] domain of media literacy	0.89 (0.61-1.29)	.53

[a]DS: dietary supplement.

[b]AA: authors and audiences.

[c]MM: messages and meanings.

[d]RR: representation and reality.

[e]Male group was used as the reference group. Thus, CI and *P* values were not reported.

Table 5. Multiple regression analysis of the frequency of DS[a] use.

Variable	Current DS use			Intention for future DS use		
	B	β	*P* value	B	β	*P* value
Gender (female)	–0.01	–.00	.93	0.02	.01	.82
Subjective health status	–0.13	–.10	.07	–0.12	–.10	.07
Perceived importance of health	0.24	.18	.001	0.24	.18	.001
AA[b] domain of media literacy	–0.12	–.07	.24	–0.18	–.11	.06
MM[c] domain of media literacy	–0.15	–.10	.02	–0.16	–.12	.011
RR[d] domain of media literacy	0.03	.02	.74	0.04	.03	.66
Model summary	R=0.20; R^2=0.04; $F_{6,460}$=3.05			R=0.22; R^2=0.05; $F_{6,460}$=3.87		

[a]DS: dietary supplement.

[b]AA: authors and audiences.

[c]MM: messages and meanings.

[d]RR: representation and reality.

Discussion

Principal Findings

This study aimed to investigate the associations between media literacy and DS use in the college-age population. The study revealed that the majority of college students were DS users and used multiple types of supplements. Furthermore, the study found that the MM domain of media literacy showed the highest correlation with DS use. Finally, perceived importance of health was positively related to current DS use and the intention for future DS use.

The study found that 338/467 (72.4%) participants reported using DSs and that 176/467 (37.7%) consumed 3 or more supplements. Among the participants, 252/467 (54.0%) took vitamin B, 191/467 (40.9%) took multivitamins, and 119/467 (25.5%) took calcium. However, few young, healthy students did not intake the required nutrient components [2]. Scholars have recommended that well-nourished adults should stop wasting money on mineral and vitamin supplements; there is no evidence to suggest that these supplements provide any advantage for well-nourished individuals and might even be harmful [31]. Therefore, it is worrying that healthy Taiwanese college students might be more likely to use DSs excessively and thereby harm their health.

As expected, participants with a lower MM media literacy were more likely to be DS users. Moreover, they were more likely to take DSs more frequently and to have higher intentions for future frequent DS use. The MM domain refers to the fact that messages are created with appealing techniques designed to influence consumers' attitudes and behaviors [20,29,30]. Mass media producers often use certain techniques to make DSs seem healthier and to promote the idea that they *optimize* nutrition. A convenient approach to maintain or improve health is an attractive strategy for individuals [1]. The MM domain of media literacy addresses these miscommunications by exposing these techniques that are designed to make behaviors seem more normative [30]. Students with adequate MM media literacy were less likely to be influenced by media messages and techniques, as they could critically evaluate the messages of DS; thus, they were more likely to be DS nonusers, to take DSs less frequently, and to have lower intentions for future frequent DS use. Therefore, increasing college students' awareness about how media messages are created to influence their attitudes and behaviors to reduce their use of DSs is a good strategy.

The study found that the AA and RR domains of media literacy were not related to DS users, their current use, or their intention for future use, thus suggesting that MM may be more important than the other 2 domains in terms of their potential relationship with DS-related behavior. These findings suggest that media literacy programs that merely portray the DS industry as manipulative of particular target markets and describe the difference between narrative media and the true effects of DS use on health may not be enough to reduce college students' DS use. However, it is unlikely that the insufficiency of such a strategy is because the AA and MM concepts are not important. Media literacy is a multifaceted construct, wherein each of the 3 components plays an important role in media literacy and reveals substantial differences in emphasis [20,30]. The AA domain is associated with how profit-driven media target specific market issues, whereas the RR domain focuses on how media information affects one's perceptions of reality.

Lucidi et al.'s study [23] found that media literacy intervention, which contains a series of interlocking cognitive processes, for example, the acquisition of knowledge structures, diverse cognitive skills, and the minimization of possible detrimental effects of the media, is effective in reducing the use of DSs (ie, creatine, carnitine, and amino acids) among adolescents. Thus, it is important to first discuss who the author of a message is (ie, the AA domain) that erases the distinction between representation and reality (ie, the RR domain) before diving into how media messages are created to affect their attitudes and behaviors (ie, the MM domain). Accordingly, a successful media literacy program relies on 3 interlocking domains (ie, the AA, MM, and RR domains). Of these, as the MM domain is the most strongly associated with DS use, it should probably be the most emphasized. In addition, young people are the most active internet users in today's world, and because the internet offers a wide variety of information regarding DSs, future studies should consider the potential influence of electronic health (eHealth) literacy and social media appeals versus advertisements on DS use among young people.

Moreover, the study showed that gender and subjective health status were not related to being a DS user, current use, or intention for future DS use. The study sample was college-age students with high access to resources. Studies have shown that individuals with higher education are more likely to use DSs than those with lower education [1,9]. Moreover, college students are among the healthiest populations [2]. The sample homogeneity may have attenuated the effect of subjective health status and gender on DS use. The associations among gender, age, education, subjective health status, and DS use are worth examining in further studies.

Finally, the study found that the perceived importance of health was not related to being a DS user; however, participants who placed the utmost importance on health were more likely to take DSs frequently and to have higher intentions for future frequent DS use. A total of 338/467 (72.4%) participants in the study were DS users. The high percentage of the population using DSs might buffer the effect of the perceived importance of health on DS users. Messages about and advertisements for DSs are attractive because of the potential for individuals to improve and maintain their health [1]. Previous studies have shown that DS users tend to take DSs to improve or maintain health [2,4,10]. Thus, Taiwanese college students who place the utmost importance on health would likely be taking DSs continuously to improve and maintain their health.

Study Limitations

This study had several limitations. For example, our main media literacy scale is a new scale that has not yet had extensive testing. However, it was based on an established model of media literacy [20]. The study found that the MM domain of media literacy was negatively related to DS use. Future studies should continue to use the same framework or other models to develop and construct other DSMLS instruments. Second, the study

found that vitamin B, multivitamins, and calcium were the popular types of supplements among Taiwanese college students. Future studies should continue to analyze the associations among the 3 domains of media literacy and the frequency of each type of DS use. Third, many detailed demographic characteristics (ie, body mass index, yearly income, diet category, parental attitudes toward DSs, and parental DS use) were not collected as part of this study. Thus, a broad range of factors that might be influential in determining college students' media literacy and DS use should be taken into consideration. Finally, this was a cross-sectional study. The intention for future DS use cannot be taken as actual DS use in the next year. Future investigations should examine the associations between media literacy and DS use longitudinally. The study, however, provides a crucial starting point for such studies.

Conclusions

Mass media have been condemned for encouraging young people to take DSs [8], and DS media literacy is an important subset of media literacy that requires independent investigation. To the best of our knowledge, this study is the first to provide evidence for the associations between DS media literacy and DS use in the college-age population. This study revealed that the majority of the college students were DS users and used multiple types of supplements. Thus, discouraging potentially unnecessary DS use is an important consideration in college health education programs.

The study found that the MM domain of media literacy had the highest correlation with low DS use. This finding suggests that to be most effective, media literacy programs designed for college students should more closely examine the various manipulative marketing techniques of media.

Furthermore, for media literacy programs to be successful, future DS-related media literacy using Primack et al.'s theoretical models can be well designed and integrated via school-based experimental interventions. It is suggested to examine the relationship between DS use outcomes and items measuring specific, individual components of media literacy, and to further examine how well the program buffers the impact of mass media on adolescents' DS use by empowering youth to actively and critically analyze and evaluate these media messages rather than being passive targets of the messages. Ultimately, this study found that participants who placed the utmost importance on health were more likely to take DSs frequently and have higher intentions for future DS use. Therefore, future studies should pay more attention to college students who place the utmost importance on health, thereby helping them make healthy choices and choose appropriate DSs according to their health situations.

Acknowledgments

This research was financially supported by the Intelligent Electronic Commerce Research Center from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan.

Conflicts of Interest

None declared.

References

1. de Jong N, Ocké MC, Branderhorst HAC, Friele R. Demographic and lifestyle characteristics of functional food consumers and dietary supplement users. Br J Nutr 2003 Feb;89(2):273-281. [doi: 10.1079/BJN2002772] [Medline: 12575912]
2. Lieberman HR, Marriott BP, Williams C, Judelson DA, Glickman EL, Geiselman PJ, et al. Patterns of dietary supplement use among college students. Clin Nutr 2015 Oct;34(5):976-985 [FREE Full text] [doi: 10.1016/j.clnu.2014.10.010] [Medline: 25466950]
3. Bailey RL, Gahche JJ, Lentino CV, Dwyer JT, Engel JS, Thomas PR, et al. Dietary supplement use in the United States, 2003-2006. J Nutr 2011 Feb;141(2):261-266 [FREE Full text] [doi: 10.3945/jn.110.133025] [Medline: 21178089]
4. Bailey RL, Gahche JJ, Miller PE, Thomas PR, Dwyer JT. Why US adults use dietary supplements. JAMA Intern Med 2013 Mar 11;173(5):355-361. [doi: 10.1001/jamainternmed.2013.2299] [Medline: 23381623]
5. Rowe S, Toner C. Dietary supplement use in women: the role of the media. J Nutr 2003 Jun;133(6):2008S-2009S. [doi: 10.1093/jn/133.6.2008S] [Medline: 12771355]
6. Conner M, Kirk SFL, Cade JE, Barrett JH. Why do women use dietary supplements? The use of the theory of planned behaviour to explore beliefs about their use. Soc Sci Med 2001 Feb;52(4):621-633. [doi: 10.1016/s0277-9536(00)00165-9] [Medline: 11206658]
7. Conner M, Kirk SFL, Cade JE, Barrett JH. Environmental influences: factors influencing a woman's decision to use dietary supplements. J Nutr 2003 Jun;133(6):1978S-1982S. [doi: 10.1093/jn/133.6.1978S] [Medline: 12771349]
8. Frison E, Vandenbosch L, Eggermont S. Exposure to media predicts use of dietary supplements and anabolic-androgenic steroids among Flemish adolescent boys. Eur J Pediatr 2013 Oct;172(10):1387-1392. [doi: 10.1007/s00431-013-2056-x] [Medline: 23748985]
9. Li YH. The Analysis of the Consumers' Behavior and Willingness to Pay for Liquid Nutritional Supplement in Taiwan's Municipalities. 2016. URL: https://hdl.handle.net/11296/279t4m [accessed 2019-01-10]

10. Nichter M, Thompson JJ. For my wellness, not just my illness: North Americans' use of dietary supplements. Cult Med Psychiatry 2006 Jun;30(2):175-222. [doi: 10.1007/s11013-006-9016-0] [Medline: 16841188]
11. Bergsma L. Media literacy and health promotion for adolescents. Journal of Media Literacy Education 2011;3(1):25-28 [FREE Full text]
12. Shaw P, Zhang V, Metallinos-Katsaras E. A content analysis of the quantity and accuracy of dietary supplement information found in magazines with high adolescent readership. J Altern Complement Med 2009 Feb;15(2):159-164. [doi: 10.1089/acm.2008.0323] [Medline: 19216659]
13. Herriman M, Fletcher L, Tchaconas A, Adesman A, Milanaik R. Dietary supplements and young teens: misinformation and access provided by retailers. Pediatrics 2017 Feb;139(2):e20161257 [FREE Full text] [doi: 10.1542/peds.2016-1257] [Medline: 28044048]
14. Or F, Kim Y, Simms J, Austin SB. Taking stock of dietary supplements' harmful effects on children, adolescents, and young adults. J Adolesc Health 2019 Oct;65(4):455-461. [doi: 10.1016/j.jadohealth.2019.03.005] [Medline: 31176525]
15. Brown JD. Media literacy has potential to improve adolescents' health. J Adolesc Health 2006 Oct;39(4):459-460. [doi: 10.1016/j.jadohealth.2006.07.014] [Medline: 16982377]
16. Chang CS, Liu EZF, Lee CY, Chen NS, Hu DC, Lin CH. Developing and validating a media literacy self-evaluation scale (MLSS) for elementary school students. Turkish Online J Educ Technol 2011;10(2):63-71 [FREE Full text]
17. Livingstone S. Media literacy and the challenge of new information and communication technologies. The Communication Review 2004;7(1):3-14. [doi: 10.1080/10714420490280152]
18. Primack BA, Sidani J, Carroll MV, Fine MJ. Associations between smoking and media literacy in college students. J Health Commun 2009 Sep;14(6):541-555 [FREE Full text] [doi: 10.1080/10810730903089598] [Medline: 19731126]
19. Hobbs R, Frost R. Measuring the acquisition of media-literacy skills. Read Res Q 2003;38(3):330-355. [doi: 10.1598/rrq.38.3.2]
20. Primack BA, Gold MA, Switzer GE, Hobbs R, Land SR, Fine MJ. Development and validation of a smoking media literacy scale for adolescents. Arch Pediatr Adolesc Med 2006 Apr;160(4):369-374 [FREE Full text] [doi: 10.1001/archpedi.160.4.369] [Medline: 16585481]
21. Bergsma LJ, Carney ME. Effectiveness of health-promoting media literacy education: a systematic review. Health Educ Res 2008 Jun;23(3):522-542. [doi: 10.1093/her/cym084] [Medline: 18203680]
22. Castellanos LM. Media Literacy and Health Behaviour Among Children and Adolescents. 2007. URL: http://www.jcsh-cces.ca/upload/Dr.%20Mantilla%20-%20Abstract%20Dr.%20Mantilla.pdf [accessed 2019-03-17]
23. Lucidi F, Mallia L, Alivernini F, Chirico A, Manganelli S, Galli F, et al. The effectiveness of a new school-based media literacy intervention on adolescents' doping attitudes and supplements use. Front Psychol 2017;8:749 [FREE Full text] [doi: 10.3389/fpsyg.2017.00749] [Medline: 28536552]
24. Goldberg L, MacKinnon DP, Elliot DL, Moe EL, Clarke G, Cheong J. The adolescents training and learning to avoid steroids program: preventing drug use and promoting health behaviors. Arch Pediatr Adolesc Med 2000 Apr;154(4):332-338. [doi: 10.1001/archpedi.154.4.332] [Medline: 10768668]
25. Harrison RA, Holt D, Pattison DJ, Elton PJ. Are those in need taking dietary supplements? A survey of 21 923 adults. Br J Nutr 2004 Apr;91(4):617-623. [doi: 10.1079/BJN20031076] [Medline: 15035689]
26. Messerer M, Johansson SE, Wolk A. Sociodemographic and health behaviour factors among dietary supplement and natural remedy users. Eur J Clin Nutr 2001 Dec 11;55:1104-1110. [doi: 10.1038/sj.ejcn.1601272]
27. Wu ML, Tu JT. SPSS and the Application and Analysis of Statistics. Taiwan: Wunan; 2014.
28. Wang WK, Wang CH. Educational Research Methods. Taiwan: Wunan; 2019.
29. Ashley S, Maksl A, Craft S. Developing a news media literacy scale. Journalism & Mass Communication Educator 2013 Jan 02;68(1):7-21. [doi: 10.1177/1077695812469802]
30. Primack BA, Hobbs R. Association of various components of media literacy and adolescent smoking. Am J Health Behav 2009;33(2):192-201 [FREE Full text] [doi: 10.5993/ajhb.33.2.8] [Medline: 18844513]
31. Guallar E, Stranges S, Mulrow C, Appel LJ, Miller III ER. Enough is enough: stop wasting money on vitamin and mineral supplements. Ann Intern Med 2013 Dec 17;159(12):850-851. [doi: 10.7326/0003-4819-159-12-201312170-00011] [Medline: 24490268]

Abbreviations

AA: authors and audiences
DS: dietary supplement
DSMLS: DS Media Literacy Scale
EFA: exploratory factor analysis
MM: messages and meanings
RR: representation and reality
VIF: variance inflation factor

Edited by G Eysenbach; submitted 02.04.20; peer-reviewed by BA Primack, CA Almenara; comments to author 12.06.20; revised version received 11.07.20; accepted 19.07.20; published 31.08.20.

Please cite as:
Yang SC, Hsu WC, Chiang CH
The Associations Among Individual Factors, Media Literacy, and Dietary Supplement Use Among College Students: Cross-Sectional Study
J Med Internet Res 2020;22(8):e19056
URL: http://www.jmir.org/2020/8/e19056/
doi:10.2196/19056
PMID:32865500

Original Paper

Cultural Impact on the Intention to Use Nursing Information Systems of Nurses in Taiwan and China: Survey and Analysis

I-Chiu Chang[1], PhD; Po-Jin Lin[1], MS; Ting-Hung Chen[2], MS; Chia-Hui Chang[3,4], PhD, RN

[1]Department of Information Management, National Chung Cheng University, Chia-Yi, Taiwan

[2]Institute of Health Policy and Management, National Taiwan University, Taipei, Taiwan

[3]Department of Nursing, Taichung Veterans General Hospital, Taichung, Taiwan

[4]Department of Nursing, Hungkuang University, Taichung, Taiwan

Corresponding Author:
Chia-Hui Chang, PhD, RN
Department of Nursing
Taichung Veterans General Hospital
1650 Taiwan Boulevard Sect 4
Taichung, 407204
Taiwan
Phone: 886 4 2359 2525
Email: cjhsnd@gmail.com

Abstract

Background: Nursing workforce shortage has emerged as a global problem. Foreign nurse importation is a popular strategy to address the shortage. The interactions between nursing staff on either side of the Taiwan Strait continue to increase. Since both nurses in Taiwan and nurses in China have adopted nursing information systems to improve health care processes and quality, it is necessary to investigate factors influencing nursing information system usage in nursing practice.

Objective: This study examined the effects of cultural and other related factors on nurses' intentions to use nursing information systems. The findings were expected to serve as an empirical base for further benchmarking and management of cross-strait nurses.

Methods: This survey was conducted in two case hospitals (one in Taiwan and one in China). A total of 880 questionnaires were distributed (n=440 in each hospital).

Results: The results showed effort expectancy had a significant effect on the intention to use nursing information systems of nurses in China (P=.003) but not nurses in Taiwan (P=.16).

Conclusions: Findings suggest nursing managers should adopt different strategies to motivate cross-strait nurses to use nursing information systems. Promoting effort expectancy is more likely to motivate nurses in China than in Taiwan. This discrepancy is probably due to the less hierarchical and more feminine society in Taiwan.

(J Med Internet Res 2020;22(8):e18078) doi:10.2196/18078

KEYWORDS

Nursing information system; intention to use; cultural differences; information literacy

Introduction

Research Background

Differences in political structures influence cultural development. Eveland et al [1] found cultural differences in political discussions and relationship strength. The natural divide between Taiwan and China, the Taiwan Strait, resulted in differences over time in their political systems and in culture and economy. Past studies on cross-cultural behavioral differences have primarily focused on the manufacturing sector [2-6]. Hofstede [7] reported that power distance was less salient in Taiwan than in mainland China. Power distance directly influences the perceived quality of medical services [8]. In other words, culture could play a key role in the perceived quality of cross-strait nursing services.

The shortage of nursing staff is a long-standing global problem. In Taiwan, this shortage is progressively becoming more serious because of Taiwan's rapidly aging population [9]. In China, the number of registered nurses surpassed 4 million at the end of 2018, but the workforce is still insufficient [10]. Recently,

interactions between nursing staff on either side of the Taiwan Strait continue to increase. However, nursing training programs are different between Taiwan and China. Furthermore, both nurses in Taiwan and nurses in China are adopting nursing information systems which are designed according to what they have learned. Therefore, the cross-strait nursing management is becoming an important issue that needs to be properly addressed.

Adopting nursing information systems can strengthen the mechanisms that underlie patient safety, improve service efficiency, improve retention rate, reduce work load, and reduce management costs. Similar to other industries in which considerable investment in information systems is required to gain the competitive advantage, the medical industry also invests significant resources in information systems to maintain quality and improve performance. Nurses are responsible for inpatients' physical and mental well-being, and they are likely to use nursing information systems frequently. Information competency is defined as the ability to find, evaluate, and use information efficiently and is important in electronic or paper-based systems. Information competency is essential for nurses to provide safer, more effective, and more efficient health care and it determines the success in implementing clinical information systems [11]. Research studies have found that nurse information competency can improve the efficient use of information systems and can enhance decision making about patient care [12]. Khezri et al [13] showed that nurses' informatics competencies have a more critical impact on patient outcomes and organizational success than the information systems did, per se.

Nursing Information Systems

Nursing information systems are designed to host all levels of data that nurses collect about nursing activities, resources, research, management, and education. The use of nursing information systems can help nurses provide (or acquire) accurate and real-time clinical information to (or from) patients, physicians, and other health care providers to ensure the provision of high-quality health care. A nursing information system is a computer-based system that collects, retrieves, stores, processes, displays, and communicates information in a timely manner to facilitate nursing care and resource management as well as providing information on patient health [14]. Once the nursing information system has been developed, several important issues emerge (eg, user acceptance, workflow standardization, and automatic operations of procedures). Nguyen et al [15] found that performance expectancy, effort expectancy, and social influence are positively correlated with intention to use nursing information systems, and nurses reported that the nursing information system could reduce their documentation time and enable them to spend more time with patients. Handayani et al [16] systematically reviewed 56 previous studies on user acceptance factors of hospital information systems and related technologies and identified perceived usefulness or performance expectancy, perceived ease of use or effort expectancy, system quality, and subjective norms or social influence (15 frequent factors) as factors.

National Culture

A thorough understanding of the background, characteristics, and values of organizational members from different cultures can offer organizations new opportunities to gain competitive advantages and avoid potential threats in information systems. Hofstede et al [17] classified culture into different categories based on six dimensions, namely, power distance, uncertainty avoidance, individualism, masculinity, long-term orientation, and indulgence and the scores on these six dimensions were computed using formulas for 76 countries. Across the Taiwan strait, the same Chinese language is spoken. However, there are cultural differences between these two regions, therefore, cross-strait staff undergo substantial cross-cultural adaptations [18]. The scores that China and Taiwan obtained on the dimensions of power distance, masculinity, uncertainty avoidance, and indulgence revealed that there were substantial differences between the two cultures. Hofstede's interpretations of the six dimensions are shown in Table 1.

In Table 1, scores between Taiwan and China are largely different in the dimensions of power distance, masculinity, uncertainty avoidance, and indulgence. Chen et al [19] later indicated that cultural differences of individualism, uncertainty avoidance, masculinity, and long-term orientation are more obvious between Taiwan and China. Lin et al [20] used masculinity, collectivism, and uncertainty avoidance to explore the influences on social media users in Taiwan acquiring and sharing health-related information.

In this study, the common dimensions of prior studies—masculinity and uncertainty avoidance—were chosen to verify the culture differences between the two groups of nursing information system users and were used as moderators to examine how their intention on using nursing information systems was influenced accordingly.

Table 1. Six-dimensional model of the national culture of Taiwan and China based on Hofstede [7].

Dimension	Taiwan (score)	China (score)
Power distance	People accept the established order without requiring further justification (58).	People believe that inequalities among people are acceptable (80).
Individualism	A collectivistic society in which long-term commitments are shared with close group "members" such as family members (17).	A collectivist culture in which people act in accordance with the interests of the group rather than personal interests (20).
Masculinity	A slightly feminine society in which the quality of life is a sign of success and standing out from the crowd is not desirable (45).	A patriarchal society that is driven by competition, achievement, and success, whereby people may prioritize work over family and leisure (66).
Uncertainty avoidance	People have a strong preference for the avoidance of uncertainty, resulting in an emotional need for rules (69).	A culture in which people are comfortable with ambiguity and believe that truth is relative (30).
Long-term orientation	People have a pragmatic long-term orientation and an ability to adapt traditions to modern contexts (93).	People believe that truth, to a great extent, depends on a given situation, context, and time (87).
Indulgence	No preference on this dimension (49).	People are restrained by social norms and believe that indulging themselves is somewhat wrong (24).

Information Literacy

Information literacy is a set of competencies and skills that help individuals effectively locate, evaluate, and use the required information. Health care organizations obtain real-time information that can be used for efficient work processes, better patient safety, clinical decision making, and effective health care. Evidence-based nursing practice requires resilient, innovative, accurate, and helpful real-time information, not only from the hospital-wide record systems with quality and performance indicators but also from advanced nursing-related research evidence, and transfers that information to existing routine nursing practice pertaining to patient care. Past studies have found that, in their workplaces, nursing staff often lack the time, resources, and tools to make use of comprehensive real-time information to make decisions regarding patient care. Kleib et al [21] found that mean informatics competency scores were significantly related to the age, educational qualification, and years of work experience of nurses. Abdrbo [22] recommended the incorporation of information competencies into nursing programs and the introduction of a major course on nursing informatics. In summary, information literacy is important for nurses on their acceptance of the nursing information system.

The unified theory of acceptance and use of technology (UTAUT) framework is widely used to facilitate research on the adoption of an information system or information technology. UTAUT [23] was proposed and validated on electronic medical records system. Since both the systems and users of nursing information systems and electronic medical records systems are different, adjustments needed to be made to the UTAUT model to incorporate our research goal. In this study, our research model was developed based on the UTAUT by adopting the key variables of performance expectancy, effort expectancy, social influence, and intention to use. Meanwhile, two aforementioned constructs, culture and information literacy, were added to explore the nurses' intention of using nursing information systems.

Methods

Participants

In this study, we selected two hospitals, Taichung Veterans General Hospital (in Taiwan) and Suzhou Tertiary Hospital (in China). Both hospitals have nursing information systems that are considerably more advanced than those of other hospitals in their respective countries. However, the nursing information systems vary considerably in their scopes. The nursing information systems of Taichung Veterans General Hospital have more functions and wider applications regarding safety and quality assurance. Both nursing information systems and their host hospitals are described.

Taichung Veterans General Hospital is a medical center in central Taiwan that has 1500 beds and 3700 employees. It can serve 7000 outpatients and 190 patients/day at the emergency room. The development of its hospital information systems began in 1970, and nursing information systems were implemented in 1993. This was the first of its kind implemented in the nursing department in Taiwan. Subsequently, a wide range of information-based systems were implemented in the department, including a web-based nursing information systems, mobile nursing stations, computerized nursing shift guidelines, a computerized discharge plan, digital signatures on digital medical records, an emergency room nursing information systems, a nutritional evaluation system, a social welfare consultation services support system, a system for patient recognition during blood transfusion, a newborn care evaluation tool, and six types of pain assessments. In 2016, this hospital's nursing information systems received the Safety and Quality Certificate Silver Award under the category of National Health and Medical Quality.

In 1988, Suzhou Tertiary Hospital was founded as the General Hospital for the Nuclear Industry and for Sino-French Friendship. It has 1976 employees, including 1653 health professionals, 18 teaching doctors, and 105 teaching postgraduates. The initial information systems implemented in this hospital were electronic health care record systems, which consisted of computerized nursing forms, vital-sign records,

and a credit management system. The first version of the nursing management information system was also implemented. Other supporting systems have also been implemented such as statistical analysis of adverse events, label printing, specimen collection and confirmation information systems, supply center information systems, the quality of health and medical records, blood transfusion system, drug tracking system, nursing quality management system, emergency information system, and catheter room information system.

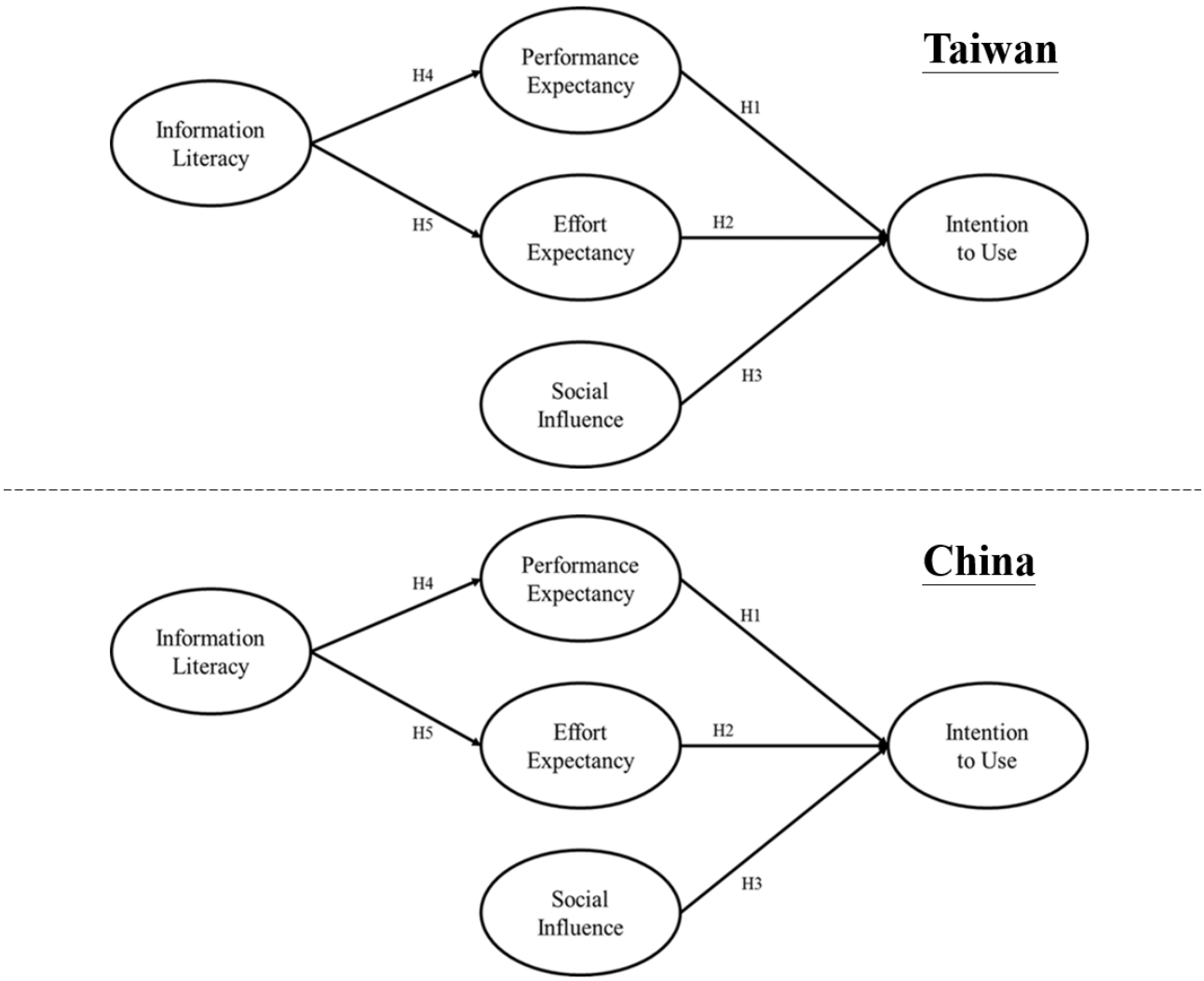
Ethical approval to conduct this study was approved by the Institutional Review Board of Taichung Veterans General Hospital (serial number: CE16137A#1). In accordance with the Institutional Review Board regulations of this study, the data are not publicly available.

Nursing staff members who had at least two years of work experience in the two aforementioned hospitals and prior experience with the nursing information systems were invited for study. All the selected nursing members were informed that their participation was voluntary, that they had the right to withdraw from the study at any time, and that their personal information remained confidential.

Study Design

Nursing information systems in the two case hospitals were not identical. The purpose of the study was to investigate system usability instead of the systems per se, hence we adopted the key constructs of the UTAUT related to system usability. Since both hospitals had a sufficient number of up-to-date facilities for the functioning of nursing information systems, we examined only 3 variables that influence users' intention of the UTAUT (ie, performance expectancy, effort expectancy, and social influence), and in addition, used information literacy and culture variables to ascertain the acceptance and usage of nursing information systems. Performance expectancy was defined as the extent to which an individual believes that using nursing information systems will help enhance job performance. Effort expectancy was defined as the ease-of-use expectancy in using nursing information systems. Social influence was defined as the extent to which an individual believes that others expect them to use the nursing information systems. The research model is shown in Figure 1. The research model was tested differently between the Taiwan and China groups; therefore, culture was used as moderator to examine all the relations of the research framework. Based on the research model shown in Figure 1, we formulated the following hypotheses.

Figure 1. Research framework.



Hypothesis 1: Nurses who believe nursing information systems will enhance their performance are more willing to use the system.

Hypothesis 2: Nurses who expected a high degree of ease in using the nursing information systems will be more willing to use the systems.

Hypothesis 3: Nurses who perceive other caregivers who use nursing information systems receiving higher levels of social influence are more willing to use the system.

Hypothesis 4: Nurses with higher levels of information literacy will have higher performance expectancies of nursing information systems.

Hypothesis 5: Nurses with higher levels of information literacy will expect higher degrees of ease in using nursing information systems.

The model was validated by data obtained separately by the two groups of nurses. Culture is a moderator that moderates the five relationships among variables of the research model. As previously mentioned, the common dimensions used in prior studies, ie, masculinity and uncertainty avoidance, were chosen to verify the cultural differences of the two groups of nursing information system users. A two-sample *t* test showed that masculinity (t_{797}=–5.15, *P*=.002) and uncertainty avoidance (t_{797}=4.45, *P*=.02) dimensions were significantly different between two groups (Multimedia Appendix 1). Therefore, the masculinity and uncertainty avoidance dimensions were used as moderators in this study.

Instrument Development

Based on the literature review, we first developed a preliminary list of measurement items. Subsequently, we modified the item contents to enhance the reliability and validity of the indicators. Participants were instructed to respond to each item on a 5-point Likert scale (with 1 = strongly disagree, 5 = strongly agree). In order to examine the validity of the questionnaire, 3 experts (2 from China and 1 from Taiwan) with substantial experience in the field of nursing information systems were invited. In addition, 2 academic experts who taught and published papers on nursing information systems developments were invited to form an expert panel. They were asked to examine the research framework, format of the measurement items, length of the instrument, and wording of the scale items. Based on the feedback and suggestions that were provided by the 5 experts, the questionnaire was modified to enhance the validity of the scale items. The modified measurement items were pretested on 3 nurses. Next, a pilot study was conducted and 15 experienced nursing information systems users in each of the two hospitals were invited to participate in the pilot study. According to Gorsuch [24], the number of participants should be at least 5 to 10 times the number of items, and the sample size should be greater than 100 whenever factors analyses are performed on the acquired data. The questionnaire consisted of 34 items related to the research model (Multimedia Appendix 1) and 6 items of demographic information including age, educational level, job title, work experience, and level of computer skills. Finally, we distributed online or paper questionnaires (N=880; Taiwan: n=440; China: n=440). SPSS statistical software (version 22.0; IBM Corp) and SmartPLS (version 3.2.6; SmartPLS GmbH) [25] were used to conduct statistical analyses. In this study, the bootstrapping method with 5000 resamples was used to determine the level of statistical significance of each hypothesis [26].

Reliability and Validity

We employed the partial least squares approach of structural equation modeling to examine the psychometric properties of the assessment and path coefficients of the structural equation. This analysis was executed using SmartPLS. Reliability analysis was conducted by computing Cronbach α and composite reliability. Scales were considered to have satisfactory reliability if the Cronbach α and composite reliability of each construct were higher than 0.8 and 0.70 [26,27], respectively. In this study, the Cronbach α and composite reliability of all the constructs exceeded 0.8. Therefore, the scales had a satisfactory reliability. Furthermore, the average variance extracted value of each construct should be greater than 0.50, the variance caused by measurement error, to have convergent validity [26]. In this study, the average variance extracted of all the constructs exceeded 0.64. Therefore, the scales had a reasonable convergent validity. Finally, the values of discriminant validity, the square root of the average variance extracted value of each construct in the model, should all be greater than the estimated correlations between the respective construct and other constructs. The square root of the average variance extracted value of each construct is shown in the Multimedia Appendix 1. Items with loadings less than 0.4 were excluded from the final analysis. The results of reliability and validity analyses are shown in Multimedia Appendix 1. The instrument demonstrated acceptable composite reliability and convergent and discriminant validity.

Results

Overview

Questionnaires with invalid responses were discarded prior to statistical analyses. Valid responses were obtained from a total of 799 participants (Taiwan: n=400, China: n=399).

Descriptive Statistics

Table 2 shows the demographic characteristics of the participants, including their age, educational level, job title, work experience, and level of computer skills.

Most nurses in our study were women (394/400, 98.5%) with an intermediate level of computer literacy (299/400, 74.8%). Nurses from the hospital in Taiwan were older (older than 30 years: 264/400, 66.0%), had a job title of nurse (305/400, 76.3%), had a bachelor's degree or higher educational qualification (358/400, 89.6%), and had longer work experience (216/400 or 54.1% had more than 10 years of work experience). In contrast, most of the nurses from the hospital in China were younger (older than 30 years: 179/399, 44.8%), had a job title of team leader (242/399, 60.7%), had a bachelor's degree or higher educational qualification (274/399, 68.7%), and had shorter work experience (>10 years: 142/399, 35.5%).

The high ratio of nurses with bachelors' degrees or higher, 89.6% (358/400) for Taiwan and 68.7% (274/399) for China, may explain their intermediate level of computer literacy, because courses of introduction to computer software or basic programming are taught as compulsory general courses at college level on both sides.

Table 2. Demographic characteristics of participants from Taiwan and China.

Characteristic, category	Taiwan (n=400), n (%)	China (n=399), n (%)
Gender		
Male	6 (1.5)	6 (1.5)
Female	394 (98.5)	393 (98.5)
Age (years)		
<25	20 (5.0)	45 (11.3)
26-30	116 (29.0)	175 (43.9)
31-35	87 (21.8)	94 (23.6)
36-40	44 (11.0)	42 (10.5)
41-45	58 (14.5)	27 (6.8)
46-50	45 (11.3)	9 (2.3)
>50	30 (7.5)	7 (1.8)
Educational level		
High school	0 (0.0)	3 (0.8)
Community college	42 (10.5)	122 (30.6)
University bachelor's degree	303 (75.8)	268 (67.2)
Master's degree or higher	55 (13.8)	6 (1.5)
Job title		
Nurse	305 (76.3)	47 (11.8)
Team leader	6 (1.5)	242 (60.7)
Vice head nurse	18 (4.5)	94 (23.6)
Head nurse or higher	13 (3.3)	16 (4.0)
Other	58 (14.5)	0 (0.0)
Work experience (years)		
1-2	0 (0.0)	11 (2.8)
2-5	72 (18.0)	77 (19.3)
5-10	112 (28.0)	169 (42.4)
10-15	67 (16.8)	66 (16.5)
>15	149 (37.3)	76 (19.0)
Level of computer skills		
High	57 (14.2)	33 (8.3)
Intermediate	299 (74.8)	305 (76.4)
Low	44 (11.0)	61 (15.3)

Structural Model Analysis

Figure 2 and Figure 3 show the estimated path coefficients and explanatory R^2 values separately for the samples from the hospitals in Taiwan and China.

Results from analyses of the sample of nurses from the hospital in Taiwan supported all our hypotheses, except hypothesis 2 (β=0.055, *t*=1.5, *P*=.16). Performance expectancy (β=0.284, *t*=5.6, *P*<.001) and social influence (β=0.533, t=10.0, *P*<.001) had a positive effect on nursing information systems usage intentions. Hypothesis 1 and hypothesis 3 were hence supported. However, effort expectancy was not significantly related to nursing information systems usage intentions. Hypothesis 2 was hence not supported. Information literacy had a positive effect on both performance (β=0.373, *t*=5.1, *P*<.001) and effort expectancy (β=0.503, *t*=8.9, *P*<.001). Hypothesis 4 and hypothesis 5 were hence supported. Overall, the model explained

65.7% (R^2=0.657) of the variance in nursing information systems usage intentions.

Results from analyses of the sample from the hospital in China supported all our hypotheses. Performance (β=0.155, *t*=3.5, *P*<.001), effort expectancy (β=0.147, *t*=2.9, *P*=.003), and social influence (β=0.550, *t*=11.1, *P*<.001) had a positive influence on nursing information systems usage intentions. Hypothesis 1, hypothesis 2, and hypothesis 3 were hence supported. Information literacy had a positive influence on both performance (β=0.306, *t*=3.3, *P*<.001) and effort expectancy (β=0.529, *t*=8.8, *P*<.001). Hypothesis 4 and hypothesis 5 were hence supported. Overall, the model explained 56.6% (R^2=0.566) of all the variance in nursing information system usage intentions. The comparison of the results of path analyses and hypothesis testing between the samples are shown in Table 3.

Figure 2. Path analysis results of Taiwan.

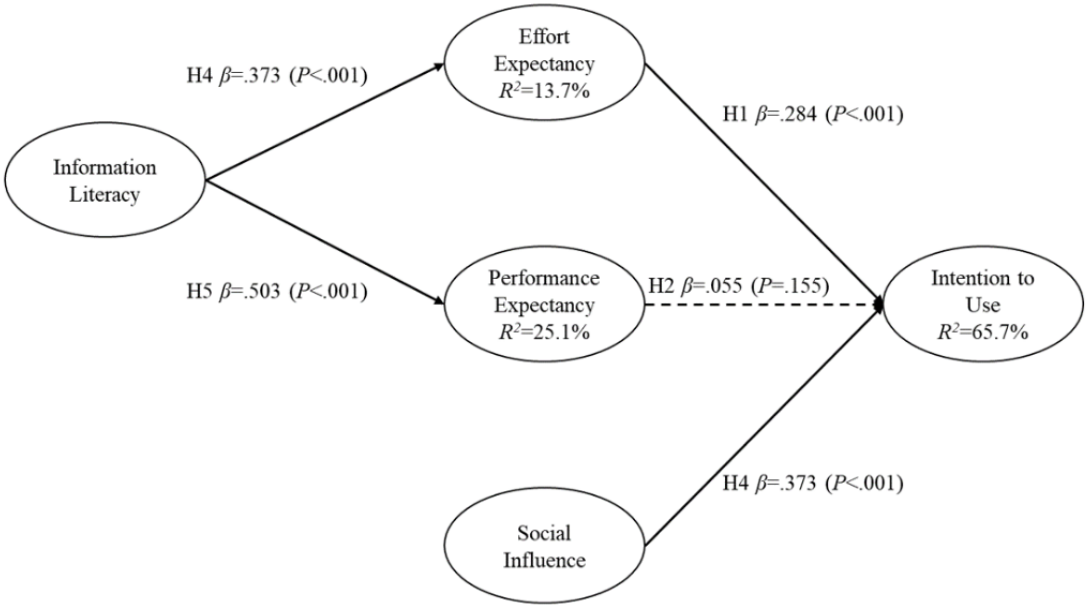


Figure 3. Path analysis results of China.

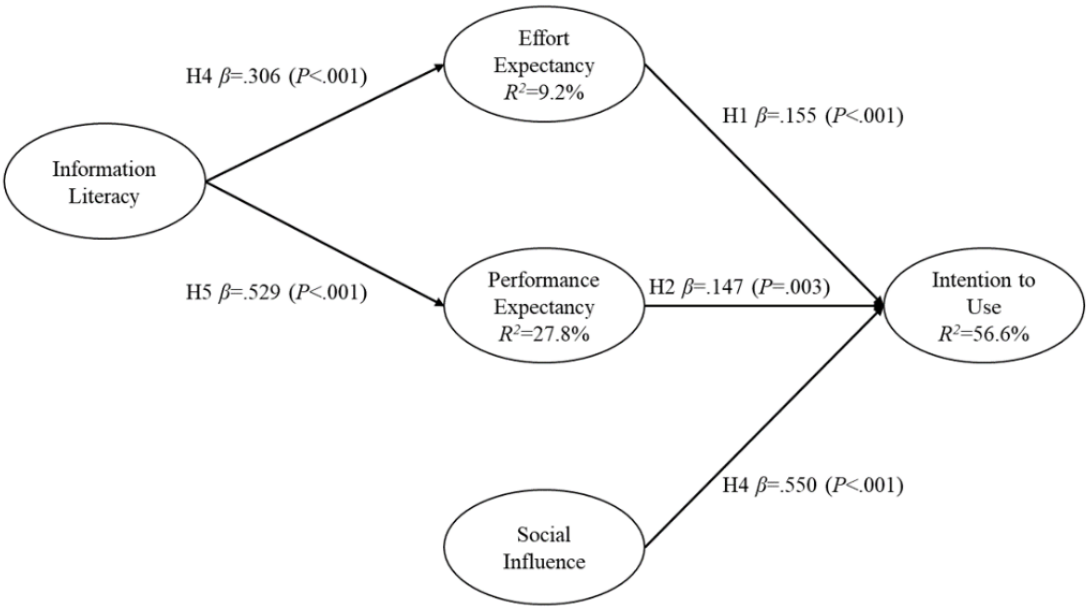


Table 3. Comparison of hypotheses test results.

Hypothesis	Nurses in Taiwan				Nurses in China			
	β	*t* value	*P* value	Decision	β	*t* value	*P* value	Decision
1. Nurses who believe the NIS[a] will enhance their performance are more willing to use the system.	0.284	5.557	<.001	Support	0.155	3.535	<.001	Support
2. Nurses who expect a high degree of ease in using the NIS will be more willing to use the system.	0.055	1.461	.16	Reject	0.147	2.959	.003	Support
3. Nurses who perceive other caregivers who use the NIS have a higher level of social influence than them are more willing to use the system.	0.533	10.021	<.001	Support	0.550	11.135	<.001	Support
4. Nurses with higher levels of information literacy will have higher performance expectancies of the NIS.	0.373	5.083	<.001	Support	0.306	3.276	<.001	Support
5. Nurses with higher levels of information literacy will expect higher degree of ease in using the NIS.	0.503	8.946	<.001	Support	0.529	8.828	<.001	Support

[a]NIS: nursing information system

All the relationships in the research framework between the groups were similar except the effect of effort expectancy on nursing information systems usage intentions was stronger with participants from the hospital in China (β=0.147, *t*=3.0, *P*=.003) than with participants from the hospital in Taiwan (β=0.055, *t*=1.5, *P*=.16). With regard to cultural effects, this indicated that, assuming all else equal, the participants from the hospital in China were more prepared to take risks (lower uncertainty avoidance) and more competitive (greater masculinity) than the participants from the hospital in Taiwan and hence showed greater ease to use a nursing information system having a significant effect on their intention to use nursing information system. In contrast, nurses from the hospital in Taiwan lived in a more risk averse, less competitive, and slightly feminine society. They would likely have used nursing information systems regardless of the degree of ease in using the system. Furthermore, nurses from the hospital in Taiwan may have more realistic expectations of the amount of effort required in using the nursing information systems.

Discussion

Culture Influences Nursing Information Systems Usage Intention

In this study, the key constructs of the UTAUT, cultural differences, and information literacy were integrated to develop a better theoretical framework through which factors affecting nursing information systems usage intentions can be examined. First, the results confirmed Nguyen et al's [15] findings using the UTAUT model. Furthermore, this study is the first of its kind in exploring cultural influences on nursing information systems usage intentions. Results revealed that performance expectancy and social influence were significant predictors of nursing information system usage intentions. In other words, nurses are likely to use a particular system if they believe that it will enhance their performance and fulfill social expectations. In addition to promoting performance expectancies, nursing managers should capitalize on the social influences from peer groups and colleagues to promote usage of nursing information systems among both nurses in Taiwan and nurses in China. It is also important for managers to be aware that the positive effects of performance expectancy on the motivation to use nursing information systems are stronger among nurses in Taiwan than among nurses in China.

However, operational ease has an effect only on users who belong to societies (eg, China) that are more accepting of inequality and people are more driven by competition. Therefore, hospital managers should promote effort expectancies to motivate nurses in China to use nursing information systems. For example, helping nurses complete their tasks more easily and improving the quality of their work should be regarded as the most important objective in nursing information system designs. Samples of nurses in Taiwan and in China showed that information literacy had a positive effect on performance and effort expectancies. This indicates that stronger information literacy promotes expectancies that the use of nursing information systems will improve performance and not be effortful. This in turn likely enhances nursing information systems usage intentions.

Research Limitations

This study collected data via both online and paper questionnaires. Since the questionnaire was self-reported by respondents, it is hard to confirm the authenticity of respondents' answers. In addition, assuming everything else being equal, culture is expected to be the main reason why effort expectancy was related to nursing information system use only in China but not in Taiwan. Since the study was conducted in an open environment instead of in an experimental laboratory, applying the results of this study may need caution.

Direction of Future Studies

Cross-strait nurse movement is mainly from mainland China to Taiwan, due to Taiwan's nurse shortage. However, this study

also found a shortage of nurses in mainland China (nursing managers and leaders) which led to nurse movement from Taiwan to mainland China. In addition, both Chinese and Taiwanese medical systems are hybrid models and each hospital may develop its own medical information systems. Therefore, not only cross-strait but also cross-hospital nurses need to adopt new nursing information systems. This study explored the impact of culture and UTAUT key variables on nurses' nursing information systems usage intentions to develop strategies and refine nursing management and care processes. Future studies can explore more factors affecting perceived ease of use and satisfaction with using nursing information system, to incorporate motivation theory, and to extend the present study findings to further promote the acceptance of nursing information systems among users and mitigate the nursing workforce shortage.

Conclusions

Managing an international workforce is a global issue and understanding cultural differences can help managers run their organizations smoothly and efficiently. In this regard, this study has two key contributions. First, we explored perceptions of and knowledge about nursing information systems of nurses in China and nurses in Taiwan. With knowledge of the differences between nurses on both sides of the Taiwan Strait, hospitals can be in a position to better encourage nurses to use nursing information systems more efficiently while providing care. Meanwhile, differences in the scope of the nursing information system between both countries can provide focus points for cross-strait on the job nurse training programs. The consequence is likely to further enhance the quality of services when hiring cross-strait nurses.

Additionally, cross-strait system vendors can make use of our findings to better understand behavioral differences in the acceptance and use of the nursing information systems between nurses in Taiwan and nurses in China. System vendors should modify their system development strategies to fit the cross-strait nursing industry better. For example, more intuitive user interfaces such as drop-down menus and popup help desks or windows are suggested to increase the degree of ease in using nursing information systems. Finally, this study is one of the first attempts to explore cultural influences on the cross-strait nursing industry. Cultural differences are often discussed within the context of the manufacturing industry, but only Li et al [28] have explored such differences in the behavioral intentions of health care employees. In this regard, this study expands the scope of this domain of research and contributes new findings to the existing literature.

Acknowledgments

This study was supported by the Taichung Veterans General Hospital (Grant TCVGH-1057407). They played no role in the design of the study, analysis, interpretation, reporting of results, or the decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supporting information.
[PDF File (Adobe PDF File), 220 KB - jmir_v22i8e18078_app1.pdf]

References

1. Eveland WP, Song H, Beck PA. Cultural variations in the relationships among network political agreement, political discussion frequency, and voting turnout. Int J Public Opin Res 2015 Apr 15:edv007. [doi: 10.1093/ijpor/edv007]
2. Chen Y, Lin T, Yen DC. How to facilitate inter-organizational knowledge sharing: the impact of trust. Information & Management 2014 Jul;51(5):568-578. [doi: 10.1016/j.im.2014.03.007]
3. Abbas M, Raja U, Darr W, Bouckenooghe D. Combined effects of perceived politics and psychological capital on job satisfaction, turnover intentions, and performance. Journal of Management 2012 Aug 10;40(7):1813-1830. [doi: 10.1177/0149206312455243]
4. Dawkins CE, Jamali D, Karam C, Lin L, Zhao J. Corporate social responsibility and job choice intentions. Business & Society 2014 Dec 22;55(6):854-888. [doi: 10.1177/0007650314564783]
5. Obeidat B, Al-Sarayrah S, Tarhini A, Al-Dmour RH, Al-Salti Z, Sweis R. Cultural influence on strategic human resource management practices: a Jordanian case study. IBR 2016 Sep 06;9(10):94. [doi: 10.5539/ibr.v9n10p94]
6. Al-Sarayrah S, Tarhini A, Obeidat BY, Al-Salti Z, Kattoua T. The effect of culture on strategic human resource management practices: a theoretical perspective. Journal of Business Management and Economic Research 2016;7(4):704-716.
7. 6-D Model of National Cultures. Hofstede, G. 2018. URL: https://www.hofstede-insights.com/product/compare-countries/ [accessed 2018-12-25]
8. Polsa P, Fuxiang W, Sääksjärvi M, Shuyuan P. Cultural values and health service quality in China. International J Health Care QA 2013 Jan 04;26(1):55-73. [doi: 10.1108/09526861311288640]
9. LIN C, HUANG C, YANG C, LU M. The relationship between work environment satisfaction and retention intention among nursing administrators in Taiwan. Journal of Nursing Research 2019;27(5):e43. [doi: 10.1097/jnr.0000000000000312]

10. Regular Press Conference's Text Record on 8 May 2019 for National Health Commission of the People's Republic of China Internet. Ministry of Publicity. 2019. URL: http://www.nhc.gov.cn/xcs/s7847/201905/75ac1db198ae4ecb8aa96c50ea853d0c.shtml [accessed 2019-12-21]
11. Hwang J, Park H. Factors associated with nurses' informatics competency. Comput Inform Nurs 2011;29(4):256-262. [doi: 10.1097/ncn.0b013e3181fc3d24]
12. Al-Hawamdih S, Ahmad MM. Examining the relationship between nursing informatics competency and the quality of information processing. Comput Inform Nurs 2018;36(3):154-159. [doi: 10.1097/cin.0000000000000379]
13. Khezri H, Abdekhoda M. Assessing nurses' informatics competency and identifying its related factors. Journal of Research in Nursing 2019 Apr 16;24(7):529-538. [doi: 10.1177/1744987119839453]
14. Saba V, McCormick K. Essentials of Nursing Informatics. New York: McGraw-Hill Professional; 2015.
15. Nguyen L, Haddad P, Moghimi H, Muhammad I, Coleman K, Redley B, et al. The acceptance of nursing information systems: an analysis using UTAUT. In: Contemporary Consumer Health Informatics. Switzerland: Springer International; 2016:347-365.
16. Handayani PW, Hidayanto AN, Budi I. User acceptance factors of hospital information systems and related technologies: systematic review. Informatics for Health and Social Care 2017 Aug 22;43(4):401-426. [doi: 10.1080/17538157.2017.1353999]
17. Hofstede G, Hofstede G, Minkov M. Cultures and organizations: software of the mind. New York: McGraw-Hill; 2010.
18. Tsai Y. A research on the cross-cultural training needs of Taiwanese medical professionals before dispatched to China - a case study of a Taiwan medical center dissertation. Kaohsiung, Taiwan: National Sun Yat-sen University; 2014.
19. Chen Y, Wu J, Chung Y. Cultural impact on trust: a comparison of virtual communities in China, Hong Kong, and Taiwan. Journal of Global Information Technology Management 2014 Sep 09;11(1):28-48. [doi: 10.1080/1097198x.2008.10856460]
20. Lin H, Ho W. Cultural effects on use of online social media for health-related information acquisition and sharing in Taiwan. Int J Hum Comput Interact 2018 Apr 10;34(11):1063-1076. [doi: 10.1080/10447318.2017.1413790]
21. Kleib M, Nagle L. Factors associated with Canadian nurses informatics competency. Comput Inform Nurs 2018 Aug 1;36(8):406-415. [doi: 10.1097/cin.0000000000000434]
22. Abdrbo AA. Nursing informatics competencies among nursing students and their relationship to patient safety competencies: knowledge, attitude, and skills. Comput Inform Nurs 2015 Nov;33(11):509-514. [doi: 10.1097/cin.0000000000000197]
23. Venkatesh V, Sykes T, Zhang X. 'Just what the doctor ordered': a revised UTAUT for EMR system adoption and use by doctors. 2011 Jan 4 Presented at: In 44 th Hawaii International Conference on System Sciences (HICSS); 2011; Kauai, HI, USA p. 1-10. [doi: 10.1109/hicss.2011.1]
24. Gorsuch R. Factor Analysis, 2nd Edition. State of New Jersey: Lawrence Erlbaum Associates; Nov 1, 1983.
25. Ringle C, Wende S, Becker J. SmartPLS 3, SmartPLS GmbH, Bönningstedt. 2015. URL: https://www.smartpls.com/ [accessed 2019-01-20]
26. Hair J, Joseph F, Hult G, Ringle C, Sarstedt M. A primer on partial least squares structural equation modeling (PLS-SEM). Thousand Oaks: Sage Publications Inc; 2016.
27. George D, Mallery P. IBM SPSS statistics 19 step by step: A simple guidereference. Boston, MA: Pearson; 2012.
28. Li KK, Chan MWH, Lee SS, Kwok KO. The mediating roles of social benefits and social influence on the relationships between collectivism, power distance, and influenza vaccination among Hong Kong nurses: A cross-sectional study. Int J Nurs Stud 2019 Nov;99:103359. [doi: 10.1016/j.ijnurstu.2019.05.007] [Medline: 31493757]

Abbreviations

UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 02.02.20; peer-reviewed by M Hurmuz, YH Chang, C Urquhart; comments to author 16.03.20; revised version received 11.05.20; accepted 13.06.20; published 12.08.20.

Please cite as:
Chang IC, Lin PJ, Chen TH, Chang CH
Cultural Impact on the Intention to Use Nursing Information Systems of Nurses in Taiwan and China: Survey and Analysis
J Med Internet Res 2020;22(8):e18078
URL: https://www.jmir.org/2020/8/e18078
doi:10.2196/18078
PMID:32784174

Original Paper

Identification of Factors Influencing the Adoption of Health Information Technology by Nurses Who Are Digitally Lagging: In-Depth Interview Study

Jacqueline A De Leeuw[1*], PhD; Hetty Woltjer[2*], MSc; Rudolf B Kool[3*], MD, PhD

[1]Department of Information Management, Radboud University Medical Center, Nijmegen, Netherlands

[2]Unit Process Improvement and Implementation, Radboud University Medical Center, Nijmegen, Netherlands

[3]Department IQ Healthcare, Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands

[*]all authors contributed equally

Corresponding Author:
Jacqueline A De Leeuw, PhD
Department of Information Management
Radboud University Medical Center
PO Box 9100
Nijmegen, 6500HB
Netherlands
Phone: 31 643914595
Email: jacqueline.deleeuw@radboudumc.nl

Abstract

Background: The introduction of health information technology (HIT) has drastically changed health care organizations and the way health care professionals work. Some health care professionals have trouble coping efficiently with the demands of HIT and the personal and professional changes it requires. Lagging in digital knowledge and skills hampers health care professionals from adhering to professional standards regarding the use of HIT and may cause professional performance problems, especially in the older professional population. It is important to gain more insight into the reasons and motivations behind the technology issues experienced by these professionals, as well as to explore what could be done to solve them.

Objective: Our primary research objective was to identify factors that influence the adoption of HIT in a sample of nurses who describe themselves as digitally lagging behind the majority of their colleagues in their workplaces. Furthermore, we aimed to formulate recommendations for practice and leadership on how to help and guide these nurses through ongoing digital transformations in their health care work settings.

Methods: In a Dutch university medical center, 10 face-to-face semi-structured interviews were performed with registered nurses (RN). Ammenwerth's FITT-framework (fit between the Individual, Task, and Technology) was used to guide the interview topic list and to formulate themes to explore. Thematic analysis was used to analyze the interview data. The FITT-framework was also used to further interpret and clarify the interview findings.

Results: Analyses of the interview data uncovered 5 main categories and 12 subthemes. The main categories were: (1) experience with digital working, (2) perception and meaning, (3) barriers, (4) facilitators, and (5) future perspectives. All participants used electronic devices and digital systems, including the electronic health record. The latter was experienced by some as user-unfriendly, time-consuming, and not supportive in daily professional practice. Most of the interviewees described digital working as "no fun at all," "working in a fake world," "stressful," and "annoying." There was a lack of general digital knowledge and little or no formal basic digital training or education. A negative attitude toward computer use and a lack of digital skills contributed to feelings of increased incompetency and postponement or avoidance of the use of HIT, both privately and professionally. Learning conditions of digital training and education did not meet personal learning needs and learning styles. A positive impact was seen in the work environment when colleagues and nurse managers were aware and sensitive to the difficulties participants experienced in developing digital skills, and when there was continuous training on the job and peer support from digitally savvy colleagues. The availability of a digital play environment combined with learning on the job and support of knowledgeable peers was experienced as helpful and motivating by participants.

Conclusions: Nurses who are digitally lagging often have had insufficient and ineffective digital education. This leads to stress, frustration, feelings of incompetency, and postponement or avoidance of HIT use. A digital training approach tailored to the

learning needs and styles of these nurses is needed, as well as an on-the-job training structure and adequate peer support. Hospital management and nurse leadership should be informed about the importance of the fit between technology, task, and the individual for adequate adoption of HIT.

(J Med Internet Res 2020;22(8):e15630) doi:10.2196/15630

KEYWORDS

qualitative research; semi-structured interview; purposive sampling; health information systems; computer user training; professional education; professional competence; registered nurses; nursing informatics

Introduction

Background

Health care has been rapidly transformed by the introduction of health information technology (HIT) [1-3]. The introduction of the electronic health record (EHR) and different eHealth devices have drastically changed the daily practice of health care professionals and the way that health care is delivered. This change will continue as robots and artificial intelligence become gradually embedded into health care [4-6]. It is generally believed that HIT adds to the safety and quality of health care and reduces morbidity and mortality. This requires broad adoption, implementation, and other changes in health care processes and structures, on an individual, national, and organizational level [7,8].

Several barriers have been identified in the implementation of HIT and the associated changes impacting different organizational levels, such as the structure of the organization, the tasks performed, the incentives given, and the way information processes are developed and organized [8,9]. The adoption of digital technology is a complex process with several influencing factors on the individual level, such as perceived ease of use and usefulness, training, helping conditions, and personality traits, as well as computer anxiety and self-efficacy [10-12]. Negative and positive emotions influence the learning process and must be acknowledged during training [13]. As such, the uptake of HIT by health professionals does not always match expectations [14,15]. There is a growing awareness and general acceptance of the need for a sociotechnical approach to the implementation of HIT, emphasizing the importance of focusing on the social aspects of HIT implementation as well as on the technical aspects of a system. The implementation process is far from linear and predictable, and the fit between work processes and tasks, information technology, and individual characteristics determines the success of implementation [9].

In the Netherlands, nurses are the largest group of registered health professionals, and the government is strongly stimulating HIT in all health care settings [7]. As such, it follows that nurses will be increasingly confronted with and involved in HIT developments. Research shows several affecting factors in the general population of nurses, such as the fact that little attention has been given to the influence of HIT on the workflow of nurses in the early stages of HIT implementation, and a lack of digital training and organizational support [16,17]. Currently, there is little research on the impact of HIT on the daily work of nurses, although it has been suggested that it has the potential to reduce health care costs and improve quality of care [18].

Until now, there have not been specific studies addressing the target group of nurses who are digitally lagging behind the majority of working nurses in the field. A literature review study [19] on issues and concerns related to the adoption and use of electronic medical records (EMR) reveals the importance of the consideration and exploration of attitudes in nurses and their personal use of information technology (IT) devices regarding EHR adoption. This study also shows that negative attitudes toward computer usage, minimal skill levels, and low levels of change readiness do not improve self-confidence in nurses regarding IT adoption. This combination of factors is more likely to be present in older nurses, suggesting that this group might need more training than others who are more comfortable around computers. Hence, the commonly applied one-size-fits-all training approach might not be effective in this situation. Another study [20], using self-assessment scales to measure computer literacy and attitudes towards computer use in registered nurses, found that 1-3% of participant scores (N=688) fell into categories representing inadequate digital skills or cyberphobia. The age of the participants in these categories was not described. Some age-correlated results were found in a study [21] with an intention-to-use survey in a group of 113 registered nurses, revealing that older nurses (ages not defined) had statistically significantly less perceived computer self-efficacy than their younger colleagues.

The aging workforce of nursing has significant implications for the near future of nursing care and health care in general, especially in relation to increasing nursing shortages [22-24]. Not only do older nurses struggle with the physical demands of nursing work, but cognitive declines are also becoming more manifest, exemplified in struggles to keep pace with paper and digital work and dealing with declines in memory. These aspects could be exacerbated by the ever-changing workplace, the speed of technological advances, and the need to continuously develop new skills. Coping with many professional demands and changes is more challenging at an older age [23,25,26]. Hence, age is viewed as a predisposing factor. Health care organizations are challenged with creating healthy working environments that stimulate change readiness and motivate nurses of all ages to continue working and become competent and digitally skilled health professionals [22,27].

In the past two decades, several studies have looked at the experiences of nurses confronted with the demands of a rapidly evolving digital health care environment [21,27-30]. Several of these studies focused on nurses' experiences with EHR implementation [21,28-30]. Research methods varied, including individual interviews, focus groups, surveys, and observations in daily practice. When individual interviews were performed,

the scope was on EHR usability and adoption. One study [28] examined EHR adoption by health care professionals by studying the way they make sense of HIT. A general finding of all these studies was that the adoption and implementation of HIT needs thorough insight and knowledge of the way different groups of health care professionals handle substantial changes in their work routines. A systematic review on the implementation of EHR in hospitals identified that organizations and their leadership should be alert to managing the balance between the technology and the work processes to improve the fit between health care professionals and the HIT they use [9].

Study Aim

There is a growing number of studies that report on the factors that influence HIT adoption in health care professionals. However, we specifically wanted to explore the experiences and needs of nurses who define themselves as digitally lagging behind the majority of their colleagues at their workplaces. This is a small and sometimes invisible group of professionals who are known to struggle with the demands of digital transformation in health care. Our aim was to uncover the views and needs of these nurses, and to deepen our understanding of the contextual and individual characteristics that affect their situation. Furthermore, we wanted to formulate recommendations for practice and leadership on how to help and guide these nurses through ongoing digital transformations in their health care work setting. We aimed to identify factors that influence the adoption of HIT in a sample of nurses who describe themselves as digitally lagging behind the majority of their colleagues at their workplaces.

Methods

Setting

The study was conducted in Radboud University Medical Center, a Dutch university hospital setting with 600 inpatient beds. In 2013, a new and fully integrated EHR was implemented hospital-wide. This hospital is one of the leading digital front runners in the Netherlands, and it is accredited with a HIMSS/EMRAM stage 7 status since 2015.

Data Collection

From November to December 2017, semi-structured interviews were conducted with 10 registered nurses. Participants were selected by purposive sampling. In August 2017, an email was sent to all nurse managers of the hospital departments asking whether they knew nurses who had difficulties working with HIT in daily practice. The nurse managers drew from information gained from regular performance interviews conducted between nurse management and individual nurse staff members, in which usage of HIT, technology, and software systems is a standard item. The nurse managers identified the nurses who met our criteria and asked them to consider taking part in our research. Those who expressed interest were contacted and informed of the details of the research and the interview. Of the nurses who were approached, 10 agreed to an interview and were included after providing written consent for study participation.

Interviews

Prior to the interview, nurses were sent an interview topic list so they could prepare themselves. Of the 10 interviews that were conducted, 9 occurred in a location outside of the hospital. The trained interviewer had no professional or private relationship with any of the participants. The interviewer made field notes during the interviews. After completing 2 interviews, the integral audio recordings of the interviews were listened to, and the topic list and interview techniques were confirmed. No further adjustments to the topic list were considered necessary. The mean duration of the interviews was 84 minutes (65 minutes minimum, 103 minutes maximum). All interviews were audio-recorded and transcribed. The manuscripts were sent to the participants for member checks and triangulation. One respondent wanted some changes made to the interview manuscript.

Interview Topic List

We identified possible themes in the current literature to explore during the interviews. A concept topic list was made, and we decided to use Ammenwerth's FITT-framework [31] as a basis for further development of the interviews. We compared our concept topic list with the different fit-axes of the framework to see if they were all broadly covered (Table 1). In addition, we discussed the topic list with several experts on information and communications technology (ICT) implementation and change management in the hospital.

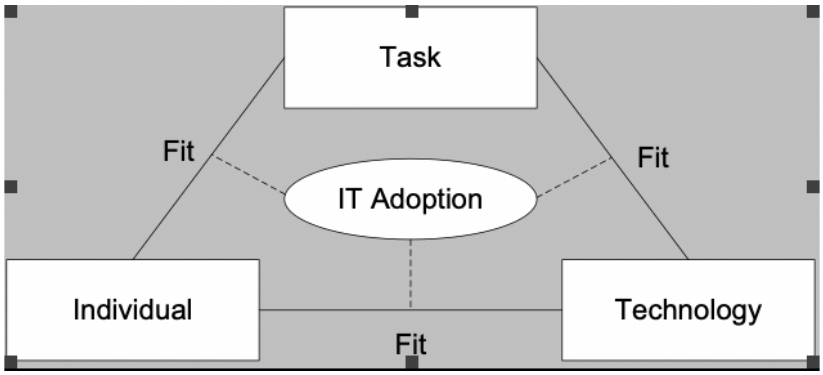
Table 1. Interview topic list compared with the FITT-framework fit-axes; FITT: fit between Individual, Task, and Technology.

INTERVIEW TOPICS	EXPLORED FIT-AXES
Digital working in health care practice	
Views and attitudes of digitization and automation in general	Individual – Technology
Views and attitudes of digitization and automation in professional practice	Individual – Technology
Experience using digital systems and devices	
Changes in daily life and professional practice after the introduction of digital systems and devices	Task – Technology
Individual learning challenges	Individual – Technology
Who or what helped in getting digital skills and competencies	Individual – Technology
Advantages, disadvantages, and pitfalls of digital working	Individual – Task
Facilitators and barriers of digital working	Individual - Technology
	Task - Technology
	Individual – Task
Outlook on the future	
Outlook towards the future of digital working and nursing practice	Task – Technology
Needs and wishes to stay connected and competent on digital developments	Individual – Technology

The FITT-framework

FITT stands for the fit between the individual, the task, and the technology. Ammenwerth developed and used this framework to evaluate the IT adoption of nurses in clinical wards by analyzing the characteristics of the users, the tasks they had to perform, and the characteristics of the technology. The framework is shown in Figure 1. Ammenwerth describes several barriers and facilitators for each of the 3 components of the framework. A poor fit might lead to frustration and, in the end, a boycott of ICT use. According to Ammenwerth, the model can help identify and analyze reasons for ICT adoption to guide the implementation process. However, it must be noted that a stable situation on the 3 fit-axes is almost impossible due to the influences of external factors, such as changing organization structures, advances in technology, and other external factors.

Figure 1. The FITT-framework of information technology (IT) adoption (Ammenwerth, 2006); FITT: fit between Individual, Task, and Technology.



Analysis

Thematic analysis was used to analyze the data [32]. The purpose of thematic analysis was to get to know the data by reading and re-reading. First, codes were generated and overarching themes were identified. Two researchers analyzed the interviews and coded the text independently of each other by assigning conceptual labels to data [33]. The 2 researchers discussed the coding on 7 occasions and discussed discrepancies until consensus was reached. A process of identifying text fragments relating to the research question was undertaken, and 183 text fragments were marked. One researcher categorized the codes about the same parts of the research question and discussed the categorization to reach consensus. This resulted in 5 main categories and 12 subthemes (Table 2). After analyzing the last 2 interviews, no additional information was presented, and saturation was reached. The interviews were analyzed and coded using the program ATLAS.ti (version 8.0, Atlas.ti Scientific Software Development GmbH) [34]. We used the COREQ (consolidated criteria for reporting qualitative research) checklist to report the results [35,36].

Table 2. Main categories, subthemes, and mapping with elements of the FITT- framework; FITT: fit between Individiaul, Task, and Technology.

Category, Subtheme	Mapping with Axes of the FITT-framework
Experience with digital working	
First experience and current use	Individual – Technology
Positive and negative experiences	Individual – Task
	Task – Technology
Perception and meaning	
Emotions, feelings	Individual – Technology
Attitude toward digital working	Individual – Technology
Barriers	
Lacking digital knowledge and skills	Individual – Task
Person-related barriers	Individual – Technology
Digital training not tailored to needs	Individual – Task
Facilitators	
Acknowledgement from management	Individual – Task
Tailored training, peer-to-peer learning	Individual – Task
Help at hand (at work and at home)	Individual – Technology
Future perspective	
General outlook to future	Individual – Technology
Future learning needs	Individual – Task

Ethics

The ethics committee of the Radboud University Medical Center waived the request to approve the study as it did not fall under the Medical Research Involving Human Subjects Act in the Netherlands.

Results

Characteristics of Participants

All 10 participants were registered nurses (RN) working in the University's medical center. Of the 10 participants, 7 were women, 3 were men, and their mean age was 56 (median 54; range 52-63). The mean duration of employment at the hospital was 28.9 years (range 20-39). Of the 10 nurses, 7 nurses worked in a nursing ward or a daycare ward, 2 worked in the outpatient clinic, and 1 combined their nursing work with administrative tasks. Table 3 presents participant characteristics.

Table 3. Characteristics of participants (N=10).

Participant Characteristic	Values
Age in years, mean (median; range)	56 (54; 52-63)
Gender, n (%)	
Male	3 (30)
Female	7 (70)
Nursing education, n (%)	
Diploma	4 (40)
Bachelor	2 (20)
Unknown	4 (40)
Years of employment at the organization, mean (median; range)	28.9 (28.9; 20-39)
Work setting, n (%)	
Non-bedside	1 (10)
Nursing ward	6 (60)
Daycare unit	1 (10)
Outpatient clinic	2 (20)
Usage of devices at home, n (%)	
Smartphone	
Used regularly[a]	9 (90)
Used rarely[b]	0 (0)
Never/no device[c]	1 (10)
PC/laptop	
Used regularly[a]	2 (20)
Used rarely[b]	5 (50)
Never/no device[c]	3 (30)
Tablet	
Used regularly[a]	2 (20)
Used rarely[b]	2 (20)
Never/no device[c]	6 (60)

[a] Regularly = at least once every day.

[b] Rarely = only now and then.

[c] Never = owns device but never uses it; no device = not in possession of device.

Category 1: Experience With Digital Working

The first experience of the participants with a computer or digital environment took place at work, at home, or at a study course. None of the participants had any intrinsic motivation or reason to start using digital systems or devices; this was always prompted by obligation or necessity. Almost half of them (4/10) were initially confronted with a computer or digital system during work for taking minutes, or registration of patient data, or other tasks (Participants 2, 4, 5, and 6).

All participants had a tablet, a laptop, or a computer at their disposal at home, and all of them used a computer professionally. If there were multiple digital devices available at home, there was always one that was preferred, but this preference varied from person to person. All but one participant (9/10) owned a smartphone. Most participants (7/10) mentioned that they checked their emails on a smartphone; however, some participants (3/10) did not know how to do so, and some (2/10) did not want to for reasons of principle. Most participants (7/10) used apps on their smartphones, like WhatsApp, a weather app, or a banking app. One participant that didn't want to use a smartphone for reasons of principle said,

> *I don't own or want to use a smartphone. I don't want to use apps or use the phone for all kinds of things. But I really feel under pressure that I still don't use*

> *and own a smartphone. I feel I really must defend and justify myself for that all the time.* [Participant 3]

Most participants (9/10) commonly used word processing as well as a search engine on the intranet or internet. The work-related applications that were mostly used were the organizational digital quality system, the EHR, and email. Rarely used or mentioned applications included Excel (used by 2 of the 10 participants), PowerPoint (used by 2 of the 10 participants), Facebook (used by 1 of the 10 participants), Skype or FaceTime (used by 2 of the 10 participants), and YouTube (used by 1 of the 10 participants). Half of the participants had typing skills.

For most of the participants (6/10), using the EHR was part of their professional practice. Reaching a basic level of proficiency with this system was not easy for them, and most still struggled. During the introduction and implementation period of the EHR, their outlook was positive, but in daily practice, they experienced the design and system functionality as user-unfriendly, time-consuming, and not supportive of the daily work process. Therefore, the EHR was experienced as a burden.

Almost all participants (9/10) said that using a computer took up valuable patient time. All were reluctant to use a computer or tablet during conversations and consultations with patients. This was perceived as an impersonal way of performing nursing care. Most of the participants (9/10) wrote their notes on paper and entered them into the system at the end of their shift rather than using the real-time documentation tools provided by the EHR. One of the participants said,

> *I think it's important that the EHR is user-friendly and intuitive, so it helps me in my work as a nurse. I often feel that I'm working for the EHR system instead of the other way around.* [Participant 4]

However, most participants (6/10) also mentioned several potential advantages of digital working, such as no more searching for paper patient records; all patient-related information in one record; easier patient handoff; better overview of the patient's situation, which improves patient safety; standardization of the process of nursing care; no problems in the readability of the handwriting; and being able to reach out to a colleague or a treatment team with 1 system click.

The participants said that social interaction and collegial conversations during breaks were hampered as everyone was busy with their smartphones. They shared that their own lack of digital skills and competencies regularly annoyed some colleagues. Two of the participants (2/10) said that they hoped for understanding and support from their colleagues (Participants 1 and 2). Several participants (5/10) said that age and aging played a role in dealing and working with digital systems. They experienced significant differences in digital skills and competencies compared to the new generation of "the digitally born." Often the participants asked their younger colleagues for help in this area.

Category 2: Perception and Meaning

The fact that digital systems play a key role in daily life and in the nursing process provoked different emotions in the participants. Most of the participants described digital working as "no fun at all," "working in a fake world," "stressful and annoying," "feeling isolated,"' "insecure," "frustrating," and "shameful."

Insecurity, anxiety, and shame were feelings mentioned by most participants (7/10). These feelings contributed to a negative self-image followed by withdrawal and the desire to become invisible (Participants 1, 2, 4, 5, 6, 7, and 8). Therefore, situations that provoked these feelings were avoided, and work was left to others; participants reported feeling trapped in a situation that was difficult to change.

> *I notice that new digital things are constantly coming up and that change is an ongoing business. Just when I think I have reached a basic skill level, a new system or functionality is there, and I feel I must start all over again. And to protect myself from feeling too stressed and frustrated, I decide to avoid that particular situation.* [Participant 10]

At times, these feelings and attitudes toward digital working were overwhelming; however, participants clearly said that they were willing to learn to become more digitally skilled out of love and passion for their nursing profession and their patients.

Category 3: Barriers

Participants mentioned several barriers to obtaining basic digital skills and competencies. There was a lack of general digital knowledge, and little or no formal digital training or education. The digital knowledge of participants was based on trial and error, and therefore, fragmented. The understanding of how ICT works was largely absent.

> *Using the Outlook agenda…I'll do it, but I really don't understand the Google and Outlook systems, and how it all works. I just don't 'understand it, and that makes me feel insecure and unsafe, as I don't know what I'm doing.* [Participant 3]

The digital language was experienced as alienating and unfamiliar. This made it difficult for participants to take part in conversations with colleagues using digital jargon. Some participants (6/10) said that the introduction of the EHR in nursing practice was a wake-up call for them to start developing their digital skills.

Several participants (4/10) talked about personal characteristics that impacted the development of digital skills and competencies. Examples included difficulty with structuring thoughts and situations, having dyslexia, and being a practical learner rather than a theoretical learner. Some nurses found it hard to ask their colleagues for help because they did not want to take up too much of others' time or were ashamed of doing so. The absence of digital help at home was reported as a shortcoming.

For the participants who had had formal digital training (including EHR training; 7/10), almost all of them (6/10) said that the content, form, and pacing of that education had not matched their personal learning needs and learning styles. Some of these participants reported too much information having been given over; a large learning group; the presence of young and

digitally savvy colleagues; a fast learning pace; content not based on the daily work process; and lack of general information about the digital system or application that was being trained for. With digital applications that were used infrequently, feelings of incompetency increased. Also, the daily workload and work stress experienced were reported as impediments to digital learning on-the-job. Hence, new digital functionality was experienced as information overload, and it increased feelings of stress and uncertainty.

Category 4: Facilitators

The participants reported several factors that could be helpful for "staying onboard digitally." These factors were related to aspects of nursing management and leadership, learning conditions in class and in the workplace, peer support and help at work, and practical digital help at home.

Participants indicated that it helped if their managers were aware of their experienced difficulties in developing digital skills. Also, knowing that other colleagues experienced similar difficulties and could be talked to about it helped to reduce mental stress. Several participants (3/10) said that managers could do more by encouraging a team culture based on collegial learning, and by facilitating and emphasizing the importance of helping each other in achieving a basic digital skills level. The concept of training on-the-job was viewed as highly effective by participants. In addition, enough time to learn and to repeat skills was viewed as an essential success factor.

> *I talked to my manager that I got stuck […] that I had difficulties using the EHR. After that, I was linked to one of our senior nurses and EHR-superuser during day shifts, and I could ask her help on EHR topics and nursing documentation whenever I wanted. That was really helpful.* [Participant 4]

Participants reported several key learning elements that were crucial to their digital learning needs. These elements were a small learning group of similarly skilled people; clear instructions parsed into learning steps; a lot of recurrent rehearsal of the learning content; practice- and process-based learning; a digital learning environment for practice; and ample time allotted for practice.

Category 5: Future Perspectives

Participants were asked how they viewed the future and what they needed to keep up with digital developments. Some of them said that concerns about the digital future were not that important to them, and some hoped that developments would pass by or stop at a certain moment (3/10; Participants 2,5,9). Some participants (2/10) were worried about their longevity in the nursing profession, and one person thought about looking for another job that was less "digitally infused" (Participant 8).

Some participants (4/10) reported that there had already been several ongoing digital developments at their workplaces but that they had reservations about using them. Some examples were patient-provider communication via the patient portal, teleconsulting, checking patient vitals with wearables (also at home), and bedside computers for patients. Some participants (5/10) believed that future hospital and nursing care would change enormously due to these developments.

At the same time, several participants (3/10) expressed a wish to be able to keep up with digital developments and to become more skillful. One participant expected improved digital skills to result in more working pleasure and satisfaction. A strategy mentioned was to use one's practical knowledge of daily workflows to get more involved in digital development projects.

> *If I hear that they are going to develop something new (digital applications), I would love to be involved, just to tell how it might be practical to work with.* [Participant 2]

Participants also mentioned that their organizations could help by defining a basic digital skills level for employees, and by providing a monitoring and examining system for defining digital skills levels. Participants were divided in their opinions regarding their own initiative and investment in becoming digitally skilled. Some (3/10) said that it was the responsibility of employees to invest in their own digital education. But most of them (7/10) would appreciate practical help in finding the right courses tailored to their digital learning needs.

Discussion

Main Findings

In this paper, we present the results of an in-depth interview study to explore and identify the experiences and needs of nurses who consider themselves to be lagging in digital competency. The study was performed to deepen our understanding of the contextual and individual characteristics that affect the situation. Furthermore, we wanted to formulate recommendations for practice, management, professional education, and research.

In general, the nurses in our study can be characterized as late adopters of technology in their personal lives and in the workplace. Some of them were averse to IT and technology. In learning to use digital devices and systems (for example, an electronic patient record), the training and learning conditions were not tailored to their needs. The learning groups were too large, and the pace of learning was too fast, with insufficient time for repetition. As late adopters of technology and digital working, the participants felt that they were not given enough time to learn on-the-job, as well as insufficient support from peers. All this added to their stress, frustration, and feelings of incompetency, resulting in a tendency to postpone or avoid the use of HIT in daily nursing practice. When interpreting the research findings in the context of the 3 dimensions of the FITT-framework, there seemed to be a suboptimal fit on all 3 dimensions. Organizational interventions and measures should focus on all 3 dimensions because the balance between them affects IT adoption.

The Context of the FITT-framework

The FITT-framework approaches IT adoption by looking at the fit between the individual, the task, and the technology. An optimal fit between the 3 framework dimensions allows for easy IT adoption. Presumably, the larger the difference between the actual fit and the planned fit, the higher number of problems that may occur during implementation of IT systems [32]. There are 3 relations of fits within this framework: (1) a fit between the individual and the technology, (2) between the individual

and the task, and (3) between the task and the technology. This framework helped us categorize interview data; however, it also showed us that it is sometimes difficult to make clear distinctions between the fit dimensions, since they are not mutually exclusive but are partly intertwined and have certain overlaps[31].

In our analysis of the fit between the individual and technology, we found our group of participants to be largely incompetent with computers, electronic devices, and software. They exhibited little enthusiasm for learning about technology or systems, and this seemed to be related to the level of digitization in the participants' personal lives. Furthermore, this group expressed reservations about utilizing an electronic device (like a computer, tablet, or smartphone) during nursing care activities. Acceptance of IT as a professional tool needs attention in this group; its use was perceived as alienating and "bad" patient-centered care (this is also a widely held opinion amongst some elderly nurses, as it goes against what they had been taught about "good nursing" [21]). This can be interpreted as a poor fit in this dimension. However, despite the intrinsic resistance to technology and systems, there is a general willingness to learn how to use basic HIT systems to keep up with the current professional and organizational developments. This presents a challenge in change management for nurse management and leadership.

In our analysis of the fit between the individual and task, we found that, although participants did mention several advantages to HIT, these advantages were not integrated into their professional competencies and daily nursing activities. The nurses struggled with changes in their work and professional collaboration with nursing colleagues due to digitalization. For example, the use of IT jargon in daily practice was reported as alienating from core nursing care language. It also negatively impacted collegiality, collaboration, and team cohesion while elevating feelings of incompetency. As the work pace is consistently high in the nursing profession, it is difficult for the participants in our study to make the required professional changes as nursing care becomes more complex and digitally infused. This issue presents a significant concern in light of the growing nursing shortages experienced in many countries all over the world [22,30]. Moreover, elderly nurses are the clinical teachers for future nursing professionals. Therefore, a possible solution lies in emphasizing the mutual learning benefits that arise from collaboration between elderly nurses and young professionals, in which the latter group teaches their mentors about digital nursing.

In our analysis of the fit between the task and technology, we found that the technology must offer sufficient functionality and performance to support a nursing task. The results of our study show a poor fit in the dimension of task-technology. For example, some participants experienced the EHR as user-unfriendly and unsuitable to the nursing task. This finding was also present in a panel study among Dutch nurses [37]. On the other hand, some participants said that after instruction and daily support from colleagues, they began to understand how the nursing process (labeled by them as a "task") was built into the EHR. Previously accustomed to a manual method of charting and documenting, this had previously gone unrecognized and misunderstood. This is a common phenomenon among elderly nurses who must make a change from manual to digital charting [21].

Implications for Practice

Several facilitating conditions seem to be of influence on IT adoption in this group of nurses digitally lagging behind the majority of their colleagues at their workplaces. The realization and interpretation of these conditions are the main implications for practice.

First, learning conditions and needs for this specific group differ from the average population. Carefully assembled training, tailored to learning styles and learning needs, is key. The content of any IT training must follow and reflect the daily work process and must be accompanied by learning on-the-job, peer support, and an adequate digital training and play environment. Second, the effect of social influences must not be underestimated when learning on-the-job and when environmental changes are occurring. Our cohort of nurses expressed sensitivity to social pressures to use particular technology or systems [21]. Consequently, nurse management could make use of this fact, for example, by appointing digital nurse role models who can act as digital coaches for their colleagues [21,28,29]. Third, some participants suggested that they should be involved in digital developments and innovations at early stages, as this would allow them to give advice about how technology and systems could be tailored to daily practices and the work process, and it would allow them to give input on the content of instructional learning materials. Health care provider participation in early-stage IT development has been suggested in other studies on IT adoption [9,27,30,31]. Lastly, nurse managers should be aware of their roles in improving the quality of the fit between the 3 fit dimensions, and be aware that it is not only individual attributes that are important. Mindfulness of the digital impact on the therapeutic relationship between nurses and patients is necessary [38]. This awareness makes fit management a constant, complex task for all who are involved with IT adoption and digital transformation. The patient, rather than the technology, must remain the central focus, and health care providers must be connected to IT personnel to ensure that the fit in the 3 dimensions is in scope from the very beginning of IT development. This will not only encourage the adoption of digital systems but will also improve ownership in any user group, which is necessary to optimize systems after they go live.

Following the research recommendations of our study, our hospital organization made several changes regarding the educational concept, form, and content of digital training and education. The component of training on-the-job and structured peer support has now been newly introduced and emphasized in daily practice.

Implications for Research

This study explored the experiences, views, and needs of digitally lagging nurses with the use of in-depth interviews. The inclusion of participants was based on the observations of nurse managers and their individual conversations with their nursing personnel.

Our study identified a group of nurses who lack basic IT knowledge and skills, and consequently have a low sense of self-efficacy and self-confidence with computers and technology use. This group perceives itself to be partly invisible and sometimes ignored (although not intentionally) because they have no common face. More research regarding this group, their profile, characteristics, and learning needs is justified, with the ultimate goal of retaining them in the nursing profession and in their organizations. Expanding this study to other health care organizations, such as general hospitals, nursing homes, or primary care organizations, may corroborate our findings.

In addition, it would be worthwhile to find out if this phenomenon exists among groups of physicians and allied health care workers. A qualitative approach would be appropriate when starting to explore views, needs, and attitudes. Another line of investigation could target management and leadership of health care organizations (eg, is there any awareness regarding this group? How are these individuals identified?). Furthermore, it may be useful to expand this research to other health care organizations, like other non-university hospitals.

We did not use any quantitative instruments to measure digital competence or attitudes toward computers use. For future research, we recommend using an additional quantitative questionnaire to assess digital competency. This would complement the results from the interview data and enhance data triangulation. Another point of attention for future research is to include the age factor of participants in the analysis of quantitative results. This would allow researchers to describe relations and correlations between age and digital competency that can provide further rationale for formulating appropriate training approaches for the target group of nurses digitally lagging, as well as other applications.

There is ongoing research regarding the use and validation of the FITT-framework of Ammenwerth. A recent publication by Prgomet [39], evaluating clinicians' use of computing devices and identifying factors affecting the optimal use of those devices, proposes an extension of the FITT-framework with the added component of "environment" to explicate the relationships between individuals, tasks, technology, and the environment in which they operate. Additional factors relating to the environment were found to affect the use of technology. These environmental attributes included, for example, physical environment, department type, organizational policies, and procedures. It would be interesting to explore these environmental aspects in further studies. In our study, we used the FITT-framework to guide the analysis of interview data. We think the framework also has the potential to be applied during HIT implementation by modifying implementation activities to the fit of the 3 axes of the model.

Study Strengths and Limitations

One of the strengths of our study was that we succeeded in identifying and including 10 nurses digitally lagging behind. Our hospital applies a policy of regular performance interviews between nurse managers and nurses, which include evaluations on job satisfaction, performance, and professional ambitions. This contributed to the identification and inclusion of our study participants. However, the group of participants might have a limited perspective on HIT, which might have been reflected in their answers to the interview questions.

Another strength is that all participants expressed they felt safe and at ease with the interviewer to share feelings and opinions, and to speak freely about worries and experiences around IT technology and system use. Therefore, the material we collected was rich in content.

A limitation of our study is that we cannot generalize the results to other groups or to the general population, as it is a single case study in one Dutch university hospital. Despite the limited generalizability, the results of our study can inform health care organizations and care managers to provide focused attention and help for this specific target group of nurses.

Conclusion

Our study explored the experiences and views of nurses who are digitally lagging. Although a fair amount of studies are available on factors influencing IT adoption among nurses, we could not find any studies targeting this particular group. The findings of this study support the assumption that this group of nurses could benefit from a tailored approach to digital training and education. The generic one-size-fits-all strategy does not suit their learning needs and styles. More research is needed to identify this target group in more detail in order to tailor interventions to their needs and goals of becoming competent digital nurses.

Acknowledgments

This study was funded by the Netherlands Federation of University Medical Centers' Citrien Fund eHealth program.

Authors' Contributions

HW contacted participants, informed them of the details of the research, and conducted the interviews. HW created a concept topic list. RBK and JDL listened to the integral audio recordings of 2 interviews, and RBK, JDL, and HW then confirmed the topic list and interview techniques. HW analyzed the interviews, coded the text, and categorized the codes after reaching consensus with RBK and JDL.

Conflicts of Interest

None declared.

References

1. Christensen C, Grossman J, Hwang J. The Innovator's prescription: a disruptive solution to the health care. New York: McGraw Hill; 2009:a.
2. Gartner. How Healthcare Providers Can Succeed In Digital Transformation. 2018. URL: https://www.gartner.com/smarterwithgartner/how-to-succeed-in-healthcare-digital-transformation/ [accessed 2019-11-02]
3. McKinsey. McKinsey on Healthcare. web page. 2019. URL: https://healthcare.mckinsey.com/digital [accessed 2020-10-30]
4. Buch VH, Ahmed I, Maruthappu M. Artificial intelligence in medicine: current trends and future possibilities. Br J Gen Pract 2018 Dec;68(668):143-144 [FREE Full text] [doi: 10.3399/bjgp18X695213] [Medline: 29472224]
5. PWC G. AI and robotics are transforming healthcare. 2019. URL: https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/transforming-healthcare.html [accessed 2019-11-06]
6. Safdari R, Ghazisaeidi M, Jebraeily M. Electronic health records: critical success factors in implementation. Acta Inform Med 2015 Apr;23(2):102-104 [FREE Full text] [doi: 10.5455/aim.2015.23.102-104] [Medline: 26005276]
7. Dutch Ministry of Health (VWS). Health IT Safety. 2017. URL: https://healthhit.gov/topic/health-it-safety [accessed 2019-06-11]
8. Taylor R, Bower A, Girosi F, Bigelow J, Fonkych K, Hillestad R. Promoting health information technology: is there a case for more-aggressive government action? Health Aff (Millwood) 2005;24(5):1234-1245. [doi: 10.1377/hlthaff.24.5.1234] [Medline: 16162568]
9. Boonstra A, Versluis A, Vos JFJ. Implementing electronic health records in hospitals: a systematic literature review. BMC Health Serv Res 2014;14:370 [FREE Full text] [doi: 10.1186/1472-6963-14-370] [Medline: 25190184]
10. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
11. Lluch M. Healthcare professionals' organisational barriers to health information technologies-a literature review. Int J Med Inform 2011 Dec;80(12):849-862. [doi: 10.1016/j.ijmedinf.2011.09.005] [Medline: 22000677]
12. Strudwick G. Predicting nurses' use of healthcare technology using the technology acceptance model: an integrative review. Comput Inform Nurs 2015 May;33(5):189-98; quiz E1. [doi: 10.1097/CIN.0000000000000142] [Medline: 25974361]
13. Kay RH. Exploring the relationship between emotions and the acquisition of computer knowledge. Computers & Education 2008 May;50(4):1269-1283. [doi: 10.1016/j.compedu.2006.12.002]
14. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
15. Westbrook JI, Braithwaite J. Will information and communication technology disrupt the health system and deliver on its promise? Med J Aust 2010 Oct 04;193(7):399-400. [Medline: 20919970]
16. Assis-Hassid S, Grosz BJ, Zimlichman E, Rozenblum R, Bates DW. Assessing EHR use during hospital morning rounds: A multi-faceted study. PLoS One 2019;14(2):e0212816 [FREE Full text] [doi: 10.1371/journal.pone.0212816] [Medline: 30802267]
17. Eley R, Fallon T, Soar J, Buikstra E, Hegney D. The status of training and education in information and computer technology of Australian nurses: a national survey. J Clin Nurs 2008 Oct;17(20):2758-2767. [doi: 10.1111/j.1365-2702.2008.02285.x] [Medline: 18808645]
18. While A, Dewsbury G. Nursing and information and communication technology (ICT): a discussion of trends and future directions. Int J Nurs Stud 2011 Oct;48(10):1302-1310. [doi: 10.1016/j.ijnurstu.2011.02.020] [Medline: 21474135]
19. Parthasarathy R, Steinbach T, Knight J, Knight L. Framework to Enhance Nurses' Use of EMR. Hosp Top 2018;96(3):85-93. [doi: 10.1080/00185868.2018.1488545] [Medline: 30277450]
20. Gürdaş Topkaya S, Kaya N. Nurses' computer literacy and attitudes towards the use of computers in health care. Int J Nurs Pract 2015 May;21 Suppl 2:141-149. [doi: 10.1111/ijn.12350] [Medline: 24804813]
21. Holtz B, Krein S. Understanding Nurse Perceptions of a Newly Implemented Electronic Medical Record System. Journal of Technology in Human Services 2011 Oct;29(4):247-262. [doi: 10.1080/15228835.2011.639931]
22. Bell LM. The ageing of the nursing workforce: what lies ahead and what we can do. Int Nurs Rev 2013 Sep;60(3):277-278. [doi: 10.1111/inr.12049] [Medline: 23961785]
23. Ryan C, Bergin M, Wells JS. Valuable yet Vulnerable-A review of the challenges encountered by older nurses in the workplace. Int J Nurs Stud 2017 Jul;72:42-52. [doi: 10.1016/j.ijnurstu.2017.04.006] [Medline: 28456111]
24. Uthaman T, Chua TL, Ang SY. Older nurses: A literature review on challenges, factors in early retirement and workforce retention. In: Proceedings of Singapore Healthcare. 2015 Oct 16 Presented at: Proceedings of Singapore Healthcare ;25(1); 2016; Singapore p. 50-55. [doi: 10.1177/2010105815610138]
25. Ang SY, Ayoob SBM, Hussain NBS, Uthaman T, Adenan H, Chiang P, et al. Challenges faced by older nurses in Singapore: a mixed methods study. Int Nurs Rev 2017 Dec;64(4):502-510. [doi: 10.1111/inr.12348] [Medline: 28093735]
26. Durosaiye IO, Hadjri K, Liyanage CL. Identifying Challenging Job and Environmental Demands of Older Nurses Within the National Health Service. HERD 2016 Apr;9(3):82-105. [doi: 10.1177/1937586715613586] [Medline: 26578540]

27. Mather C, Cummings E, Gale F. Nurses as Stakeholders in the Adoption of Mobile Technology in Australian Health Care Environments: Interview Study. JMIR Nursing 2019 Aug 09;2(1):e14279. [doi: 10.2196/14279]
28. Jensen TB, Aanestad M. How Healthcare Professionals "Make Sense" of an Electronic Patient Record Adoption. Information Systems Management 2006 Dec 22;24(1):29-42. [doi: 10.1080/10580530601036794]
29. Markham RB, Stashak PW, Prescott B, Amsbaugh DF, Baker PJ. Generation of low-dose paralysis in the absence of the ability to secrete antibody. J Immunol 1978 Mar;120(3):986-990. [Medline: 24665]
30. Vehko T, Hyppönen H, Puttonen S, Kujala S, Ketola E, Tuukkanen J, et al. Experienced time pressure and stress: electronic health records usability and information technology competence play a role. BMC Med Inform Decis Mak 2019 Aug 14;19(1):160 [FREE Full text] [doi: 10.1186/s12911-019-0891-z] [Medline: 31412859]
31. Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Med Inform Decis Mak 2006;6:3 [FREE Full text] [doi: 10.1186/1472-6947-6-3] [Medline: 16401336]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Corbin J, Strauss A. Grounded theory research: Procedures, canons, and evaluative criteria. Qual Sociol 1990;13(1):3-21 [FREE Full text] [doi: 10.1007/bf00988593]
34. Sandelowski M. The Sage Encyclopedia Of Qualitative Research Methods. The Sage encyclopedia of qualitative research methods: Sage Publications, Inc; 2019.
35. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
36. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
37. de VAJE, Fleuren MAH, Bekkema N, Francke AL. Successful implementation of new technologies in nursing care: a questionnaire survey of nurse-users. BMC Med Inform Decis Mak 2011;11:67 [FREE Full text] [doi: 10.1186/1472-6947-11-67] [Medline: 22032728]
38. Burkoski V. Nursing Leadership in the Fully Digital Practice Realm. Nurs Leadersh (Tor Ont) 2019 May;32(SP):8-15. [doi: 10.12927/cjnl.2019.25818] [Medline: 31099743]
39. Prgomet M, Georgiou A, Callen J, Westbrook J. Fit Between Individuals, Tasks, Technology, and Environment (FITTE) Framework: A Proposed Extension of FITT to Evaluate and Optimise Health Information Technology Use. Stud Health Technol Inform 2019 Aug 21;264:744-748. [doi: 10.3233/SHTI190322] [Medline: 31438023]

Abbreviations

EHR: electronic health record
EMR: electronic medical records
FITT: Fit between Individuals, Task, and Technology
HIT: health information technology
ICT: information and communications technology
IT: information technology

Edited by G Eysenbach; submitted 25.07.19; peer-reviewed by G Strudwick, S Thiebes; comments to author 03.10.19; revised version received 13.01.20; accepted 25.06.20; published 14.08.20.

Please cite as:
De Leeuw JA, Woltjer H, Kool RB
Identification of Factors Influencing the Adoption of Health Information Technology by Nurses Who Are Digitally Lagging: In-Depth Interview Study
J Med Internet Res 2020;22(8):e15630
URL: http://www.jmir.org/2020/8/e15630/
doi:10.2196/15630
PMID:32663142

Original Paper

Rejected Online Feedback From a Swiss Physician Rating Website Between 2008 and 2017: Analysis of 2352 Ratings

Stuart McLennan[1,2], MBHL, PhD

[1]Institute of History and Ethics in Medicine, Technical University of Munich, Munich, Germany

[2]Institute for Biomedical Ethics, University of Basel, Basel, Switzerland

Corresponding Author:
Stuart McLennan, MBHL, PhD
Institute of History and Ethics in Medicine
Technical University of Munich
Ismaninger Straße 22
Munich, 81675
Germany
Phone: 49 89 4140 4041
Email: stuart.mclennan@tum.de

Abstract

Background: Previous research internationally has only analyzed publicly available feedback on physician rating websites (PRWs). However, it appears that many PRWs are not publishing all the feedback they receive. Analysis of this rejected feedback could provide a better understanding of the types of feedback that are currently not published and whether this is appropriate.

Objective: The aim of this study was to examine (1) the number of patient feedback rejected from the Swiss PRW Medicosearch, (2) the evaluation tendencies of the rejected patient feedback, and (3) the types of issues raised in the rejected narrative comments.

Methods: The Swiss PRW Medicosearch provided all the feedback that had been rejected between September 16, 2008, and September 22, 2017. The feedback were analyzed and classified according to a theoretical categorization framework of physician-, staff-, and practice-related issues.

Results: Between September 16, 2008, and September 22, 2017, Medicosearch rejected a total of 2352 patient feedback. The majority of feedback rejected (1754/2352, 74.6%) had narrative comments in the German language. However, 11.9% (279/2352) of the rejected feedback only provided a quantitative rating with no narrative comment. Overall, 25% (588/2352) of the rejected feedback were positive, 18.7% (440/2352) were neutral, and 56% (1316/2352) were negative. The average rating of the rejected feedback was 2.8 (SD 1.4). In total, 44 subcategories addressing the physician (n=20), staff (n=9), and practice (n=15) were identified. In total, 3804 distinct issues were identified within the 44 subcategories of the categorization framework; 75% (2854/3804) of the issues were related to the physician, 6.4% (242/3804) were related to the staff, and 18.6% (708/3804) were related to the practice. Frequently mentioned issues identified from the rejected feedback included (1) satisfaction with treatment (533/1903, 28%); (2) the overall assessment of the physician (392/1903, 20.6%); (3) recommending the physician (345/1903, 18.1%); (4) the physician's communication (261/1903, 13.7%); (5) the physician's caring attitude (220/1903, 11.6%); and (6) the physician's friendliness (203/1903, 10.6%).

Conclusions: It is unclear why the majority of the feedback were rejected. This is problematic and raises concerns that online patient feedback are being inappropriately manipulated. If online patient feedback is going to be collected, there needs to be clear policies and practices about how this is handled. It cannot be left to the whims of PRWs, who may have financial incentives to suppress negative feedback, to decide which feedback is or is not published online. Further research is needed to examine how many PRWs are using criteria for determining which feedback is published or not, what those criteria are, and what measures PRWs are using to address the manipulation of online patient feedback.

(J Med Internet Res 2020;22(8):e18374) doi:10.2196/18374

KEYWORDS

physician rating websites; patient satisfaction; participatory medicine; patient feedback

Introduction

There remains relevant unwarranted variation in health care systems and deficiencies regarding all key aspects of health care [1]. Members of the public, however, have traditionally had few ways of knowing who the "good" health care organizations and professionals are [2]. As part of a wider move toward transparency, public reporting activities have been developed in a number of countries with the aim of providing information about health care organizations or professionals to the public to correct this asymmetry of information in order to inform patient decision-making and drive quality improvement [3-6].

One type of public reporting activity that has been developed in recent decades is physician rating websites (PRWs) [7-10]. PRWs represent a "bottom-up" approach to public reporting, allowing users to post ratings and comments regarding their physician as a source of information for others [11-14]. Although patients have always been able to share their opinions about their physicians with others, the ability to share these opinions via the internet and social media now means that these opinions have the potential to reach a far wider audience. With a growing number of patients utilizing the internet in relation to health care [15], it is expected that PRWs will play an increasingly important role.

A recent systematic search of PRWs internationally analyzed 143 different websites from 12 countries [16] and found that the majority (76.9%) of websites provided options to give feedback both on a predefined quantitative rating scale and as narrative comments. Previous research internationally has often focused on analyzing the ratings and comments publicly available on PRWs. This research has reported that many PRWs have incomplete lists of physicians, a low number of physicians rated, and a low number of ratings per physician that are overwhelmingly positive, which has raised concerns about the representativeness, validity, and usefulness of information on PRWs [14,17]. Furthermore, the medical profession has often expressed concerns that feedback on PRWs will be manipulated for "doctor bashing" or defamation [10].

The first PRWs in Switzerland were established in 2008, at the same time as many international PRWs. However, in comparison with other countries that have established PRWs, there has been limited research conducted on PRWs in Switzerland. This author recently conducted a study involving a random stratified sample of 966 physicians generated from the regions of Zürich and Geneva [18,19]. Selected physicians were searched on a total of four websites (OkDoc, Medicosearch, DocApp, and Google) between November 2017 and July 2018, and it was recorded whether the physician could be found. Moreover, the physician's rating, the number of ratings and narrative comments, and the text of narrative comments were recorded. As far as the author is aware, this was the first inclusion of Google in a study examining physician ratings internationally.

With regard to the frequency of quantitative ratings and narrative comments on Swiss PRWs, similar issues as those identified in the international literature were found. Many of the selected physicians could not be identified (the proportion of physicians who could be identified ranged from 42.4% on OkDoc to 87.3% on DocApp), few of the identifiable physicians had been rated quantitatively (4.5% on DocApp to 49.8% on Google) or received a narrative comment (4.5% on DocApp to 31.2% on Google) at least once, rated physicians had, on average, a low number of quantitative ratings (1.47 ratings on OkDoc to 3.74 rating on Google) and narrative comments (1.23 comments on OkDoc to 3.03 comments on Google), and all three websites that allowed quantitative ratings had very positive average ratings on a 5-star rating scale (DocApp, 4.71; Medicosearch, 4.69; and Google, 4.41) [18].

With regard to the contents of narrative comments, it was found that the selected physicians had a total of 849 comments [19]. Narrative comments were analyzed and classified according to a theoretical categorization framework previously developed by Emmert et al [10]. In total, 43 subcategories addressing the physician, staff, and practice were identified. None of the PRWs' comments covered all 43 subcategories of the categorization framework; comments on Google covered 86% of the subcategories, those on Medicosearch covered 72%, those on DocApp covered 60%, and those on OkDoc covered 56%. In total, 2441 distinct issues were identified within the 43 subcategories of the categorization framework; 83.65% of the issues were related to the physician, 6.63% were related to the staff, and 9.70% were related to the practice. Overall, 95% of the subcategories of the categorization framework and 81.60% of the distinct issues identified were concerning aspects of performance (interpersonal skills of the physician and staff, infrastructure, and organization and management of the practice) considered assessable by patients [19]. Furthermore, this research raised concerns that user feedback is being suppressed by Swiss PRWs [18,19], which risks undermining the overall aim of PRWs of providing a reliable source of unbiased information regarding patients' experiences and satisfaction with physicians.

As far as this author is aware, previous research internationally has only analyzed publicly available feedback on PRWs. However, as it appears that many PRWs are not publishing all the feedback they receive. Analysis of this rejected feedback could provide a better understanding of the types of feedback that are currently not published and whether this is appropriate. This study therefore aimed to examine (1) the number of patient feedback rejected from the Swiss PRW Medicosearch, (2) the evaluation tendencies of the rejected patient feedback, and (3) the types of issues raised in the rejected narrative comments. Gaining a better understanding of feedback rejected by PRWs may help to identify issues in the way PRWs are currently determining which feedback are and are not to be publicly published.

Methods

Sample

Switzerland is a Central European country with a population of about 8.4 million people and 4 official languages (German, French, Italian, and Romansh). The Swiss health care system is highly complex and decentralized, and all Swiss residents are required to purchase basic mandatory health insurance that is offered by competing nonprofit insurers. Mandatory health

insurance covers most general practitioner and specialist services, and people not enrolled in managed care plans generally have free choice of professionals [18]. The first PRWs in Switzerland, OkDoc and Medicosearch, were established in 2008. A systematic web-based search conducted in June 2016 identified that the websites DocApp and Google also allow users to view quantitative ratings and/or narrative comments about Swiss physicians in a structured manner without having to open an account or log onto the website [18,19]. It appears that other websites have also subsequently started to allow users to view quantitative ratings and/or narrative comments about Swiss physicians (eg, DeinDoktor and Doctena) [19]. Nevertheless, out of the dedicated Swiss PRWs, Medicosearch appears to be one of the best established and used [18,19]. Medicosearch allows users to search for physicians by location and specialty. Physician profiles provide general information about the physician (specialties, languages spoken, and contact details). In recent years, Medicosearch has shifted its business strategy toward online appointments, where physicians pay a fee and their booking systems are integrated with Medicosearch, allowing patients to book an appointment with a physician directly on Medicosearch. Users can also leave reviews of physicians, but Medicosearch requires both a quantitative rating (5-star rating scale) and a narrative comment in every patient feedback. Although Medicosearch allows negative comments, it informs the concerned physician before publishing it on the website, so that the physician can decide whether to activate the negative feedback function. If the physician refuses, the feedback function is deactivated, removing positive comments as well [19].

As part of a larger project examining Swiss PRWs, the author approached the CEO of Medicosearch, Beat Burger. Discussions confirmed that Medicosearch does not publish all the feedback they receive. The author enquired about the possibility of receiving these rejected feedback for analysis. Medicosearch agreed to provide the rejected feedback to the author in an anonymous form. On October 24, 2017, Medicosearch sent the author an excel file that included all the feedback that Medicosearch had rejected between September 16, 2008, and September 22, 2017. The details of the rated physicians were not included. Medicosearch did not provide any reasons for why the feedback were rejected. The following data were imported into a Statistical Package for the Social Sciences (SPSS version 26 for Windows, IBM Corporation) file: the date the feedback was created, the quantitative rating out of 5, and the narrative comments provided under “title” and “description.” In early 2019, the author inquired with Medicosearch whether it would be possible to receive the rejected feedback from September 23, 2017, to December 31, 2018. Medicosearch told the author on April 11, 2019, that owing to new data-protection rules, Medicosearch had deleted all rejected feedback, and therefore, it was not possible to provide an updated file.

Data Analysis

Medicosearch uses a 5-star rating scale; a rating of 4 to 5 stars was considered a positive rating, 3 stars was considered a neutral rating, and 1 to 2 stars was considered a negative rating. The content of each narrative comment was analyzed and classified by the author according to a theoretical categorization framework of physician-, staff-, and practice-related issues. The categorization framework from Emmert et al was initially used [10], with modifications where necessary. This included removing categories that were not identified in the comments, adding categories that were identified but were not adequately covered by the previous framework, and separating categories (eg, friendliness and caring attitude) that were discussed in comments as distinct issues. Narrative comments were analyzed in their original language. Descriptive statistics included means and standard deviations for continuous variables and percentages for categorical variables. To analyze whether differences existed between German and French comments, chi-squared tests were used. All analyses were performed with the significance level α set to .05 and two-tailed tests, using SPSS version 26.

Results

Characteristics of Ratings

Between September 16, 2008, and September 22, 2017, Medicosearch rejected a total of 2352 patient feedback (Table 1).

The majority of feedback rejected (1754/2352, 74.6%) had narrative comments in German (Table 2). Other rejected feedback had narrative comments in French (275/2352, 11.7%), Italian (31/2352, 1.3%), English (12/2352, 0.5%), and Spanish (1/2352, 0.04%). However, 11.9% (279/2352) of the rejected feedback only provided a quantitative rating with no narrative comment.

Overall, 25% (588/2352) of the quantitative ratings were positive, 18.7% (440/2352) were neutral, and 56% (1316/2352) were negative (Table 3). Additionally, the average rating of the rejected feedback was 2.8 (SD 1.4).

Table 1. Distribution of rejected feedback according to year.

Distribution of comments (year)[a] (N=2352), n (%)									
2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
26 (1.1)	259 (11.0)	232 (9.9)	344 (14.6)	392 (16.7)	321 (13.6)	268 (11.4)	142 (6.0)	236 (10.0)	132 (5.6)

[a]From September 16, 2008, to September 22, 2017.

Table 2. Distribution of rejected feedback according to language.

Language (N=2352), n (%)					
German	French	Italian	English	Spanish	Missing
1754 (74.6)	275 (11.7)	31 (1.3)	12 (0.5)	1 (0.04)	279 (11.9)

Table 3. Quantitative rating evaluation results.

Measure	Language[a]						Total (N=2352), n (%)
	German (N=1754), n (%)	French (N=275), n (%)	Italian (N=31), n (%)	English (N=12), n (%)	Spanish (N=1), n (%)	Missing (N=279), n (%)	
Evaluation							
Positive	399 (22.7)	81 (29.5)	4 (12.9)	4 (33.3)	1/1 (100)	99 (35.5)	588 (25.0)
Neutral	296 (16.9)	59 (21.5)	6 (19.4)	1 (8.3)	0 (0)	77 (26.3)	440 (18.7%)
Negative	1065 (60.4)	135 (49.1)	21 (67.7)	7 (58.3)	0 (0)	88 (30)	1316 (56%)
Missing	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	8 (2.9)	8 (0.3)
Average rating (SD)	2.7 (1.3)	3.0 (1.4)	2.4 (1.3)	2.9 (1.5)	5.0	3.2 (1.4)	2.8 (1.4)

Of the 2073 ratings that provided narrative comments, analysis found that a total of 170 comments were not feedback from a patient concerning the physician, staff, or practice (92 comments were not comprehensible, 29 comments were explicitly labelled as test ratings, 15 comments were about the PRW, 10 comments reported that the person had not visited the physician yet, 10 comments simply reported that the physician's details were not up to date, eight comments were abusive, four comments were second-hand reports, two comments were asking for advice regarding their or their family member's condition, and two comments were not concerning a visit to a physician). Consequently, this feedback was excluded from the categorization framework.

The 1903 included narrative comments had a mean length of 158 characters (SD 214), ranging from 1 to 2788 characters. There was a significant difference in the mean character length between positive comments (mean 88, SD 130) and negative comments (mean 193, SD 241) (t_{1314}=−11, P<.001). There was no significant difference in the mean character length between German comments (mean 158, SD 205) and French comments (mean 154, SD 206) (t_{1862}=0.2, P=.82).

Categorization of Issues

Content analysis of the included 1903 narrative comments identified 44 subcategories addressing the physician (n=20), staff (n=9), and practice (n=15) (Textbox 1).

Textbox 1. Categorization framework.

Physician (n=20)

Satisfaction with treatment; overall assessment; recommendation; communication; caring attitude; friendliness; competence; treatment cost/billing; being taken seriously; time spent with patient; trust; professionalism; cooperation with medical specialists; alternative medicine; telephone availability; privacy; health insurance differentiation; patient involvement; individualized service; child friendliness

Staff (n=9)

Friendliness; overall assessment; service/assistance; communication; professionalism; availability by telephone; time spent with patient; health insurance differentiation; trust

Practice (n=15)

Overall assessment; waiting time within the practice; atmosphere; organization; ability to get appointment; equipment; recommendation; consultation hours; location; waiting room entertainment; parking space; availability by telephone; privacy; barrier-free access; online appointment

In total, 3804 distinct issues were identified within the 44 subcategories of the categorization framework; 75% (2854/3804) of the issues were related to the physician, 6.4% (242/3804) were related to the staff, and 18.6% (708/3804) were related to the practice (Table 4). The most frequent issue mentioned in the rejected comments was satisfaction with treatment (533/1903, 28%); 73.2% (390/533) of these ratings were negative. Other frequently mentioned issues regarding the physician were as follows: 20.6% (392/1903) of comments provided an overall assessment of the physician (53.8% negative); 18.1% (345/1903) provided a recommendation regarding the physician (76.2% negative); 13.7% (261/1903) referred to the physician's communication (77.8% negative); 11.6% (220/1903) referred to the physician's caring attitude (75.9% negative); and 10.6% (203/1903) referred to the physician's friendliness (73.9% negative). In relation to staff issues, the most frequently mentioned issue was regarding the staffs' friendliness (109/1903; 5.7%); 43.1% of these ratings

were negative. Concerning practice issues, the frequently mentioned issues were as follows: 15.5% (295/1903) of the comments provided an overall assessment (38.6% positive), 8.1% (155/1903) referred to the waiting time within the practice (58.7% negative), and 5% (96/1903) referred to the atmosphere of the practice (40.6% positive).

However, there were some significant differences between German and French comments in a number of subcategories (Multimedia Appendix 1). For instance, German comments referred significantly more often to the physician's friendliness (χ^2_1=5.9, *P*=.01), being taken seriously by the physician (χ^2_1=8.5, *P*=.002), staff friendliness (χ^2_1=7.0, *P*=.005), waiting time in the practice (χ^2_1=6.3, *P*=.01), and practice atmosphere (χ^2_1=6.6, *P*=.007).

Table 4. Categorization of issues.

Issue	Total (N=1903), n (%)	Quantitative rating evaluation		
		Positive, n (%)	Neutral, n (%)	Negative, n (%)
Physician				
Satisfaction with treatment	533 (28.0)	82 (15.4)	61 (11.4)	390 (73.2)
Overall assessment	392 (20.6)	122 (31.0)	59 (15.1)	211 (53.8)
Recommendation	345 (18.1)	35 (10.1)	47 (13.6)	263 (76.2)
Communication	261 (13.7)	39 (14.9)	19 (7.3)	203 (77.8)
Caring attitude	220 (11.6)	40 (18.2)	13 (5.9)	167 (75.9)
Friendliness	203 (10.6)	35 (17.2)	18 (8.9)	150 (73.9)
Treatment cost/billing	173 (9.1)	8 (4.6)	30 (17.3)	135 (78.0)
Competence	170 (8.9)	43 (25.3)	13 (7.6)	114 (67.1)
Being taken seriously	141 (7.4)	10 (7.1)	10 (7.1)	121 (85.8)
Time spent with patient	136 (7.1)	17 (12.5)	18 (13.2)	101 (74.3)
Trust	133 (6.9)	44 (33.1)	11 (8.3)	78 (58.6)
Professionalism	97 (5.1)	11 (11.3)	6 (6.2)	80 (82.5)
Cooperation with medical specialists	13 (0.7)	4 (30.8)	0	9 (69.2)
Alternative medicine	8 (0.4)	6 (75.0)	0	2 (25.0)
Telephone availability	8 (0.4)	1 (12.5)	3 (37.5)	4 (50.0)
Privacy	7 (0.4)	0	2 (28.6)	5 (71.4)
Health insurance differentiation	6 (0.3)	0	0	6 (100)
Patient involvement	5 (0.3)	0	1 (20.0)	4 (80.0)
Individualized service	2 (0.1)	0	0	2 (100)
Child friendliness	1 (0.04)	0	0	1 (100)
Staff				
Friendliness	109 (5.7)	39 (35.8)	23 (21.1)	47 (43.1)
Overall assessment	60 (3.2)	19 (31.7)	18 (30.0)	23 (38.3)
Service/assistance	31 (1.6)	11 (35.5)	3 (9.7)	17 (54.8)
Communication	22 (1.2)	4 (18.2)	7 (31.8)	11 (50.0)
Professionalism	10 (0.5)	2 (20.0)	2 (20.0)	6 (60.0)
Availability by telephone	6 (0.3)	1 (16.7)	4 (66.7)	1 (16.7)
Time spent with patient	2 (0.1)	1 (50)	0	1 (50.0)
Health insurance differentiation	1 (0.04)	0	1 (100)	0
Trust	1 (0.04)	0	0	1 (100)
Practice				
Overall assessment	295 (15.5)	114 (38.6)	90 (30.5)	91 (30.8)
Waiting time within practice	155 (8.1)	22 (14.2)	42 (27.1)	91 (58.7)
Atmosphere	96 (5.0)	39 (40.6)	29 (30.2)	28 (29.2)
Organization	37 (1.9)	7 (18.9)	11 (29.7)	19 (51.4)
Ability to get appointment	36 (1.9)	4 (11.1)	11 (30.6)	21 (58.3)
Equipment	31 (1.6)	8 (25.8)	10 (32.3)	13 (41.9)
Recommendation	25 (1.3)	4 (16.0)	5 (20.0)	16 (64.0)
Consultation hours	8 (0.4)	3 (37.5)	3 (37.5)	2 (25.0)
Location	7 (0.4)	2 (28.6)	2 (28.6)	3 (42.9)

Issue	Total (N=1903), n (%)	Quantitative rating evaluation		
		Positive, n (%)	Neutral, n (%)	Negative, n (%)
Waiting room entertainment	5 (0.3)	4 (80.0)	1 (20.0)	0
Parking space	4 (0.2)	4 (100)	0	0
Availability by telephone	3 (0.2)	1 (33.3)	1 (33.3)	1 (33.3)
Privacy	3 (0.2)	0	3 (100)	0
Barrier-free access	2 (0.1)	0	1 (50.0)	1 (50.0)
Online appointment	1 (0.04)	0	1 (100)	0

Discussion

Principal Findings

As far as this author is aware, this is the first study internationally to examine feedback that has been rejected from a PRW. The key findings of this study are as follows: (1) the Swiss PRW Medicosearch rejected a total of 2352 patient feedback between September 16, 2008, and September 22, 2017, (2) just over half of all the rejected feedback were negative, and (3) the most frequently mentioned issue in the rejected feedback was satisfaction with treatment. Medicosearch has shown a lot of transparency in providing this rejected feedback for analysis. It is, however, unclear why the majority of the feedback were rejected. This is problematic and raises concerns that online patient feedback are being inappropriately manipulated.

Medicosearch did not provide the reasons why it rejected the feedback, and as far as this author is aware, Medicosearch does not use formal criteria for determining which feedback should be published or rejected. Of the 2352 ratings rejected by Medicosearch, 170 comments were excluded from the categorization framework for various reasons. For example, the feedback were not comprehensible, were explicitly labelled as test ratings, were about the PRW, etc. These would also appear to be legitimate reasons for Medicosearch to reject the feedback. However, the appropriateness of rejecting the remaining 92.7% of feedback is less clear, particularly as they appear to be qualitatively the same as the published feedback for a sample of Swiss physicians recently analyzed [19].

Twelve percent of the rejected feedback only provided a quantitative rating with no narrative comment. Medicosearch requires that both a quantitative rating and a narrative comment be provided in every patient feedback, and this is likely the reason for rejecting this feedback. Narrative comments often provide a richer source of information than quantitative ratings [10]; however, making narrative comments mandatory seems inappropriate. Some patients may not be willing or able to describe what happened in a narrative comment, but may still want to share their satisfaction with their physician with others. It is unclear why these ratings should simply be excluded because the patient did not want to also write a narrative comment.

In terms of the evaluation tendencies of online patient feedback, previous Swiss and international research has found that the published online patient feedback on PRWs is overwhelmingly positive [8,10,14,17-32]. However, recent research also raised concerns that negative feedback is being suppressed by Swiss PRWs [19]. For instance, the PRW OkDoc explicitly states on its website that any negative comments will be deleted, and while Medicosearch allows negative comments, it informs the concerned physician before publishing it online, so the physician can decide whether to activate the negative feedback function [19]. There was therefore an expectation that the majority of the rejected feedback would be negative. However, this analysis of 2352 rejected feedback from Medicosearch found that just over half of all rejected feedback were negative, and the average rejected rating was 2.8 out of 5.

The proportion of rejected negative feedback, however, is substantially higher than the proportion of negative feedback that has been published in the international literature [8,10,14,17-32]. Analysis of the published feedback for a sample of Swiss physicians also reported that only 4.3% (10/234) of the feedback published on Medicosearch was negative and that the average rating was 4.68 out of 5. It is unclear why there is such a large discrepancy between published and rejected negative feedback. It has previously been suggested that Switzerland's restrictive legal framework regarding data protection may be having a big impact on the types of online patient feedback that are published [18]. However, it may also be that PRWs like Medicosearch are also deciding themselves not to publish a lot of the negative feedback they receive owing to conflicts of interest. Medicosearch has shifted its business strategy toward online appointments, where physicians pay a fee and their booking systems are integrated with Medicosearch, which allows patients to book an appointment with a physician directly on Medicosearch. Consequently, Medicosearch will likely be reluctant to upset paying physicians by publishing too much negative feedback, as their business is now reliant on physicians using their online appointment system.

Users of PRWs can also manipulate online patient feedback, and there is some indication that physicians or practice staff sometimes pose as patients on PRWs to post either positive comments about themselves or negative comments about competitors [33]. Indeed, 25% (588/2352) of the rejected feedback were positive, and it is possible some of these were rejected because Medicosearch suspected that these were fake reviews. However, without a clear and consistent way to determine which feedback is rejected, there is a danger that feedback will be inappropriately rejected.

With regard to the contents of narrative comments, the most frequently mentioned issues identified from the rejected feedback included (1) satisfaction with treatment; (2) the overall

assessment of the physician; (3) recommending the physician; (4) the physician's communication; (5) the physician's caring attitude; and (6) the physician's friendliness. In comparison, the top five mentioned issues identified in the analysis of the published feedback for a sample of Swiss physicians were (1) the overall assessment of the physician and the physician's competence; (2) the physician's communication; (3) recommending the physician; (4) the physician's friendliness; and (5) the physician's caring attitude [19]. This suggests that online patient feedback is raising similar issues, regardless of whether it is published or rejected. Indeed, just like in the analysis of the published feedback for a sample of Swiss physicians [19], it is important to recognize that 95% (42/44) of the subcategories of the categorization framework and 81.5% (3101/3804) of the distinct issues identified were concerning aspects of performance (interpersonal skills of the physician and staff, infrastructure, and organization and management of the practice) that are considered to be assessable by patients.

If online patient feedback is going to be collected, there needs to be clear policies and practices about how this is handled. It cannot be left to the whims of PRWs, who may have financial incentives to suppress negative feedback, to decide which feedback is or is not published online. It has previously been recommended that "there is a need for consensus-based criteria that applies to all Swiss PRWs for determining which comments are and are not to be publicly published and which are clearly publicized so users of PRWs are aware of it" [19]. This analysis of 2352 rejected feedback from Medicosearch further highlights the need for such a consensus-based criteria. To support this, further research is needed to examine how many Swiss PRWs are using criteria for determining which feedback is published or not, what those criteria are, and what measures PRWs are using to address the manipulation of online patient feedback. It appears that research examining these issues would be helpful in most countries that have PRWs.

Limitations

This study has some limitations that should be taken into account when interpreting the results. First, it is unknown how many patient feedback Medicosearch received in total during the time period covered. The author contacted Medicosearch asking for this information but never received a response, and the information is not freely available on the website. It would be helpful to know the proportion of patient feedback that is being rejected. Second, the sample of rejected feedback was only taken from one Swiss PRW. Although Medicosearch is one of the oldest and most used Swiss PRWs, it is unclear how generalizable the results are to other PRWs and other countries. Future research examining whether PRWs are using criteria for determining which feedback is published or not should include all Swiss PRWs. Third, the specialty and sociodemographic information of the rated physicians are unknown, and there may be important differences between the different specialties and physicians. Finally, the sociodemographic information of the rating patients is unknown and may not be representative of Swiss patients in general.

Acknowledgments

This work was funded by the Swiss Academy of Medical Sciences' Käthe-Zingg-Schwichtenberg-Fonds, which had no role in the project design; in the collection, analysis, or interpretation of data; in the writing of the article; or in the decision to submit the paper for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Categorization of issues by language.
[DOCX File , 25 KB - jmir_v22i8e18374_app1.docx]

References

1. Institute of Medicine. Best care at lower cost: The path to continuously learning health care in America. Washington, DC: The National Academies Press; 2013.
2. Paterson R. The Good Doctor: What Patients Want. Auckland, New Zealand: Auckland University Press; 2010.
3. Faber M, Bosch M, Wollersheim H, Leatherman S, Grol R. Public reporting in health care: how do consumers use quality-of-care information? A systematic review. Med Care 2009 Jan;47(1):1-8. [doi: 10.1097/MLR.0b013e3181808bb5] [Medline: 19106724]
4. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. JAMA 2000 Apr 12;283(14):1866-1874. [doi: 10.1001/jama.283.14.1866] [Medline: 10770149]
5. Fung CH, Lim Y, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. Ann Intern Med 2008 Jan 15;148(2):111-123. [doi: 10.7326/0003-4819-148-2-200801150-00006] [Medline: 18195336]
6. Berger ZD, Joy SM, Hutfless S, Bridges JF. Can public reporting impact patient outcomes and disparities? A systematic review. Patient Educ Couns 2013 Dec;93(3):480-487. [doi: 10.1016/j.pec.2013.03.003] [Medline: 23579038]

7. Emmert M, Sander U, Esslinger AS, Maryschok M, Schöffski O. Public reporting in Germany: the content of physician rating websites. Methods Inf Med 2012;51(2):112-120. [doi: 10.3414/ME11-01-0045] [Medline: 22101427]
8. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013 Aug 06;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
9. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013 Feb 01;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
10. Emmert M, Meier F, Heider A, Dürr C, Sander U. What do patients say about their physicians? an analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
11. Hennig-Thurau T, Gwinner KP, Walsh G, Gremler DD. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? Journal of Interactive Marketing 2004 Jan;18(1):38-52. [doi: 10.1002/dir.10073]
12. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
13. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014 Mar 31;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
14. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
15. Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371 [FREE Full text] [doi: 10.3322/canjclin.53.6.356] [Medline: 15224975]
16. Rothenfluh F, Schulz PJ. Content, Quality, and Assessment Tools of Physician-Rating Websites in 12 Countries: Quantitative Analysis. J Med Internet Res 2018 Jun 14;20(6):e212 [FREE Full text] [doi: 10.2196/jmir.9105] [Medline: 29903704]
17. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jun;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]
18. McLennan S. Quantitative Ratings and Narrative Comments on Swiss Physician Rating Websites: Frequency Analysis. J Med Internet Res 2019 Jul 26;21(7):e13816 [FREE Full text] [doi: 10.2196/13816] [Medline: 31350838]
19. McLennan S. The Content and Nature of Narrative Comments on Swiss Physician Rating Websites: Analysis of 849 Comments. J Med Internet Res 2019 Sep 30;21(9):e14336 [FREE Full text] [doi: 10.2196/14336] [Medline: 31573918]
20. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013 Aug 28;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
21. Black E, Thompson L, Saliba H, Dawson K, Black NM. An analysis of healthcare providers' online ratings. Inform Prim Care 2009;17(4):249-253 [FREE Full text] [doi: 10.14236/jhi.v17i4.744] [Medline: 20359403]
22. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011 Nov 16;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
23. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012 Feb 24;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
24. Ellimoottil C, Hart A, Greco K, Quek ML, Farooq A. Online reviews of 500 urologists. J Urol 2013 Jun;189(6):2269-2273. [doi: 10.1016/j.juro.2012.12.013] [Medline: 23228385]
25. Sobin L, Goyal P. Trends of online ratings of otolaryngologists: what do your patients really think of you? JAMA Otolaryngol Head Neck Surg 2014 Jul;140(7):635-638. [doi: 10.1001/jamaoto.2014.818] [Medline: 24876073]
26. Murphy GP, Awad MA, Osterberg EC, Gaither TW, Chumnarnsongkhroh T, Washington SL, et al. Web-Based Physician Ratings for California Physicians on Probation. J Med Internet Res 2017 Aug 22;19(8):e254 [FREE Full text] [doi: 10.2196/jmir.7488] [Medline: 28830852]
27. Strech D, Reimann S. [German language physician rating sites]. Gesundheitswesen 2012 Aug;74(8-9):e61-e67. [doi: 10.1055/s-0031-1297254] [Medline: 22189671]
28. Emmert M, Gerstner B, Sander U, Wambach V. Eine Bestandsaufnahme von Bewertungen auf Arztbewertungsportalen am Beispiel des Nürnberger Gesundheitsnetzes Qualität und Effizienz (QuE). Gesundh ökon Qual manag 2013 Jul 9;19(04):161-167. [doi: 10.1055/s-0033-1335899]
29. McLennan S, Strech D, Reimann S. Developments in the Frequency of Ratings and Evaluation Tendencies: A Review of German Physician Rating Websites. J Med Internet Res 2017 Aug 25;19(8):e299 [FREE Full text] [doi: 10.2196/jmir.6599] [Medline: 28842391]

30. Liu JJ, Matelski JJ, Bell CM. Scope, Breadth, and Differences in Online Physician Ratings Related to Geography, Specialty, and Year: Observational Retrospective Study. J Med Internet Res 2018 Mar 07;20(3):e76 [FREE Full text] [doi: 10.2196/jmir.7475] [Medline: 29514775]
31. Greaves F, Pape UJ, Lee H, Smith DM, Darzi A, Majeed A, et al. Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. J Med Internet Res 2012 Oct 17;14(5):e146 [FREE Full text] [doi: 10.2196/jmir.2280] [Medline: 23076301]
32. Lagu T, Goff SL, Hannon NS, Shatz A, Lindenauer PK. A mixed-methods analysis of patient reviews of hospital care in England: implications for public reporting of health care quality data in the United States. Jt Comm J Qual Patient Saf 2013 Jan;39(1):7-15. [doi: 10.1016/s1553-7250(13)39003-5] [Medline: 23367647]
33. Bodkin H. 2018. GPs are posing as patients and posting 'fake reviews' online, health chiefs reveal, The Telegraph. URL: https://www.telegraph.co.uk/news/2018/07/16/gps-posting-fake-reviews-online-health-chiefs-reveal/ [accessed 2019-06-28]

Abbreviations

PRW: physician rating website

Edited by G Eysenbach; submitted 22.02.20; peer-reviewed by E Brunson, P Schulz, F Kaliyadan; comments to author 15.04.20; revised version received 22.05.20; accepted 11.06.20; published 03.08.20.

Please cite as:
McLennan S
Rejected Online Feedback From a Swiss Physician Rating Website Between 2008 and 2017: Analysis of 2352 Ratings
J Med Internet Res 2020;22(8):e18374
URL: https://www.jmir.org/2020/8/e18374
doi:10.2196/18374
PMID:32687479

Original Paper

Effect of Adding Telephone-Based Brief Coaching to an mHealth App (Stay Strong) for Promoting Physical Activity Among Veterans: Randomized Controlled Trial

Laura J Damschroder[1], MS, MPH; Lorraine R Buis[2], PhD; Felicia A McCant[3], MSSW; Hyungjin Myra Kim[1], PhD; Richard Evans[1], MSc; Eugene Z Oddone[3,4], MHS, MD; Lori A Bastian[5,6], MPH, MD; Gwendolyn Hooks[1], MA; Reema Kadri[2], MLIS; Courtney White-Clark[3], MS; Caroline R Richardson[2], MD; Jennifer M Gierisch[3,4,7], MPH, PhD

[1]Veterans Affairs Center for Clinical Management Research, Ann Arbor Healthcare System, Ann Arbor, MI, United States

[2]University of Michigan, Department of Family Medicine, Ann Arbor, MI, United States

[3]Veterans Affairs Durham Center of Innovation to Accelerate Discovery and Practice Transformation, Durham Veterans Affairs Health Care System, Durham, NC, United States

[4]Division of General Internal Medicine, Department of Medicine, Duke University Medical Center, Durham, NC, United States

[5]Veterans Affairs Pain Research, Informatics, Multimorbidities, and Education Center, Veterans Affairs Connecticut, West Haven, CT, United States

[6]Division of General Internal Medicine, Department of Medicine, Yale University, West Haven, CT, United States

[7]Department of Population Health Sciences, Duke University Medical Center, Durham, NC, United States

Corresponding Author:
Laura J Damschroder, MS, MPH
Veterans Affairs Center for Clinical Management Research
Ann Arbor Healthcare System
2215 Fuller Rd (152)
Ann Arbor, MI, 48105
United States
Phone: 1 7348453603
Email: laura.damschroder@va.gov

Abstract

Background: Though maintaining physical conditioning and a healthy weight are requirements of active military duty, many US veterans lose conditioning and rapidly gain weight after discharge from active duty service. Mobile health (mHealth) interventions using wearable devices are appealing to users and can be effective especially with personalized coaching support. We developed *Stay Strong*, a mobile app tailored to US veterans, to promote physical activity using a wrist-worn physical activity tracker, a Bluetooth-enabled scale, and an app-based dashboard. We tested whether adding personalized coaching components (*Stay Strong+Coaching*) would improve physical activity compared to *Stay Strong* alone.

Objective: The goal of this study is to compare 12-month outcomes from *Stay Strong* alone versus *Stay Strong+Coaching.*

Methods: Participants (n=357) were recruited from a national random sample of US veterans of recent wars and randomly assigned to the *Stay Strong* app alone (n=179) or *Stay Strong+Coaching* (n=178); both programs lasted 12 months. Personalized coaching components for *Stay Strong+Coaching* comprised of automated in-app motivational messages (3 per week), telephone-based human health coaching (up to 3 calls), and personalized weekly goal setting. All aspects of the enrollment process and program delivery were accomplished virtually for both groups, except for the telephone-based coaching. The primary outcome was change in physical activity at 12 months postbaseline, measured by average weekly Active Minutes, captured by the Fitbit Charge 2 device. Secondary outcomes included changes in step counts, weight, and patient activation.

Results: The average age of participants was 39.8 (SD 8.7) years, and 25.2% (90/357) were female. Active Minutes decreased from baseline to 12 months for both groups (P<.001) with no between-group differences at 6 months (P=.82) or 12 months (P=.98). However, at 12 months, many participants in both groups did not record Active Minutes, leading to missing data in 67.0% (120/179) for *Stay Strong* and 61.8% (110/178) for *Stay Strong+Coaching*. Average baseline weight for participants in *Stay Strong* and *Stay Strong+Coaching* was 214 lbs and 198 lbs, respectively, with no difference at baseline (P=.54) or at 6 months (P=.28) or 12 months (P=.18) postbaseline based on administrative weights, which had lower rates of missing data. Changes in the number of steps recorded and patient activation also did not differ by arm.

Conclusions: Adding personalized health coaching comprised of in-app automated messages, up to 3 coaching calls, plus automated weekly personalized goals, did not improve levels of physical activity compared to using a smartphone app alone. Physical activity in both groups decreased over time. Sustaining long-term adherence and engagement in this mHealth intervention proved difficult; approximately two-thirds of the trial's 357 participants failed to sync their Fitbit device at 12 months and, thus, were lost to follow-up.

Trial Registration: ClinicalTrials.gov NCT02360293; https://clinicaltrials.gov/ct2/show/NCT02360293

International Registered Report Identifier (IRRID): RR2-10.2196/12526

(J Med Internet Res 2020;22(8):e19216) doi:10.2196/19216

KEYWORDS

exercise; veterans; smartphones; wearable physical activity tracker; behavior change; mobile phone; online; app; mobile app; wearable

Introduction

Background

Adequate levels of physical activity (PA) reduce the risk of many diseases including diabetes, obesity, cardiovascular disease, dementia, and many cancers. Furthermore, adequate levels contribute to significant improvements in quality of life by improving sleep and physical function, preventing falls, and improving pain management. Inadequate PA is one of the top drivers of premature death. Despite strong evidence for PA's beneficial effects, most adult men (74%) and women (81%) in the United States do not meet national recommendations for PA levels [1] (150 minutes of moderate activity or 90 minutes of vigorous activity per week [2]).

A lack of PA is especially prevalent among veterans of the US Armed Forces. The type and intensity of physical activities that veterans engage in too often abruptly reduces as they transition from active duty to postdeployment civilian life. The relatively unstructured nature of postdeployment life and illnesses or injuries sustained during military service may contribute to this shift in activity levels [3]. Younger veterans involved in the Afghanistan and Iraq wars may also have work-life balance issues related to child and older adult care issues, and challenges integrating into civilian life because of high physical and mental health burdens (eg, chronic pain, mental illness, substance abuse) [4,5]. Additionally, in one large cohort, 65.8% of men and 46.7% of women were overweight or obese at their first postdeployment visit in the Veteran Health Administration, an additional barrier to engaging in PA [6].

One potential strategy for increasing PA among veterans is using readily available consumer-grade wearable PA sensors and monitoring devices. A solid foundation of evidence demonstrates shorter-term effectiveness of internet-mediated interventions [7-17], particularly when combined with wearable PA sensors, tailored motivational messaging, and coaching [18-20]. By leveraging the broad availability of mobile sensors, we can increase access to interventions aimed at increasing PA levels [21-23]. The evidence-base for mobile health (mHealth) interventions, however, is largely based on small trials with short-term follow-up [13,21]; longer-term engagement with mHealth programs is not always sustained [24-26]. Human-based or automated health coaching, which has produced positive lifestyle changes across a wide range of populations [26-29], has potential for sustaining longer-term adherence, engagement, and outcomes.

Several studies using mobile apps to promote weight loss and increase PA have reported high rates of attrition, with participants dropping out after the first month [11]. Without active adherence and engagement by participants, the mHealth intervention is unlikely to be effective [30]. A potential strategy to increase adherence and engagement with, and effectiveness of mHealth interventions, is to add health coaching, including telephone-based lifestyle coaching delivered by humans [31-35]. Health coaching is a patient-centered, collaborative model grounded in theories of health behavior change where coaches work in partnership with patients using motivational interviewing, goal setting, and problem solving as key strategies. Across a wide variety of populations, health coaching has produced positive impacts on lifestyle modifications [27-29]. In a recent trial, a relatively low-intensity dose (2 coaching calls at 1 and 4 weeks postbaseline) of telephone-based coaching in conjunction with use of an online risk assessment tool resulted in increased engagement in lifestyle change programs and increased patient activation in a trial among US veteran participants, compared to use of the online risk assessment alone [27-29]. Thus, adding coaching features such as telephone-based human coaching, extended with personalized automated messaging to help address barriers and provide motivation and personalized weekly goals, may further enhance the impact of mHealth interventions.

Study Objective

The objective of this study is to determine whether PA levels would improve at 12 months with a wearable activity tracker combined with health coaching versus a wearable activity tracker alone, among US veterans of recent Afghanistan or Iraq wars. PA levels were measured by Active Minutes, as recorded by the Fitbit Charge 2, a consumer-grade mobile PA sensor that was widely available at the time of this study.

Methods

Study Design

This is a comparative effectiveness randomized controlled trial comparing two 12-month programs: *Stay Strong*, an mHealth intervention using a wearable activity tracker with a dashboard available through a smartphone app along with a

Bluetooth-enabled weight scale, versus *Stay Strong+Coaching*, comprising *Stay Strong* plus human health coaching provided over the telephone and in-app automated weekly personalized PA goals, motivational messages, and personalized weekly PA goals. A summary of the methods is provided here; our published protocol provides more details [36].

Recruitment

A stratified random sample of administrative medical record data for US veterans of recent wars including Operation Enduring Freedom (OEF), Operation Iraqi Freedom (OIF), and Operation New Dawn (OND) residing within the United States was used to identify potentially eligible individuals. We oversampled women to ensure they comprised at least 20% of participants. Inclusion criteria were online confirmation of OEF, OIF, or OND veteran status; identifying a Veterans Health Administration (VHA) health care provider as being responsible for their medical care; having interest in starting a PA program within the next 30 days; having access to a computer with internet connection and a working USB port; having a smartphone with a compatible iOS or Android operating system; and being younger than 65 years (because the interventions were targeted to OEF, OIF, and OND veterans who would typically be younger than 65 years). Individuals were excluded if they reported that a health care provider had told them that it was currently unsafe to exercise in an unsupervised or unmonitored setting, had a history of eating disorders or a BMI<20, were not competent to consent for themselves to a research study, or wore a PA sensor within the last 30 days.

Invitation letters briefly describing the study included a URL with an individualized code to access an online portal. This was the first study within the Veterans Affairs (VA) that relied completely on online technology-mediated approaches for recruitment, consent, Health Insurance Portability Accountability Act (HIPAA) authorization, enrollment, delivery of the interventions, and conduct of the program (see [37] for content). Telephonic support was available to participants as needed for technical support.

After consent, participants completed a baseline survey. At the end of the survey, online instructions asked participants to download and install the *Stay Strong* app on their smartphone via the Google Play (Android smartphone users) or Apple (iPhone users) stores. When *Stay Strong* was successfully installed, a package was shipped containing instructions, their Fitbit Charge 2 device, a Bluetooth-enabled weight scale, and a USB dongle for syncing their Fitbit device using a USB-enabled computer. Individuals were instructed to sync their Fitbit device via the Fitbit Connect software using a Bluetooth-enabled dongle that was plugged into a USB port on a computer. This configuration was necessary to comply with VHA data security and confidentiality standards. Syncing was typically completed within minutes of initiation. Individuals were also instructed to configure their Bluetooth-enabled weight scale by pairing their smartphone with the scale. In-app instructions were provided to walk the user through step-by-step so that data received via Bluetooth connection would be recorded by the *Stay Strong* app and displayed by the in-app dashboard. Synching typically lasted seconds.

Participants authorized Vibrent, Inc (the developer of the *Stay Strong* server platform) [38] to sync and access their Fitbit data. Their Fitbit device had a “Do Not Remove” sticker, covering the device’s display during the baseline period before randomization.

When at least 5 valid days of data (days when at least 5 Lightly Active Minutes were recorded) within a 7-day period were synced to the study server, the individual was randomized in a 1:1 ratio, to *Stay Strong* or *Stay Strong+Coaching*. All the study staff were blinded to the randomization list that was generated by the study biostatistician. After the participant was assigned to an arm, their smartphone-based mobile app was updated to reflect their assigned program (*Stay Strong* or *Stay Strong+Coaching*) and they were instructed to remove the sticker that covered their Fitbit display.

Follow-up surveys were administered at 6- and 12-months postbaseline. Respondents were mailed a US $25 Amazon gift card for each completed follow-up survey. All participants kept their Fitbit device and scale after their program ended.

Institutional Review Board Approval and Ethical Considerations

Ethical oversight was provided by VHA’s Central Institutional Review Board that approved the protocol. A copy of the approved study protocol is available online [37]. Participants were randomized between October 11, 2017, and May 31, 2018. The last participants finished their program on July 9, 2019.

Interventions

Table 1 lists program components for the two trial arms. The *Stay Strong* programs both lasted 12 months. Designs were informed by experiences of veterans in a prior study [39] and by the self-regulation theory and the information-motivation-behavioral skills model [40-43], which describe processes of behavior change mediated through goal attainment and skills mastery, and acknowledges the central role of self-efficacy in sustaining change in PA [41,44-47]. Our published protocol provides more details about behavior change techniques incorporated into the programs [36]. Detailed functional requirements, screenshots, along with the full library of messages sent to participants are available online [37]. The Fitbit Charge 2 provided detailed minute-by-minute self-monitoring information through objective measurement of PA. A veterans’ work group provided input into the logistics of intervention delivery and enrollment processes. Feedback on an early version of *Stay Strong+Coaching* was elicited from a convenience group of testers employed at the VA and veterans who served on an advisory panel for this grant, several of whom were OEF, OIF, or OND veterans. No changes were made to *Stay Strong* during the course of this trial.

Table 1. *Stay Strong* components by trial arm.

Component	*SS*[a] arm	*SS+Coaching* arm	Intensity	Duration	Mode
Objective physical activity monitoring (Fitbit Charge 2 and data visualizations within *SS*)	✓	✓	Fitbit worn daily, and data syncing at least 1/week	1 year	Fitbit worn on wrist; data visualizations available within *SS* app
Weight self-monitoring (scale and weight data visualizations within *SS*)	✓	✓	Weight measured weekly with data syncing at least 1/week	1 year	Data visualizations available within *SS* app
Administrative message reminders (reminders for Fitbit and weight scale syncing, adverse event reporting, and data assessments)	✓	✓	One message less than 230 characters	As needed over 1 year	Push notification on phone
Automated personalized goal setting		✓	Weekly, based on previous weeks' physical activity data	1 year	Abbreviated phone push notification and message with image within *SS* app
Automated messages: nonpersonalized		✓	1 message up to 225 characters, 3/week	1 year	Abbreviated phone push notification and full message with visual image within SS app
Automated messages: personalized based on self-reported barriers		✓	1 message up to 225 characters, 3/week	1 year	In-app and smartphone notification with image
Telephone-based lifestyle coaching		✓	Up to 30 min	2 calls plus an optional 3rd call in the first 9 weeks	Telephone

[a]*SS*: *Stay Strong*.

Stay Strong Intervention

The *StayStrong* program lasted 12 months and comprised of a Fitbit device to capture PA, a Bluetooth-enabled weight scale, and a smartphone app with a dashboard showing key metrics over time (Active Minutes, miles, steps, stairs, and heart rate zone). The Fitbit Charge 2 device is a wrist-worn PA monitoring device that continuously logs PA. Participants were encouraged to wear the Fitbit device during waking hours for the study duration and to upload device data at least weekly via the Fitbit Connect software and USB port on their computer. The smartphone app displayed PA data in 1- to 4-week increments (or most recent valid week).

Data from the Bluetooth-enabled weight scale (A&D Deluxe Connected Weight Scale UC-352BLE) was synced with *Stay Strong* or could be manually entered. The smartphone app displayed weight data in 1- to 4-week increments. Participants were asked to weigh themselves and sync their scale at least weekly during the duration of their program.

During the study period, all participants received automated administrative messages including reminders to report adverse events every 90 days and reminders to complete 6- and 12-month survey assessments.

Stay Strong+Coaching Intervention

In addition to the *Stay Strong* components, participants received personalized coaching, which comprised of automated in-app motivational messages (3/week for the duration of the 12-month program), telephone-based human health coaching (up to 3 calls, spaced over the first 9 weeks), and personalized weekly goal setting. The coaching telephone calls were designed to motivate participants by helping them develop goals and action plans to achieve Fitbit-derived PA goals, problem solve barriers to achieving PA goals, and understand features of the *Stay Strong* app, with an emphasis on interpreting Fitbit PA data shown in their dashboard to monitor their progress. PA goals were computed by increasing each new daily PA goal by 5 Active Minutes based on previous week's (or most recent) synced Active Minutes, not to exceed 60 Active Minutes per day. Participants received three messages per week delivered within the app and via push notification to the smartphone; most were nonpersonalized, but a subset were personalized to address specific barriers identified by each participant. To maintain interest and engagement in the messages throughout the 12 months, we randomly timed in-app message delivery during the day Monday through Saturday. Messages comprised of a maximum 225 characters and were designed to help participants stay engaged and learn more about topics including: exercise, healthy eating, initiating behavior change, pain, inspirational quotes, maintaining behavior change, weight loss and management, heart rate monitoring, and appropriate athletic gear. Additionally, at baseline and 6 months, participants were asked to choose up to four barriers that most prevented increasing PA from a list of 11 prespecified barriers (lack of time, social influence, lack of energy, lack of willpower or motivation, fear of injury or pain, lack of resources, family obligations, weather conditions, depression, accountability or external motivation, and disability). The prespecified list was developed based on work by Sallis and colleagues [48,49] and

highlighted by the Centers for Disease Control and Prevention, plus the addition of disability.

Outcomes

The primary outcome was Active Minutes per week, as recorded by the Fitbit Charge 2 device, for 12 months following randomization. To report PA levels (eg, Active Minutes, steps), participants synced their Fitbit device as often as desired. We encouraged participants to sync at least once per week. Active Minutes is a proprietary measure that captures the number of minutes of continuous moderate-to-vigorous exercise when sustained for at least 10 minutes [50]. Secondary outcomes included step counts (reported through syncing the Fitbit device), weight loss, and patient activation. Weight was to be recorded by the Bluetooth-enabled scale; participants were encouraged to record their weight at least weekly. However, most participants did not sync their scales and, thus, did not provide weights. Therefore, for the comparison of weight change rates, we conducted an alternative analysis with weights captured by the VA administrative medical record data, and baseline weights were self-reported at the time of enrollment. Patient activation was assessed by online questionnaire at baseline and at 6 and 12 months postbaseline using the self-reported 13-item Patient Activation Measure (PAM) [51]. The PAM assesses an individual's knowledge, skills, beliefs, and confidence for managing their own health. PAM scores have high construct validity and have been positively associated with engagement in healthier lifestyle behaviors [52].

Sample Size

Our sample size calculations were based on unpublished data from a pilot study. A 10-minute differential improvement at 12 months was set as a minimal clinically important difference. We anticipated a baseline mean of 53 minutes, a standard deviation of 28 minutes in both treatment groups, and r=0.46 correlation between baseline and 12 months. Because of the lengthy enrollment process, we expected up to 50% dropout during the consent and preparation phases, and assumed 25% dropout after enrollment during the 12-month program. We aimed to randomize 350 patients (175 per group) to detect a 10-minute difference in improvement at 12 months based on a 5% significance level 2-sided test using analysis of covariance (ANCOVA) with 90% power.

Statistical Analysis

Primary analyses were based on intent-to-treat focusing on the effect of *Stay Strong+Coaching* compared to *Stay Strong* alone on change in PA from baseline to 12 months postrandomization. Women who self-reported pregnancy at any of the three primary assessment times (baseline, 6-month, 12-month) were not included in the analyses for PA and weight; 1 participant indicated pregnancy at the 6-month follow-up. Summary statistics (eg, means, medians, and proportions) were used to describe all study variables including outcome measures for overall study participants, by study arm, and at each of the three primary assessment times. Adjusted between-arm difference in Active Minutes at both 6 and 12 months were compared and estimated based on a mixed model using data at the three primary assessment times. Between-arm comparison was also done using a mixed model with all longitudinally assessed weekly averages of Active Minutes as dependent variables. The model included time (weeks since randomization), treatment arm (*Stay Strong+Coaching* arm) indicator, an interaction of time by treatment arm as primary predictors, and random intercepts and slopes. The model was also adjusted for baseline Active Minute goal and stratification factors of sex and smartphone operating system. A test of significant slope of the interaction term was used to assess if the rate of change in Active Minutes over the study period differed between treatment arms. Secondary outcomes of steps and weights were analyzed similarly using data at primary assessment times, as well as full weekly data, and patient activation was analyzed using data at three primary assessment times.

Several alternative analyses were conducted to ensure consistency in our main results. Due to high skewness of the PA data, we modeled weekly Active Minutes after log-transformation and step counts after taking the square root. For PA data, we also used robust regression based on the median, minimizing the sum of the absolute deviation from the estimate of the center.

Additionally, to account for a substantial amount of missing follow-up data, primary analyses were repeated using weighted likelihood methods to give more weights to individuals who were more likely to miss 12-month outcomes. Weights were estimated from a penalized likelihood (least absolute shrinkage and selection operator) logistic regression model, with missing 12-month data as the response variable and with baseline sociodemographic characteristics as predictors of missing 12-month data. "Do Not Remove" stickers covered each participant's Fitbit screen to prevent feedback that may motivate higher levels of PA even with instructions to maintain normal levels of activity. Alternative analyses were conducted to test for "reactivity," where participants may have increased their activity levels despite these precautions. This "reactivity" often manifests as unusually high or low activity levels with use of a new device like the Fitbit Charge 2 used in this study. If this occurred, PA may decline to previously normal levels by the second week. To assess reactivity, we re-estimated the between-arm difference in Active Minutes at 12 months after replacing baseline data with the second week data. The a priori level for statistical significance was a 2-sided P<.05. For all analyses, R version 3.6.0 (R Foundation for Statistical Computing) was used. All code and detailed results are available online at [37].

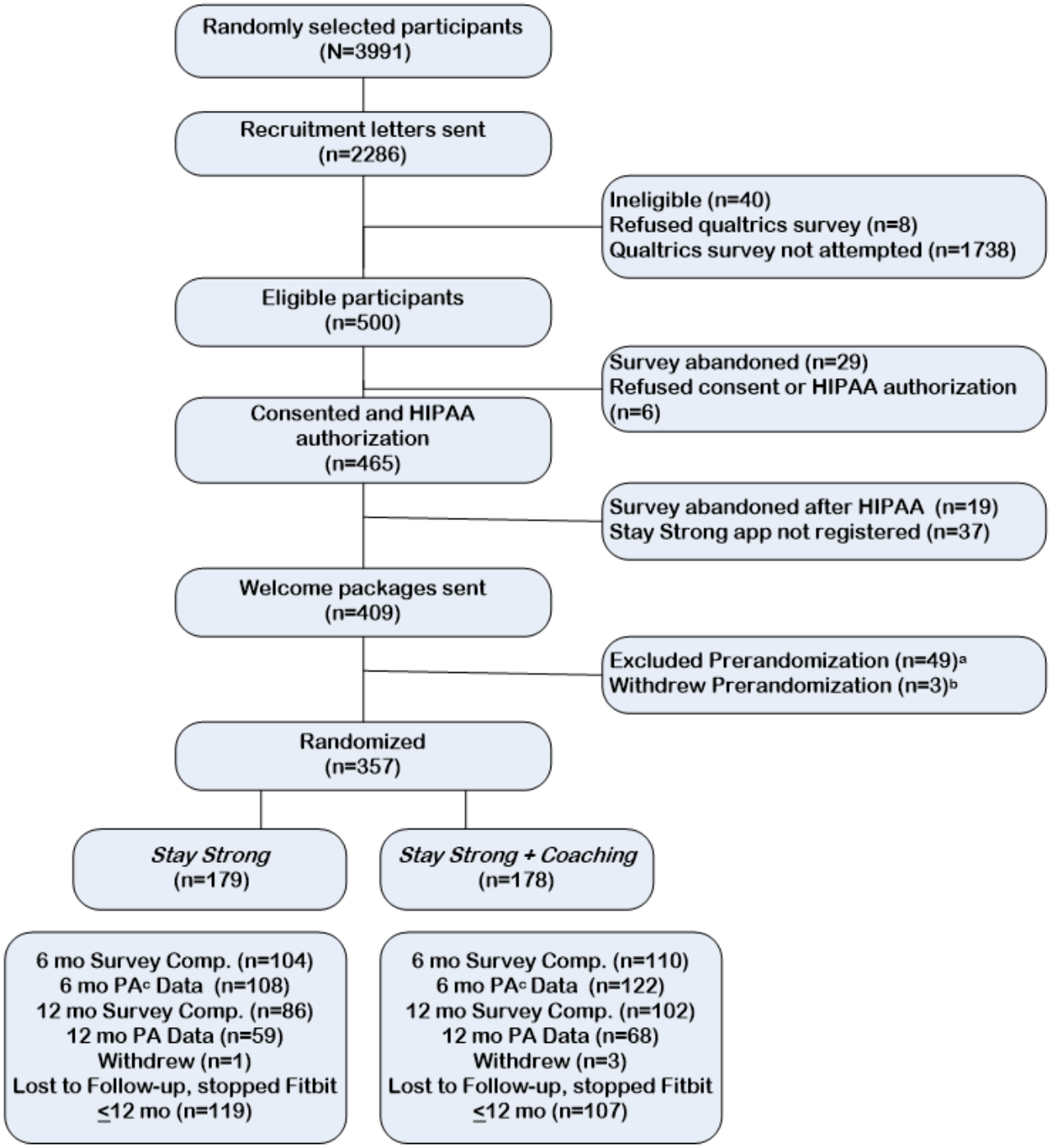
Results

Participant Characteristics

Letters with the VHA letterhead along with a logo specially designed for *Stay Strong* were mailed to 2286 randomly selected US veterans, of whom 17.9% (409) were eligible, completed consent, provided HIPAA authorization, and to whom welcome packages were sent containing their Bluetooth scale and Fitbit device along with instructions for use (see Figure 1). Of those 409 participants, 357 (87.3%) successfully set up their devices and synced a week of valid PA data and were randomized to *Stay Strong* (n=179) or *Stay Strong+Coaching* (n=178).

Table 2 shows baseline participant characteristics overall and by study arm. Of the 357 participants, the average age was 39.8 years, 90 (25.2%) were female, 231 (64.7%) were non-Hispanic White, 248 (69.5%) were married, 215 (60.2%) had children in the household, 156 (40.9%) had a bachelor's or higher college degree, and 48 (13.4%) were current smokers. Additionally, based on administrative medical record data, 191 (53.5%) were diagnosed with posttraumatic stress disorder (PTSD), 118 (35%) had diagnoses of clinically significant depression, 155 (43%) reported moderate or severe pain, 95 (26.6%) reported clinically significant alcohol misuse based on Alcohol Use Disorders Identification Test Version C score, and 204 (57.1%) were obese (BMI≥30).

Figure 1. CONSORT flow of recruitment through randomization.



Note: a: Unable to sync Fitbit (n=12), did not pair with study account (n=12), unable to setup device due to secure environment (n=1), unable to contact (n=9), unable to setup Fitbit with Stay Strong app (n=10), unable to comply or follow study procedures (n=5); b: Changed mind about participating (n=3); c: PA: physical activity

Table 2. Baseline characteristics of participants (N=357).

Characteristics	Overall (N=357)	*SS*[a] (n=179)	*SS+Coaching* (n=178)
Sociodemographic factors			
Sex (female), n (%)	90 (25.2)	43 (24.0)	47 (26.4)
Age (years), mean (SD)	39.8 (8.7)	40.4 (9.0)	39.2 (8.4)
Race/ethnicity, n (%)			
Non-Hispanic White	231 (64.7)	114 (63.7)	117 (65.7)
Non-Hispanic Black	47 (13.2)	27 (15.1)	20 (11.2)
Other	79 (22.1)	38 (21.2)	41 (23.0)
Married, n (%)	248 (69.5)	128 (71.5)	120 (67.4)
Have children in household, n (%)	215 (60.4)	117 (65.7)	98 (55.1)
Education, n (%)			
High school graduate/equivalent (GED[b]) or less	23 (6.4)	7 (3.9)	16 (9.0)
Some college, trade/vocational, associate's degree	188 (52.7)	98 (54.7)	90 (50.6)
Bachelor's degree	94 (26.3)	47 (26.4)	47 (26.3)
Postgraduate work or graduate degree	52 (14.6)	27 (15.1)	25 (14.0)
Full-time employment, n (%)	202 (56.6)	97 (54.2)	105 (59.0)
Inadequate income[c], n (%)	49 (13.7)	27 (15.1)	22 (12.4)
Health and comorbidities			
General health[d], n (%)			
Excellent	16 (4.5)	16 (4.5)	8 (4.5)
Very good	74 (20.7)	74 (20.7)	36 (20.2)
Good	150 (42.0)	150 (42.0)	71 (39.9)
Fair	97 (27.2)	97 (27.2)	50 (28.1)
Poor	20 (5.6)	20 (5.6)	13 (7.3)
Moderate/severe pain[e], n (%)	155 (43.4)	75 (41.9)	80 (44.9)
Diabetes[f], n (%)	27 (7.6)	16 (8.9)	11 (6.2)
Hypertension[f], n (%)	63 (17.6)	31 (17.3)	32 (18.0)
Posttraumatic stress disorder diagnosis[f], n (%)	191 (53.5)	97 (54.2)	94 (52.8)
Depression scale (PHQ-8[g]), median (IQR)	6.5 (3.0-12.0)	6.0 (3.0-11.0)	7.0 (3.0-13.0)
Clinically significant depression[h], n (%)	118 (34.9)	57 (33.9)	61 (35.9)
Lifestyle and related factors			
Patient Activation Measure, mean (SD)[i]	70.2 (11.9)	70.3 (12.2)	70.1 (11.7)
AUDIT-C[j]≥4 (male); AUDIT-C≥3 (female), n (%)	95 (26.6)	42 (23.5)	53 (29.8)
Current smoker[k], n (%)	48 (13.4)	22 (12.3)	26 (14.6)
Weight (lbs), mean (SD)	210.4 (44.9)	208.5 (39.5)	212.3 (49.8)
BMI (kg/m^2), mean (SD)	31.1 (5.7)	30.9 (5.3)	31.2 (6.2)
<25	38 (10.6)	16 (8.9)	22 (12.4)
25-29	115 (32.2)	57 (31.8)	58 (32.6)
30-34	110 (30.8)	62 (34.6)	48 (27.0)
35-39	70 (19.6)	34 (19.0)	36 (20.2)

Characteristics	Overall (N=357)	*SS*[a] (n=179)	*SS+Coaching* (n=178)
≥40	24 (6.7)	10 (5.6)	14 (7.9)
Technology use			
Comfort using the internet[l], mean (SD)	32.6 (3.8)	32.4 (4.0)	32.8 (3.6)
Phone type: iPhone[m], n (%)	174 (48.7)	89 (49.7)	85 (47.8)
Physical activity device use[n], n (%)	164 (45.9)	74 (41.3)	90 (50.6)
Prior experience with Fitbit use, n (%)	95 (26.6)	44 (24.6)	51 (28.7)

[a]*SS*: *Stay Strong*.

[b]GED: General Education Development.

[c]Responded yes to "Must cut back on things to pay bills or have difficulty paying bills at the end of the month."

[d]12-Item Short-Form Health Survey [53].

[e]Pain intensity ≥4 out of 11-point scale.

[f]See [37] for specific diagnosis codes used for defining this category.

[g]PHQ-8: Patient Health Questionnaire Depression Scale.

[h]PHQ-8≥10 [54].

[i]Patient Activation Measure-13 score [51].

[j]AUDIT-C: Alcohol Use Disorders Identification Test Version C.

[k]Combination of "Smoke every day" and "Smoke some days."

[l]Score range from 7 to 35; a higher score corresponds to more comfort.

[m]In comparison to Android.

[n]Prior experience with a physical activity device (eg, Fitbit, Apple Watch).

At baseline, 103/179 (57.9%) and 96/178 (53.9%) of participants in *Stay Strong* and *Stay Strong+Coaching* recorded more than 150 minutes of weekly Active Minutes (*P*=.46). All participants recorded Active Minutes during their baseline week, but by 12 months postrandomization, most participants were not synching their Fitbit, thus PA data were not available; specifically, though 230/357 (64.4%) participants provided synced data at 6 months, only 127 (35.6%) did so at 12 months.

Primary Outcomes

Adjusted mean Active Minutes, based on a repeated measures ANCOVA model, showed no between-arm differences at 6 months (*P*=.82) or 12 months (*P*=.98). Mean weekly Active Minutes reported by Fitbit devices declined in both arms. A mixed model based on weekly longitudinal Active Minutes data revealed Active Minutes decreased significantly over the 12 months in the *Stay Strong* group (weekly slope=–3.04, *P*<.001), with no significant difference in the rate of decrease in the two study arms (*P*=.40 for the interaction of time by the *Stay Strong+Coaching* arm indicator). Multiple alternate analytic models resulted in similar findings with neither clinically, nor statistically, significant differences in Active Minutes between study arms. For example, a mixed model weighted by the estimated probability of missing 12-month data showed significantly decreasing Active Minutes over time (*P*<.001), but no difference in the rate of decrease in Active Minutes between the two arms (*P*=.37). We also tested and adjusted for "reactivity," given the high baseline levels of Active Minutes recorded to answer the question. We first assessed whether participants increased their normal PA levels in their baseline week despite masking feedback on their Fitbit device by covering their Fitbit display with a "Do Not Remove" sticker that prevented users from seeing and reacting to PA levels recorded by the Fitbit. If reactivity was present, then we would expect PA levels to decrease in the following weeks [55,56]. Our analyses revealed that Active Minutes decreased slightly from the first to the second week; however, analysis where second week data replaced the baseline data did not alter findings.

Secondary Outcomes

We found no significant differences between arms in any secondary outcomes including step counts (*P*=.08), weight (*P*=.55), or patient activation (*P*=.98) at the 12-month follow-up (Table 3). The between-arm difference in the predicted mean at 12 months was 1009 steps per day, adjusting for sex, type of smartphone, and baseline goal. For step counts per day, averaged over a week, crude means declined from 8163 steps per day at baseline to 5736 at 12 months in the *Stay Strong* arm, and from 7571 to 5638 in the *Stay Strong+Coaching* arm. Multiple alternate models based on weekly step counts all showed significantly decreasing step counts over the 12 months in the *Stay Strong* group (*P*<.001), with no significant difference in the rate of decrease in the two study arms.

Table 3. Primary outcome of Active Minutes and secondary outcomes of steps, weight, and patient activation for each major assessment time.

Outcomes	*Stay Strong* (n=179), mean[a] (95% CI)	*Stay Strong+Coaching* (n=178), mean[a] (95% CI)	Between-group difference, mean[b] (SE)	*P* value[c]
Active Minutes per week				
Baseline (N=357)	255 (216-295)	240 (201-280)	8.63 (23.0)	.71
6-month (n=230)	225 (175-276)	234 (186-281)	–10.0 (24.2)	.68
12-month (n=127)	190 (121-258)	199 (136-263)	–29.4 (41.2)	.48
Steps per week				
Baseline (N=357)	8163 (7531-8795)	7571 (6938-8205)	152 (465)	.74
6-month (n=230)	6351 (5537-7165)	6563 (5797-7328)	–841 (551)	.13
12-month (n=127)	5736 (4635-6837)	5638 (4612-6663)	–1009 (583)	.08
Weight (lbs)				
Baseline (N=357)	214 (200-228)	198 (182-215)	–3.5 (4.8)	.46
6-month (n=97)	217 (205-228)	206 (193-219)	–3.3 (4.9)	.49
12-month (n=65)	221 (206-235)	217 (203-232)	–3.1 (5.3)	.55
Patient activation				
Baseline (n=315)	70.4 (68.5-72.2)	70.1 (68.3-72.0)	0.20 (1.35)	.88
6-month (n=198)	68.0 (65.7-70.4)	66.9 (64.5-69.3)	0.91 (1.65)	.58
12-month (n=171)	69.4 (66.8-72.1)	69.2 (66.7-71.6)	–0.04 (1.77)	.98

[a]Crude means. n in the first column represent the number of participants with available data for crude means.

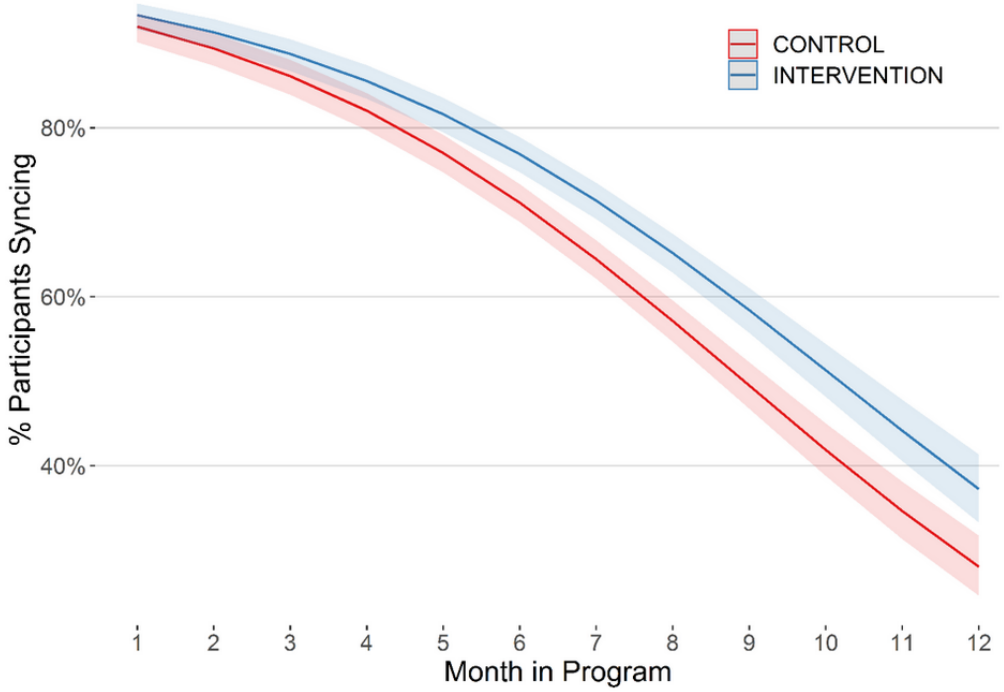
[b]Calculated as the estimated marginal mean difference (*Stay Strong* group – *Stay Strong+Coaching* group) based on a model fit using all available data (n=179 for *Stay Strong* and n=178 for *Stay Strong+Coaching*) and adjusting for baseline goal, sex, and operating system type for all outcomes except for patient activation, which relies on n's listed in the first column for between-group difference and *P* values.

[c]For between-group difference, adjusted for comparing a family of 3 estimates.

Adherence

Of the 178 *Stay Strong+Coaching* arm participants, 70.8% (n=126) completed at least 2 coaching calls, and 56.7% (n=101) completed all 3 phone calls. However, participants in both groups increasingly failed to sync their Fitbit devices over their 12-month program (Figure 2). At 9 weeks, soon after coaching ended, there was no difference in syncing rates between the two groups (*P*=.14). By 6 months postbaseline, 60.3% (108/179) and 68.5% (122/178) of participants in *Stay Strong* and *Stay Strong+Coaching*, respectively, synced their Fitbit data. This difference was reflected by participants in *Stay Strong+Coaching* having higher odds of syncing their data compared to participants in *Stay Strong* (OR 1.36, 95% CI 1.17-1.58; *P*<.001). This difference was not sustained at 12 months postbaseline: rates were comparably low with 33.0% (59/179) and 38.2% (68/178) in *Stay Strong* and *Stay Strong+Coaching*, respectively, syncing their Fitbit data.

Figure 2. Percentage of participants who synced their Fitbit within the last 30 days by month and arm.



Satisfaction With Stay Strong

The 12-month satisfaction survey was completed by 51.8% (185/357) of participants across the two programs, 89.7% (166/185) reported being "likely" or "extremely likely" to recommend their program to another veteran, and 69.7% (129/185) agreed or strongly agreed that their program was a benefit to their overall health. Overall, when asked if they did *not* like wearing their Fitbit, only 13.5% (25/185) strongly agreed or agreed. Though only 24.9% (46/185) of respondents strongly agreed or agreed that they found it difficult to sync their Fitbit using a desktop computer, 82.2% (152/185) strongly agreed or agreed that they would rather sync their Fitbit data using a smartphone than a desktop computer.

Discussion

Summary of Findings

This is the first completed large-scale trial of an mHealth intervention using wearable PA tracking devices (Fitbit) and a smartphone app among veterans. Adding automated and phone-based human coaching to the *Stay Strong* mHealth program (*Stay Strong+Coaching*) did not improve PA levels compared to baseline, nor compared to the *Stay Strong* program alone among US veterans of recent wars. Specifically, the rate of changes did not show difference between arms in Active Minutes, step counts, patient activation, and weight at 6 months or 12 months. In *Stay Strong+Coaching*, 70.8% (126/178) of participants completed at least 2 of 3 planned coaching calls in the first 9 weeks of the program but, like participants in *Stay Strong*, significantly decreased the frequency of syncing their Fitbit device to the point where over 60% (230/357) of trial participants had missing PA data 12 months postbaseline.

Program Adherence and Missing Data

The completion of coaching calls was high: 70.8% (126/178) completed at least 2 calls, and over half (101/178, 56.7%) completed all 3 phone calls, even with the third call being optional. Other than the coaching calls for *Stay Strong+Coaching*, participants completed their program without human interaction. We attempted to reach participants who had not synced their Fitbit data within 7 days of the 12-month program ending; up to 9 phone calls were made over a 3-week period, with one follow-up letter. Sustaining long-term adherence to and engagement with mHealth interventions without human contact is challenging [26,57,58]. The rate of data syncing was relatively high at 3 months (317/357, 88.8%; Figure 2) in our trial, which is the time period evaluated in many published mHealth trials. However, the percentage of participants who synced their data by 12 months was low (127/357, 35.6%) for both arms. One potential explanation for this is that participants were asked to use a dongle plugged into a computer's USB port with Fitbit Connect software to bypass direct syncing using Fitbit's proprietary app installed on their smartphone. This process was not as efficient as using Fitbit's app directly for syncing their device; in fact, the Connect software is no longer supported. Ethics oversight required use of the Connect software, however, to minimize the possibility of personal information (eg, name, locations, information from contacts lists stored on their smartphone) being accessed and stored by Fitbit. At 6 months, participants in *Stay Strong+Coaching* were more likely to sync their Fitbit data, suggesting that the added telephone-based health coaching with automated weekly personalized goal messages and personalized and standard motivational messaging may have helped retain participants for a longer period.

Participant Characteristics and mHealth Interventions

Stay Strong was targeted specifically to OEF, OIF, and OND veterans who tend to be below-average age compared to the general veteran population. The average age of our participants was under 40 years old; this is much younger than the general veteran population, more than half of whom are over 60 [59]. Recruitment goals were met more quickly than initially planned, indicating a high level of initial enthusiasm for the study and potential ease of virtual enrollment procedures. However, despite high baseline PA levels, our participants reported lower quality of life (117/357, 32.8% reported fair or poor health) compared to the general US adult population (typically below 20%) [60,61]. Our study participants also had a high burden of mental health and other comorbidities; one-third to over half reported clinical depression, moderate or severe pain, or had a PTSD diagnosis. These conditions all present potential challenges to maintaining or increasing PA levels [4]. Other research has identified potential risks of developing mHealth interventions that are too complex, that may be inattentive to user needs and capabilities, and may leave vulnerable patients behind [62], perpetuating health disparities. Less healthy and poorer individuals may be least likely to use interventions using wearable devices [63].

mHealth Interventions for Physical Activity

PA levels decreased by 41-65 active minutes at 12 months compared to baseline. However, PA declines among *Stay Strong+Coaching* participants largely occurred after the first 6 months, while *Stay Strong* participants continuously declined throughout their 12-month program. Thus, the additional lifestyle coaching support may have helped sustain PA levels longer compared to *Stay Strong* alone when any possible lasting effects subsided. Our findings are consistent with others who found that PA decreases over time in studies employing accelerometers as an intervention strategy [26]. We found no evidence of reactivity where participants may have been motivated to increase their PA levels during their baseline week even with a “Do Not Remove” sticker that covered their device so they would not see their data; on average, PA did not significantly decrease in the second week postbaseline.

This study marks a significant contribution to the mHealth literature. Our negative findings should be viewed in context of having an active comparator: both arms of the trial provided devices and a smartphone app to support PA. Further, this was a randomized trial of a relatively long-term program (12 months), drawing from a large national sample of OEF, OIF, and OND veterans who consented and enrolled online with no in-person assessments or interactions. *Stay Strong* was designed based on a fully described theoretical framework [36], which few apps do [64]. Furthermore, we followed participants for 12 months, which is longer than most other published trials [9,12,26,57,58]. Further development and testing are needed to continue to find interventions to help people increase and, importantly, sustain PA levels. Higher intensity and dose of human coaching may help. All human coaching was completed within the first 9 weeks; timing calls based on synced data (eg, when PA decreases or a participant fails to sync in a period of time) may help bolster levels when an individual is waning in their efforts or encountering new challenges, or increasing the number of calls over a longer period of time. Further, it is important to note that the content of the coaching calls may need to shift over time as participants lapse in and out of maintenance or receive new threats to their lifestyle modifications. Thus, behavioral strategies used to initiate behavior change, like increasing PA, are likely different from those needed to sustain gains over time. Our theoretical model was based on supports needed to initiate behavior change. We did not implement human coaching supports to maintain changes. A recent systematic review of behavior change techniques supports this hypothesis [65]. Although goal setting and self-monitoring of behaviors were important in both short- and long-term behavioral change, long-term behavior change also benefited from additional behavioral supports such as giving feedback on the outcome of the behavior, adding objects into the participant’s environment, receiving social support, and problem solving. These long-term techniques are likely hard to communicate or practice without providing human coaching over a longer time.

Human Coaching to Strengthen mHealth Interventions

To our knowledge this is one of the first studies to assess the addition of coaching components on PA as an add-on component to an mHealth intervention with objective self-monitoring and feedback. The goals of our coaching strategy were to aid initial engagement and help keep interest in the mHealth intervention fresh and interesting for participants so they would continue to participate and, thus, enhance impact of the mHealth program [66]. Therefore, we frontloaded human coaching to occur within the first 9 weeks of the program. Much of the literature compares multimodal such as coaching + mHealth + objective self-monitoring to usual care or weak, inactive comparators such as an educational comparator [26,67]. Such designs make it impossible to tease out the independent contribution of coaching to mHealth engagement. Moreover, in much of the literature, mHealth plays a supportive role in the intervention with coaching as the central component. In *Stay Strong*, the mHealth platform is the central intervention component and human coaching is subordinate (ie, only 3 sessions in first 9 weeks of a 12-month mHealth program). Other studies have demonstrated that approaches that integrate coaching have more robust outcomes, and this was a central hypothesis in our study. The current literature is not adequate to address the independent contribution of coaching on mHealth interventions aimed at increasing PA [26,67].

Role of Motivational Messages

Our barrier-specific messaging was based on a twice-administered survey (baseline and again 6 months later) of barriers such as lack of time, asking participants to choose up to four possible PA barriers they would encounter. This allows targeting messages to specific barriers. However, more recent advances with microtailoring based on season, geographic location, momentary mood, personal characteristics, employment and parenting demands, or other life circumstances would provide more actionable, meaningful, and potentially more motivating messages. Further tailoring to PA levels may also be effective, such as messaging when there is a gap in

synced data [68], and the addition of more strategies to motivate and engage [69]. A challenge with mHealth interventions is that the novelty of the intervention may be motivating for a short period, but after the novelty wears off, the interventions lose their effectiveness. This is true for messaging as well; though we varied the time and day of our messages to help make them "fresh," they were not timed based on any specific attributes or preferences of the participant. Another component to consider is the addition of an online community to increase engagement [70,71], though one systematic review only found trials that lasted 14 weeks or less [72], well before our participants, 64% (230/357) of whom were still syncing at 6 months, stopped syncing. Additionally, we did pilot approaches with veterans and used their feedback to guide development of the *Stay Strong* interventions. User-centered design approaches that more deeply involve potential participants in design through evolving rounds of development [73-75] to inform outcomes [76,77], information displays, and message content and timing may result in higher intervention durability and better outcomes.

Study Limitations

This trial has several limitations. This trial was designed to assess outcomes between two mHealth interventions (*Stay Strong* vs *Stay Strong+Coaching*); this design precluded our ability to assess and, thus, compare change in PA among veterans without any mHealth intervention. However, nearly half of participants reported prior experience using a PA device, indicating its widespread use, which makes it challenging to require participants *not* to use a device while participating in a trial. Our primary outcome was Active Minutes, a proprietary measure captured by the Fitbit device that captures moderate or vigorous exercise levels in bouts of at least 10 minutes. This metric may have been confusing and, thus, demotivating for some participants who may not have fully understood why they were not getting "credit" for exercise if they failed to get their heart rate up for a long enough period. On the other hand, Fitbit also displayed step counts, which is a well-known and commonly used metric. Our findings are based on a minority of participants who synced data at 12 months postbaseline. Syncing frequency was our only indication of adherence to the *Stay Strong* app. Unfortunately, we did not have the ability to build in other measures of adherence or engagement at the participant level including, for example, time spent in the app. Multiple alternative models did not reveal any clear bias between participants who were lost to follow-up versus those who were included in our outcome analyses. Baseline PA levels were quite high among our study participants; over half met the minimum standard of 150 moderate or vigorous minutes of PA per week at baseline, which was surprising, given earlier indications of low PA levels among veterans [3]. Exploratory analyses did not support the possibility that participants may have increased their PA at baseline compared to a true "usual" level, even with a "Do Not Remove" sticker covering their Fitbit display.

Conclusions

Although research has shown mHealth to have potential for promoting health behavior change, long-term participant adherence to study protocols and sustained engagement with mHealth interventions remains a challenge [24-26]. Our trial results have important implications for future research in this arena. Over 12 months, participant adherence to study protocols across both *Stay Strong* programs declined over time, as did PA levels. Although more *Stay Strong+Coaching* participants synced their Fitbit at 6 months compared to *Stay Strong* alone, we found no significant differences in PA between groups at 9 weeks, shortly after coaching ended for the *Stay Strong+Coaching* participants, nor at the end of the program (12 months). If we had less loss to follow-up at 12 months, we may have seen intervention effects. Continuing to develop ways to optimize content and type of automated and intensifying human health coaching informed by evidence-based behavior change techniques are strategies to explore to realize the full potential of mHealth.

Acknowledgments

This study was supported by the Department of VA, VHA, Office of Research and Development, and Health Services Research and Development Grant # CRE 12-305, as well as the VA Ann Arbor Center for Clinical Management Research and the Center of Innovation to Accelerate Discovery and Practice Transformation (CIN 13-410) at the Durham VA Health Care System. Study design, conduct, and all statistical analyses were completed by the study team independent of the funder and vendors involved in the study. We are grateful to the team at Vibrent, Inc who developed the *Stay Strong* apps for both arms and the back-end server functions that enabled the study team and coaches to capture and interact with participant data. A special thank you to Lilly Pritula with the Department of Family Medicine at the University of Michigan who graciously agreed to prepare the manuscript with her formatting and editing prowess.

Conflicts of Interest

None declared.

References

1. Giroir BP, Wright D. Physical activity guidelines for health and prosperity in the United States. JAMA 2018 Nov 20;320(19):1971-1972. [doi: 10.1001/jama.2018.16998] [Medline: 30418473]
2. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical activity guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]

3. Buis LR, Kotagal LV, Porcari CE, Rauch SAM, Krein SL, Richardson CR. Physical activity in postdeployment Operation Iraqi Freedom/Operation Enduring Freedom veterans using Department of Veterans Affairs services. JRRD 2011;48(8):901. [doi: 10.1682/JRRD.2010.08.0144]
4. Haskell SG, Mattocks K, Goulet JL, Krebs EE, Skanderson M, Leslie D, et al. The burden of illness in the first year home: do male and female VA users differ in health conditions and healthcare utilization. Womens Health Issues 2011;21(1):92-97 [FREE Full text] [doi: 10.1016/j.whi.2010.08.001] [Medline: 21185994]
5. Possemato K, Wade M, Andersen J, Ouimette P. The impact of PTSD, depression, and substance use disorders on disease burden and health care utilization among OEF/OIF veterans. Psychological Trauma Theory Res Pract and Policy 2010;2(3):218-223. [doi: 10.1037/a0019236]
6. Rosenberger PH, Ning Y, Brandt C, Allore H, Haskell S. BMI trajectory groups in veterans of the Iraq and Afghanistan wars. Prev Med 2011 Sep;53(3):149-154 [FREE Full text] [doi: 10.1016/j.ypmed.2011.07.001] [Medline: 21771610]
7. Bhardwaj N, Wodajo B, Gochipathala K, Paul ID, Coustasse A. Can mHealth revolutionize the way we manage adult obesity? Perspect Health Inf Manag 2017;14(Spring):1a [FREE Full text] [Medline: 28566984]
8. Buchholz SW, Wilbur J, Ingram D, Fogg L. Physical activity text messaging interventions in adults: a systematic review. Worldviews Evid Based Nurs 2013 Aug;10(3):163-173. [doi: 10.1111/wvn.12002] [Medline: 23746267]
9. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344. [doi: 10.1038/oby.2010.208] [Medline: 20847736]
10. Changizi M, Kaveh MH. Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population-a systematic review. Mhealth 2017;3:51. [doi: 10.21037/mhealth.2017.08.06] [Medline: 29430455]
11. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
12. Lee J, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
13. Pfaeffli Dale L, Dobson R, Whittaker R, Maddison R. The effectiveness of mobile-health behaviour change interventions for cardiovascular disease self-management: A systematic review. Eur J Prev Cardiol 2016 May;23(8):801-817. [doi: 10.1177/2047487315613462] [Medline: 26490093]
14. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight. J Cardiovasc Nurs 2013;28(4):320-329. [doi: 10.1097/jcn.0b013e318250a3e7]
15. Wang J, Wang Y, Wei C, Yao N, Yuan A, Shan Y, et al. Smartphone interventions for long-term health management of chronic diseases: an integrative review. Telemed J E Health 2014 Jun;20(6):570-583. [doi: 10.1089/tmj.2013.0243] [Medline: 24787747]
16. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
17. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? an evidence review. J Med Internet Res 2016 Oct 31;18(11):e287. [doi: 10.2196/jmir.5692] [Medline: 27806926]
18. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
19. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
20. Richardson CR, Newton TL, Abraham JJ, Sen A, Jimbo M, Swartz AM. A meta-analysis of pedometer-based walking interventions and weight loss. Ann Fam Med 2008;6(1):69-77 [FREE Full text] [doi: 10.1370/afm.761] [Medline: 18195317]
21. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, et al. Current science on consumer use of mobile health for cardiovascular disease prevention. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/cir.0000000000000232]
22. Lewis BA, Napolitano MA, Buman MP, Williams DM, Nigg CR. Future directions in physical activity intervention research: expanding our focus to sedentary behaviors, technology, and dissemination. J Behav Med 2017 Feb;40(1):112-126 [FREE Full text] [doi: 10.1007/s10865-016-9797-8] [Medline: 27722907]
23. Triantafyllidis A, Filos D, Claes J, Buys R, Cornelissen V, Kouidi E, et al. Computerised decision support in physical activity interventions: a systematic literature review. Int J Med Inform 2018 Mar;111:7-16. [doi: 10.1016/j.ijmedinf.2017.12.012] [Medline: 29425636]
24. Birkhoff SD, Smeltzer SC. Perceptions of smartphone user-centered mobile health tracking apps across various chronic illness populations: an integrative review. J Nurs Scholarsh 2017 Jul;49(4):371-378. [doi: 10.1111/jnu.12298] [Medline: 28605151]

25. Buis LR, Hirzel L, Turske SA, Des Jardins TR, Yarandi H, Bondurant P. Use of a text message program to raise type 2 diabetes risk awareness and promote health behavior change (part I): assessment of participant reach and adoption. J Med Internet Res 2013 Dec 19;15(12):e281 [FREE Full text] [doi: 10.2196/jmir.2928] [Medline: 24356329]
26. Goode AP, Hall KS, Batch BC, Huffman KM, Hastings SN, Allen KD, et al. The impact of interventions that integrate accelerometers on physical activity and weight loss: a systematic review. Ann Behav Med 2017 Feb;51(1):79-93 [FREE Full text] [doi: 10.1007/s12160-016-9829-1] [Medline: 27565168]
27. Kivelä K, Elo S, Kyngäs H, Kääriäinen M. The effects of health coaching on adult patients with chronic diseases: a systematic review. Patient Educ Couns 2014 Nov;97(2):147-157. [doi: 10.1016/j.pec.2014.07.026] [Medline: 25127667]
28. Oliveira JS, Sherrington C, Amorim AB, Dario AB, Tiedemann A. What is the effect of health coaching on physical activity participation in people aged 60 years and over? A systematic review of randomised controlled trials. Br J Sports Med 2017 Oct;51(19):1425-1432. [doi: 10.1136/bjsports-2016-096943] [Medline: 28320732]
29. Olsen JM, Nesbitt BJ. Health coaching to improve healthy lifestyle behaviors: an integrative review. Am J Health Promot 2010 Sep;25(1):e1-e12. [doi: 10.4278/ajhp.090313-lit-101]
30. Azar KMJ, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
31. Oddone EZ, Gierisch JM, Sanders LL, Fagerlin A, Sparks J, McCant F, et al. A coaching by telephone intervention on engaging patients to address modifiable cardiovascular risk factors: a randomized controlled trial. J Gen Intern Med 2018 Sep;33(9):1487-1494 [FREE Full text] [doi: 10.1007/s11606-018-4398-6] [Medline: 29736750]
32. Donnelly JE, Goetz J, Gibson C, Sullivan DK, Lee R, Smith BK, et al. Equivalent weight loss for weight management programs delivered by phone and clinic. Obesity (Silver Spring) 2013 Oct;21(10):1951-1959. [doi: 10.1002/oby.20334] [Medline: 23408579]
33. Gold DB, Anderson DR, Serxner SA. Impact of a telephone-based intervention on the reduction of health risks. Am J Health Promot 2000;15(2):97-106. [doi: 10.4278/0890-1171-15.2.97] [Medline: 11194702]
34. Perri MG, Limacher MC, Durning PE, Janicke DM, Lutes LD, Bobroff LB, et al. Extended-care programs for weight management in rural communities: the treatment of obesity in underserved rural settings (TOURS) randomized trial. Arch Intern Med 2008 Nov 24;168(21):2347-2354 [FREE Full text] [doi: 10.1001/archinte.168.21.2347] [Medline: 19029500]
35. U.S. Preventive Services Task Force. Behavioral counseling in primary care to promote a healthy diet: recommendations and rationale. Am J Prev Med 2003 Jan;24(1):93-100. [doi: 10.1016/s0749-3797(02)00581-0] [Medline: 12554028]
36. Buis LR, McCant FA, Gierisch JM, Bastian LA, Oddone EZ, Richardson CR, et al. Understanding the effect of adding automated and human coaching to a mobile health physical activity app for Afghanistan and Iraq veterans: protocol for a randomized controlled trial of the Stay Strong intervention. JMIR Res Protoc 2019 Jan 29;8(1):e12526. [doi: 10.2196/12526]
37. CCMRcodes / StayStrong 100. GitHub. URL: https://github.com/ccmrcodes/StayStrong
38. Vibrent Health. URL: https://www.vibrenthealth.com/ [accessed 2020-07-22]
39. Holtz B, Krein SL, Bentley DR, Hughes ME, Giardino ND, Richardson CR. Comparison of veteran experiences of low-cost, home-based diet and exercise interventions. J Rehabil Res Dev 2014;51(1):149-160. [doi: 10.1682/jrrd.2013.04.0088]
40. Amico KR, Harman JJ, O'Grady MA. Attrition and related trends in scientific rigor: a score card for ART adherence intervention research and recommendations for future directions. Curr HIV/AIDS Rep 2008 Nov;5(4):172-185. [doi: 10.1007/s11904-008-0026-0] [Medline: 18838057]
41. Ashford S, Edmunds J, French DP. What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis. Br J Health Psychol 2010 May;15(Pt 2):265-288. [doi: 10.1348/135910709X461752] [Medline: 19586583]
42. Fisher JD, Amico KR, Fisher WA, Harman JJ. The information-motivation-behavioral skills model of antiretroviral adherence and its applications. Curr HIV/AIDS Rep 2008 Nov;5(4):193-203. [doi: 10.1007/s11904-008-0028-y] [Medline: 18838059]
43. Steinhardt MA, Dishman RK. Reliability and validity of expected outcomes and barriers for habitual physical activity. J Occup Med 1989 Jun;31(6):536-546. [doi: 10.1097/00043764-198906000-00011] [Medline: 2786559]
44. DuVall C, Dinger MK, Taylor EL, Bemben D. Minimal-contact physical activity interventions in women: a pilot study. Am J Health Behav 2004;28(3):280-286. [doi: 10.5993/ajhb.28.3.9] [Medline: 15152887]
45. Kelly S, Melnyk BM, Belyea M. Predicting physical activity and fruit and vegetable intake in adolescents: a test of the information, motivation, behavioral skills model. Res Nurs Health 2012 Apr;35(2):146-163. [doi: 10.1002/nur.21462] [Medline: 22262049]
46. Pearson MR, Prince MA, Bravo AJ. Moderators of the effects of alcohol protective behavioral strategies: three attempts of replication and extension. Subst Use Misuse 2017 Jun 07;52(7):939-949. [doi: 10.1080/10826084.2016.1267222] [Medline: 28426365]
47. Rhodes RE, Dickau L. Moderators of the intention-behaviour relationship in the physical activity domain: a systematic review. Br J Sports Med 2013 Mar;47(4):215-225. [doi: 10.1136/bjsports-2011-090411] [Medline: 22278998]
48. Sallis JF, Hovell MF. Determinants of exercise behavior. Exerc Sport Sci Rev 1990;18:307-330. [Medline: 2192898]
49. Sallis JF, Hovell MF, Hofstetter CR. Predictors of adoption and maintenance of vigorous physical activity in men and women. Prev Med 1992 Mar;21(2):237-251. [doi: 10.1016/0091-7435(92)90022-a] [Medline: 1579558]

50. What are Active Zone Minutes or active minutes on my Fitbit device? Fitbit, Inc. URL: https://help.fitbit.com/articles/en_US/Help_article/1379 [accessed 2020-07-22]
51. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005 Dec;40(6 Pt 1):1918-1930 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00438.x] [Medline: 16336556]
52. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do increases in patient activation result in improved self-management behaviors? Health Serv Res 2007 Aug;42(4):1443-1463 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00669.x] [Medline: 17610432]
53. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
54. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
55. Clemes SA, Deans NK. Presence and duration of reactivity to pedometers in adults. Med Sci Sports Exerc 2012 Jun;44(6):1097-1101. [doi: 10.1249/MSS.0b013e318242a377] [Medline: 22595985]
56. Matevey C, Rogers LQ, Dawson E, Tudor-Locke C. Lack of reactivity during pedometer self-monitoring in adults. Meas Phys Education Exerc Sci 2006 Mar;10(1):1-11. [doi: 10.1207/s15327841mpee1001_1]
57. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR mHealth uHealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
58. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
59. National Center for Veterans Analysis and Statistics. US Department of Veterans Affairs. URL: https://www.va.gov/vetdata/Veteran_Population.asp [accessed 2020-07-28]
60. Idler EL, Russell LB, Davis D. Survival, functional limitations, and self-rated health in the NHANES I Epidemiologic Follow-up Study, 1992. First National Health and Nutrition Examination Survey. Am J Epidemiol 2000 Nov 01;152(9):874-883. [doi: 10.1093/aje/152.9.874] [Medline: 11085400]
61. McGee DL, Liao Y, Cao G, Cooper RS. Self-reported health status and mortality in a multiethnic US cohort. Am J Epidemiol 1999 Jan 01;149(1):41-46. [doi: 10.1093/oxfordjournals.aje.a009725] [Medline: 9883792]
62. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 2016 Dec 01;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
63. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 03;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
64. Cowan LT, Van Wagenen SA, Brown BA, Hedin RJ, Seino-Stephan Y, Hall PC, et al. Apps of steel: are exercise apps providing consumers with realistic expectations?: a content analysis of exercise apps for presence of behavior change theory. Health Educ Behav 2013 Apr;40(2):133-139. [doi: 10.1177/1090198112452126] [Medline: 22991048]
65. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
66. Carter DD, Robinson K, Forbes J, Hayes S. Experiences of mobile health in promoting physical activity: a qualitative systematic review and meta-ethnography. PLoS One 2018;13(12):e0208759 [FREE Full text] [doi: 10.1371/journal.pone.0208759] [Medline: 30557396]
67. Connelly J, Kirk A, Masthoff J, MacRury S. The use of technology to promote physical activity in Type 2 diabetes management: a systematic review. Diabet Med 2013 Dec;30(12):1420-1432. [doi: 10.1111/dme.12289] [Medline: 23870009]
68. Höchsmann C, Infanger D, Klenk C, Königstein K, Walz SP, Schmidt-Trucksäss A. Effectiveness of a behavior change technique-based smartphone game to improve intrinsic motivation and physical activity adherence in patients with type 2 diabetes: randomized controlled trial. JMIR Serious Games 2019 Feb 13;7(1):e11444 [FREE Full text] [doi: 10.2196/11444] [Medline: 30758293]
69. Tang MY, Smith DM, Mc Sharry J, Hann M, French DP. Behavior change techniques associated with changes in postintervention and maintained changes in self-efficacy for physical activity: a systematic review with meta-analysis. Ann Behav Med 2019 Aug 16;53(9):801-815. [doi: 10.1093/abm/kay090] [Medline: 30534971]
70. Resnick PJ, Janney AW, Buis LR, Richardson CR. Adding an online community to an internet-mediated walking program. Part 2: strategies for encouraging community participation. J Med Internet Res 2010 Dec 17;12(4):e72 [FREE Full text] [doi: 10.2196/jmir.1339] [Medline: 21169161]
71. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010 Dec 17;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]

72. Petersen JM, Prichard I, Kemps E. A comparison of physical activity mobile apps with and without existing web-based social networking platforms: systematic review. J Med Internet Res 2019 Aug 16;21(8):e12687 [FREE Full text] [doi: 10.2196/12687] [Medline: 31420956]
73. Carbonnel F, Ninot G. Identifying frameworks for validation and monitoring of consensual behavioral intervention technologies: narrative review. J Med Internet Res 2019 Oct 16;21(10):e13606 [FREE Full text] [doi: 10.2196/13606] [Medline: 31621638]
74. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
75. Patel S, Arya M. The BUS Framework: a comprehensive tool in creating an mHealth app utilizing behavior change theories, user-centered design, and social marketing. J Mob Technol Med 2017 Apr;6(1):39-45 [FREE Full text] [doi: 10.7309/jmtm.6.1.6] [Medline: 29104674]
76. Nittas V, Lun P, Ehrler F, Puhan MA, Mütsch M. Electronic patient-generated health data to facilitate disease prevention and health promotion: scoping review. J Med Internet Res 2019 Oct 14;21(10):e13320 [FREE Full text] [doi: 10.2196/13320] [Medline: 31613225]
77. Vaida B. Patient-centered outcomes research: early evidence from a burgeoning field. Health Aff (Millwood) 2016 Apr;35(4):595-602. [doi: 10.1377/hlthaff.2016.0239] [Medline: 27044957]

Abbreviations

ANCOVA: analysis of covariance
HIPAA: Health Insurance Portability Accountability Act
mHealth: mobile health
OEF: Operation Enduring Freedom
OIF: Operation Iraqi Freedom
OND: Operation New Dawn
PA: physical activity
PAM: Patient Activation Measure
PTSD: posttraumatic stress disorder
VA: Veterans Affairs
VHA: Veterans Health Administration

Edited by G Eysenbach; submitted 08.04.20; peer-reviewed by J Petersen, N Bhardwaj, V Nittas; comments to author 01.05.20; revised version received 11.06.20; accepted 07.07.20; published 04.08.20.

Please cite as:
Damschroder LJ, Buis LR, McCant FA, Kim HM, Evans R, Oddone EZ, Bastian LA, Hooks G, Kadri R, White-Clark C, Richardson CR, Gierisch JM
Effect of Adding Telephone-Based Brief Coaching to an mHealth App (Stay Strong) for Promoting Physical Activity Among Veterans: Randomized Controlled Trial
J Med Internet Res 2020;22(8):e19216
URL: http://www.jmir.org/2020/8/e19216/
doi:10.2196/19216
PMID:32687474

Original Paper

Cancer Survivors' Receptiveness to Digital Technology–Supported Physical Rehabilitation and the Implications for Design: Qualitative Study

Sine Rossen[1,2], MSc, PhD; Lars Kayser[3], MD, PhD; Jette Vibe-Petersen[2], MD; Jesper Frank Christensen[1], MSc, PhD; Mathias Ried-Larsen[1], MSc, PhD

[1]Centre for Physical Activity Research, Rigshospitalet, Copenhagen, Denmark

[2]Copenhagen Centre for Cancer and Health, Municipality of Copenhagen, Copenhagen, Denmark

[3]Department of Public Health, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Sine Rossen, MSc, PhD
Copenhagen Centre for Cancer and Health
Municipality of Copenhagen
Nørre Allé 45
Copenhagen, DK-2200
Denmark
Phone: 45 60298218
Email: kg58@kk.dk

Abstract

Background: Physical activity is associated with a positive prognosis in cancer survivors and may decrease the risk of adverse effects of treatment. Accordingly, physical activity programs are recommended as a part of cancer rehabilitation services. Digital technology may support cancer survivors in increasing their level of physical activity and increase the reach or efficiency of cancer rehabilitation services, yet it also comes with a range of challenges.

Objective: The aim of this qualitative study was to explore cancer survivors' receptiveness to using digital technology as a mode of support to increase their physical activity in a municipality-based cancer rehabilitation setting.

Methods: Semistructured interviews were conducted with 11 cancer survivors (3 males, 8 females, age range 32-82 years) who were referred for cancer rehabilitation and had participated in a questionnaire survey using the Readiness and Enablement Index for Health Technology (READHY) questionnaire. Data analysis was based on the content analysis method.

Results: Two themes were identified as important for the interviewees' receptiveness to using digital technology services in connection with their physical activity during rehabilitation: their attitude toward physical activity and their attitude toward digital technology–assisted physical activity. Our results indicated that it is important to address the cancer survivors' motivation for using technology for physical activity and their individual preferences in terms of the following: (1) incidental or structured (eg, cardiovascular and strength exercises or disease-specific rehabilitative exercises) physical activity; (2) social or individual context; and (3) instruction (know-how) or information (know-why).

Conclusions: The identified preferences provide new insight that complements the cancer survivors' readiness level and can likely help designers, service providers, and caregivers provide solutions that increase patient receptiveness toward technology-assisted physical activity. Combining digital technology informed by cancer survivors' needs, preferences, and readiness with the capacity building of the workforce can aid in tailoring digital solutions to suit not only individuals who are receptive to using such technologies but also those reluctant to do so.

(J Med Internet Res 2020;22(8):e15335) doi:10.2196/15335

KEYWORDS

cancer; rehabilitation; physical activity; digital technology

Introduction

Increased physical activity among cancer survivors is associated with increased survival and a lower risk of cancer recurrence, particularly among breast and colorectal cancer survivors [1-5]. Physical activity is also associated with decreased cancer-related fatigue and sleep disturbances [6], increased health-related quality of life [7], and decreased treatment-related adverse effects [5,8]. The general recommendation for adults is at least 150 minutes per week of moderate-intensity physical activity [9]. Accordingly, the incorporation of physical activity is recommended in cancer rehabilitation services [10,11]. Participating in rehabilitation is a challenge for some cancer survivors. In the United States, up to 42% of cancer survivors do not meet the recommendations for physical activity [12]. In Copenhagen, Denmark, only 48% of those referred to municipality-based rehabilitation participate in the group-based strength and cardiovascular exercise that is available. Studies have shown that reasons for not participating in exercise interventions can be both cancer-specific (ie, symptoms such as fatigue and pain) and situational/environmental (eg, distance to exercise facilities, time of day classes are held, or other commitments) [13,14].

It is estimated that the number of cancer survivors in need of cancer rehabilitation will increase in the coming years due to increasing cancer incidence and higher survival rates [15,16]. The increasing prevalence of cancer survivors and scarce resources (ie, increasing costs and declining health professional workforce) challenge rehabilitative services, for instance, in the form of longer waiting lists and shorter appointment times per patient for in-person services [17].

Digital technology can perhaps help cancer survivors increase their level of physical activity as technology can serve to resolve the geographical and logistical obstacles associated with traditional programs requiring in-person supervision and communication, increasing the efficiency of rehabilitative services [18-20]. Examples of digital technology interventions are applications offered via smartphones and websites, as well as wearables that can be used to instruct, monitor, or motivate physical activity [21-27]. When introducing digital services, it may be important that the provider or service organization understands how and to what extent the digital technology may be beneficial for the individual. Such a stratification will play an important role in requirement specifications for digital solution providers [28]. In addition, when identifying and excluding people unable to take advantage of technology, it may be possible to allocate additional resources for in-person supported services for this group.

We previously reported how cancer survivors referred to rehabilitation can be stratified into four distinct profiles according to their health technology readiness using the Readiness and Enablement Index for Health Technology (READHY) [29], an instrument based on the eHealth Literacy Framework [30] and the eHealth Literacy Questionnaire [31], supplemented by the social dimensions of the Health Literacy Questionnaire [32] and by the self-management dimensions of the Health Education Impact Questionnaire [33]. The dimensions relating to electronic health (eHealth) literacy described user knowledge and skills, the intersection between users and technologies, and users' experience of systems. The dimensions relating to self-management add knowledge about the individuals' ability to handle their condition and emotional response. The social dimensions add knowledge about the individuals' social context (ie, support from family and friends or health professionals) [29]. The four identified READHY profiles (Figure 1) differ regarding their receptiveness to physical activity rehabilitation supplemented by digital technology. In this context, receptiveness is based on whether the interviewees can imagine supplementing physical exercise with technology (eg, a smartphone, computer, or smartwatch). This means that receptiveness not only addresses their technology readiness but also their intention to perform or their attitude toward physical activity [34].

Figure 1. Four health technology readiness profiles and their receptiveness to supplement physical activity during rehabilitation with technology. The READHY scale ranges from 1 (Strongly disagree) to 4 (Strongly agree). Within each dimension, the average READHY scores are color coded relative to the other profiles from red (lowest score) to green (highest score). Adapted from [29,34]. eHLQ: eHealth Literacy Questionnaire; HeiQ: Health Education Impact Questionnaire; HLQ: Health Literacy Questionnaire; READHY: Readiness and Enablement Index for Health Technology.

Parameters			Profiles			
			1	2	3	4
READHY dimensions, mean						
	Self-management					
		HeiQ3 (Self-monitoring and insight)	2.74	2.83	3.15	3.21
		HeiQ4 (Constructive attitudes and approaches)	2.85	2.88	3.51	3.39
		HeiQ5 (Skills and Technique Acquisition)	2.65	2.79	3.25	3.32
		HeiQ8 (Emotional distress; reversed)	2.60	2.62	3.18	2.98
	Support					
		HLQ1 (Feeling understood and supported by healthcare providers)	3.01	2.94	3.50	3.46
		HLQ4 (Social support for health)	3.01	3.19	3.74	3.68
	eHealth literacy					
		eHLQ1 (Using technology to process health information)	1.53	2.75	2.31	3.52
		eHLQ2 (Understanding of health concepts and language)	2.68	2.91	3.18	3.61
		eHLQ3 (Ability to actively engage with digital services)	1.70	3.00	2.89	3.64
		eHLQ4 (Feel safe and in control)	2.88	2.90	3.21	3.56
		eHLQ5 (Motivated to engage with digital services)	1.81	2.65	2.35	3.42
		eHLQ6 (Access to digital services that work)	2.09	2.74	2.77	3.36
		eHLQ7 (Digital services that suit individual needs)	1.74	2.60	2.42	3.29
Receptiveness, %						
	Receptive		23.7	83.1	63.6	84.2
	Unreceptive		76.3	16.9	36.4	15.8

The quantitative approach provided important information about the four different profiles and their level of readiness for technology. The stratification of users can serve designers and service providers in addressing the various profiles of service users in relation to their overall characteristics. However, patient receptiveness to technology-supported physical activity cannot be explained solely by their technology readiness level. Other aspects, such as the cancer survivors' underlying assumptions and reasoning, may facilitate or constitute a barrier to usage of health technology [35], and should be more thoroughly investigated. Therefore, the purpose of this study is to enhance knowledge on receptiveness to technology-assisted physical activity in cancer survivors and to give a broad spectrum of cancer survivors a voice by taking their various levels of readiness into consideration. Consequently, our research questions are the following:

1. What are the cancer survivors' assumptions and reasoning regarding health technology in relation to physical activity?
2. How may these assumptions influence their receptiveness to using digital technology in relation to their physical activity during rehabilitation?
3. Do needs and preferences vary between the profiles identified?

A qualitative approach can provide a more nuanced description of the profiles [36], enabling us to identify differences that can be used to tailor services and identify those who may not benefit from technology in physical activity rehabilitation programs.

Methods

The methods section is reported according to the Consolidated Criteria for Reporting Qualitative Research (COREQ) [37].

Setting

This study took place at the Copenhagen Centre for Cancer and Health in Copenhagen, Denmark. In 2007, the main responsibility for the rehabilitation of patients with cancer in Denmark was transferred from the regional level (hospital management) to the municipal level, in accordance with the current trend of offering more people-centered services [38]. The Copenhagen Centre for Cancer and Health provides interdisciplinary rehabilitation of patients with cancer that includes group-based strength and cardiovascular training and individually tailored rehabilitative exercises and health education, as well as individual counseling in relation to health problems, diet, occupation, and economic issues. Since cancer survivors can be referred to rehabilitation in their initial treatment period, many of them are undergoing active treatment while participating in rehabilitation programs.

Study Design and Interview Participants

This qualitative study is part of a large cross-sectional study exploring health technology readiness in 305 patients with

cancer referred to public rehabilitation services at the Copenhagen Centre for Cancer and Health [29,34]. Briefly, participants did the READHY questionnaire measuring self-management, social support, and eHealth literacy, in addition to filling out a background information questionnaire. Subsequent cluster analysis identified four cluster profiles differing regarding their health technology readiness profiles and sociodemographic variables, such as age, education, number of chronic conditions, and receptiveness to technology in relation to physical activity [29,34]. Our goal was to invite three participants from each of the four profiles for interviews. Potential interviewees were purposely selected for interviews by the first author, SR (female), based on their READHY profile, sex, and age, without any other characteristics taken into consideration, including diagnosis. In total, 23 (12 females, 11 males) of the 305 patients were invited by mail and then contacted by phone after 7 days by a staff member at the Copenhagen Centre for Cancer and Health to inquire about participation. Those who agreed to participate (n=11, 8 females, 3 males) were contacted by SR to make an interview appointment. Of those who did not participate (n=12), 5 could not be reached after 3 attempts by phone, and 7 declined participation. Reasons for declining were not requested. Interviewees received oral and written information about the study and gave written informed consent before being interviewed. Interviewees were interviewed between June 2017 and October 2017 by SR. Interviews were conducted in an interview room at the center (n=9) or at the interviewee's own home (n=2). No one else was present during the interviews, and interviewees were not compensated for their participation. The interviewer knew which READHY profile the interviewee belonged to but focused on keeping an open mind during the interviews. No repeat interviews were conducted. The interviewer is a postdoctoral fellow at the Centre for Physical Activity Research at Rigshospitalet, University of Copenhagen Hospital, and was supervised by author LK from the Department of Public Health, University of Copenhagen. The semistructured interview guide with prompts was designed to enable a deeper exploration of our previous findings and to provide us with insight into the cancer survivors' underlying assumptions and reasonings regarding their receptiveness to using technology in relation to physical activity during rehabilitation. As a result, we based the questions on the READHY questionnaire addressing self-management, including the impact of their condition; social support from health professionals and their personal network; and their eHealth literacy. We also included questions on exercise/physical activity and digital technology–assisted physical activity (Multimedia Appendix 1). A digital audio recording of each interview, which lasted 15 to 45 minutes (mean 28 minutes), was made and transcribed verbatim.

Data Analysis

Interviews were analyzed using directed content analysis [39,40]. All interview transcripts were read and reread by SR to ensure familiarity with the data. We used an abductive approach [40]. The first step was deductive based on the READHY framework; a codebook [41] was constructed based on three concepts in the READHY framework (self-management, support, and eHealth literacy) and receptiveness to technology-assisted physical activity (Multimedia Appendix 2). SR then marked all the passages in the transcripts that appeared to relate to the codes. NVivo 12 (QSR International) was used to organize and code the data. After coding, SR and LK identified two themes. In the second step, which was inductive, SR and LK revisited the text and codes to identify subcategories within the themes and achieved consensus using abstraction and interpretation processes [40]. In the abstraction process, it became evident that when discussing physical activity in connection with rehabilitation, interviewees did not distinguish between exercise and specialized disease-specific exercises. For the interpretation, the category physical activity was revisited by SR using the following: (1) the term "rehabilitative exercise" for specialized cancer-specific exercises to alleviate disease or treatment-related problems (eg, mobilization of scar tissue, self-managed manual lymph drainage, and swallowing exercises for dysphagia) and (2) Caspersen and colleagues' [42] definition of physical activity ("bodily movement produced by skeletal muscles that results in energy expenditure"), which can be further classified as incidental or structured. Structured physical activity or exercise is a subset of physical activity that is planned, structured, and repetitive, with the purpose of improving fitness or health [42,43]. Incidental physical activity is unstructured activity that is part of daily living at work or at home, such as walking or cycling for transport, climbing stairs, and doing housework [44]. SR and LK selected the quotations presented in this study.

Ethics

According to Danish law, formal ethical approval was not required because no biological material was obtained in the study. The study was conducted in accordance with the Helsinki Declaration and approved by the Danish Data Protection Agency (2015-55-0630). All participants were informed about the study before starting the interviews, received a participant information sheet, and were informed that their participation was voluntary, that they were ensured anonymity, and that all data would be handled confidentially. Written informed consent was obtained.

Results

Overview

In total, 8 women and 3 men were interviewed; they ranged from 32 to 82 years of age and represented all four READHY profiles (Table 1). Based on their statements, the interviewees were further characterized as being users (or nonusers) of digital technology for physical activity if they stated that they used or had used (or not) smartphone apps, websites, or wearables for physical activity. Interviewees generally owned technological devices, eg, smartphones, tablets, or computers, using them for various everyday purposes. Of the 11 interviewees, 5 were receptive to using technology for physical activity during rehabilitation. In total, 5 of the 11 participated in strength and cardiovascular exercise at the center (75 minutes twice weekly for 16 weeks). The interviewees represented 8 different International Classification of Diseases Version 10 cancer sites: lip, oral cavity, and pharynx (C00-C14; n=1); digestive organs (C15-C26; n=2); respiratory and intrathoracic organs (C30-C39;

n=1); breast (C50; n=3); urinary tract (C64-C68; n=1); eye, brain, and other parts of central nervous system (C69-C72; n=1); lymphoid, hematopoietic, and related tissue (C81-C96; n=1), and ill-defined, secondary, and unspecified site (C76-C80; n=1). At the time of the interview, 2 participants were still undergoing active treatment. Most of the noninterviewees were male (8 males, 4 females). They ranged in age from 28 to 70 years (mean 56 years), and generally owned smartphones, tablets, or computers, using them for various everyday purposes (Multimedia Appendix 3).

Table 1. Characteristics of interview participants.

ID	Sex	Age	READHY profile[a]	Physical activity digital technology usage[b]	Receptive[c]	Technology ownership[c,d]			Usage[c]	Purpose of using technology[c,e]				
						S	T	C		W	IS	Cm	P	E
1	F	82	1	Nonuser	No			✓	Several times daily				✓	
2	F	69	1	Nonuser	No	✓			Several times daily			✓		
3	F	43	2	Nonuser	Yes	✓	✓	✓	Several times daily	✓	✓	✓	✓	
4	F	71	2	User	Yes	✓	✓	✓	Several times daily		✓		✓	✓
5	F	65	3	User	No	✓	✓	✓	Several times daily	✓	✓	✓	✓	
6	M	32	3	User	Yes	✓	✓	✓	Several times daily	✓	✓	✓		✓
7	F	63	3	Nonuser	No	✓	✓	✓	Once per day		✓	✓		
8	F	66	3	User	Yes	✓	✓	✓	Several times daily		✓	✓	✓	✓
9	M	64	4	Nonuser	No	✓	✓	✓	Several times daily	✓	✓	✓	✓	
10	M	36	4	User	Yes	✓		✓	Several times daily	✓	✓	✓	✓	
11	F	67	4	Nonuser	No		✓	✓	A few times per week			✓		

[a]READHY: Readiness and Enablement Index for Health Technology.

[b]This is based on the interviewees' statements.

[c]These are based on questionnaire data reported in [34].

[d]S: smartphone; T: tablet; C: computer.

[e]W: work; IS: information seeking; Cm: communication; P: practicality; E: exercise.

During the content analysis, we identified four subcategories as important to understanding the interviewees' assumptions and reasonings in relation to digital technology-assisted rehabilitation, divided into the following two themes: (1) attitude toward physical activity within the subcategories incidental and structured physical activity and social relations (ie, with other cancer survivors, exercise participants, or health professionals when performing physical activity); and (2) attitude toward technology-assisted physical activity within the subcategories motivation and prerequisites. Table 2 provides a sample of quotes illustrating the two themes and related subcategories.

Table 2. Quotes illustrating the "attitude toward physical activity" and "attitude toward technology-assisted physical activity" themes.

Attitudes and subcategories	Profile 1 quotes	Profile 2 quotes	Profile 3 quotes	Profile 4 quotes
Attitude toward physical activity				
Incidental or structured (exercise) physical activity				
	I've never done that [exercised]. Well, I did gymnastics back in the day. I've done sports, right? I've skied a lot. Slalom, up and down white mountains, that sort of stuff, you know? I don't do that anymore, but I haven't had the need. You didn't do that when I was a kid. You didn't exercise, you climbed trees and stuff, you know? [Female, profile 1, ID1] *It can be hard to motivate yourself to exercise in this [cancer disease], and what you're going through with the disease and stuff. So, it's fine [exercising at the center], and it got me started exercising. I'm quite happy about that, even though I think every time, "Oh no," you know? But you're quite contented when you leave and think "That's probably why I feel fine." Because, as I said to you, I get chemo and I don't feel any side effects. […] I don't want to go to a fitness center.* [Female, profile 1, ID2]	*I've been extended for two more months [team-based exercise at the center] because they know if they let me go, I won't exercise. […]. Because, I'm sorry to say, I don't like to do exercise. […] I think it hurts. I think it's hard. I hate sweating.* [Female, profile 2, ID3] *I was in a really good physical shape before […] I'm good at riding my bike to go for a swim in the morning [at the beach]; I do winter swimming, and I'm active.* [Female, profile 2, ID4]	*I dance at a fairly high level and do yoga on a fairly high level.* [Female, profile 3, ID5] *I feel agile enough, but, but, I should, you know [do regular exercise]? And I don't like those centers at all [fitness centers]. […] I felt so happy about coming here [rehabilitation center], so it wasn't a problem getting up in the morning in the middle of the winter and biking here twice a week. […] I haven't done much exercise. I really haven't attended anything other than yoga. No hard physical exercise. I've just rushed around in my everyday life, you see?* [Female, profile 3, ID7] *I used to work out every other day before. Strength exercises. […] [The physiotherapist] gave me some exercises I can do at home and I have an app called 7-minute workout. That actually works really well to get a little exercise. Biking to work, going for long walks*. [Male, profile 3, ID6] *For the past 30 years I've exercised my left side to keep it going [rehabilitative exercises]. And I just continued that on the right side. Little exercises at first and gradually more, you know? […] Even on days when you think you haven't walked much, and we've just been at home. We've walked 6-7,000 steps anyway, you know? Well, we have a little yard. I take care of it myself, and there are a lot of stairs.* [Female, profile 3, ID8]	*I haven't been able to do badminton and tennis like I used to. But I've walked or biked to work. […] At work we have exercise facilities. So, I could go there after work or during my lunch break or whatever suited me. […] In the beginning you were sick. You had to do it. In some way, doing it was more legitimate. Now I'm not sick; then there's so much other stuff.* [Male, profile 4, ID10] *I don't want to do exercise [in a gym] in the summer when I have a small plot of land. I do plenty in the yard, mowing the lawn, cutting hedges, biking back and forth, and other stuff. I have a deal with the fitness center that I take the summer months off. And then from October to May I do exercise [in the fitness center].* [Female, profile 4, ID11] *I went to a fitness center before, and I continued after the operation. And used the exercises I got from here [the rehabilitation center]. […] In the summer, I don't go [to the fitness center]. I bike, and I play golf and other stuff. I think I get enough exercise. But, in October, I start [going to the fitness center].* [Male (profile 4, ID9]
Social relations				

Attitudes and subcategories	Profile 1 quotes	Profile 2 quotes	Profile 3 quotes	Profile 4 quotes
	When you're walking, of course you have to be aware of not just trudging along, but I usually think it's cozier walking with someone. [Female, profile 1, ID2]	*[…] If it's on my calendar, if I had an appointment with you. Well, then I would do it. It could also just be a personal trainer who leads me through the first program. Just to say, "Well okay, she's shown up." Just like you do [rehabilitation center]. Then I would go.* [Female, profile 2, ID3] *[…] I can't see myself sitting at home doing exercise, I'm too social for that. […] I appreciate the other women who I exercise with.* [Female, profile 2, ID4]	*[…] there's a sense of safety in the almost family-like atmosphere when you go through the door. […] You see people's hair grow, and they get color in their cheeks, and you see someone who is dragging themselves along, and you think, "I'm glad that's not how I'm feeling anymore." […] When you sweat together, then you have something to share. And you have a little chat afterwards over a cup of coffee.* [Female, profile 3, ID7] *For me, it's important that we're doing it together [exercising].* [Female profile 3, ID5]	
Attitude toward technology-assisted physical activity				
Motivation				
	If anyone paces me, tells me what to do, I get annoyed and obstinate. I'll decide that myself. And I'll do it in my own pace. Well, I would say no thanks. But I haven't tried. But I could imagine that was how I would react. Yes, yes [laughs]. [...] I can just say, "Okay, I can delete the app." Or I can say, "Hey, nobody's there." So, it probably has to be more personal, if I have to do it. [Female, profile 2, ID3] *It's hypothetical, right? Because, again, will I actually do it? [use, eg, an app to exercise].* [Female, profile 2, ID4]	*I thought, "that might be fun." The other day, I was walking with my daughter and she said to me, "Mom, we've walked 4.6 miles." "Super," I replied. You can see that on the, what's it called, the GPS thing. […] Because it tells you how far you've walked and stuff, you know? And I think that's clever. But stuff like my heartrate and all that. I don't want that.* [Female, profile 1, ID2]	*Deep down, I want to use it because I can see that it's convenient for many people. But you wouldn't get the social dimension that you get in this building, you know? [rehabilitation center]. […] I have a hard time finding a place I want to go and continue my exercise. If I knew, I would spend half an hour and turn on my phone every morning. Why don't I do that? Then, I wouldn't have to do it at five pm, in the rain, in November, but in the morning when it suits me the best.* [Female profile 3, ID7] *I don't think I would want to do that at all [use an app for exercise]. […] My attitude towards using my body is that it has to be fun. And, it has to be entertaining, and it has to be nice. For me, I'm not interested in how high my pulse gets. Or how many pounds I lift. [..] I don't think an app should decide how I should move. […]* [Female, profile 3, ID5]	*If it works [an app], I might consider using it. But I don't know what I should use it for [because she keeps herself physically active all the time]. [Female, profile 4, ID11]* *I would ask the instructor [fitness instructor], "Would you do a training/exercise program for me?" I haven't thought about using an app. […] It might be a good idea. I just haven't thought about it. I always run around with a piece of paper in my bag with my exercises on it.* [Male, profile 4, ID9]
Prerequisites				

Attitudes and subcategories	Profile 1 quotes	Profile 2 quotes	Profile 3 quotes	Profile 4 quotes
		Well, it should be what I'm writing on this paper now. Date, how many pounds, and how many times I've lifted, and stuff like that. […] There's this fantastic game. It's a mental exercise app. And there's this curve that shows that I didn't do well today. It motivates me by saying at 9 am, "It's time to do your games." And I think that's exciting. And I can see that something's happening, and I can see when it isn't. Then it's a bit fun, and I can also see if I'm progressing. [Female, profile 2, ID3]	*I think, if it were something I could do at home, some sort of morning program. I know someone who has an aerobic step in front of their tv, you know? And then they get it done in the morning. It doesn't matter what you do. I would easily have the time to do that for half an hour in the morning.* [Female profile 3, ID 7] *If I had to use an app, then maybe it could be a yoga app […] Yoga is a guided activity, and you can do it yourself, but you have to know exactly how to combine them [the exercises] if you want to do something you're not used to.* [Female, profile 3, ID5] *You should be able to adjust the settings to, for example, I want to do it in the morning, at noon, and in the evening, like I do now. And then it goes "beep, beep," remember your exercises [rehabilitative exercises]. And you could bring it to work. It's always in your pocket, you see? […] It could be smart if you could call or send a message to your main contact person [at the rehabilitation center]. […] It's good exercise, where you get support and you have like a coach saying to do ten push-ups. The app [7-minute workout] actually says, "Do ten push-ups." Then you can put it on your bookcase, clear some space in theliving room, and do all the exercises. That works really well. And you can do it at 5:30 am or 3:45 pm.* [Male, profile 3, ID6] *Maybe more enthusiasm. It's hard to explain, but one thing is that you're shown the exercises. But I also need to know why. […] Why is this exercise better than another one […] A little more depth. And I think technology would be good at/for that.* [Female, profile 3, ID8]	*It should contain instructions on the exercises I should do. I mean, how they look. Describe how they look. So that you can see what you're supposed to do, you know? […] That would make sense to me. That would be clever, because I already use my phone to listen to music when I exercise. Then I could just check, "What's the next exercise?" That would be fantastic.* [Male, profile 4, ID9] *These exercises are especially good; you should do them like this and this." I mean, how much you should do or how hard it should be, but it should make sense to you. It has to be something where I've been asked or assessed, "This is important for you." […] If there are three days a week where I should do it, what happens if I skip a day? Should I do more then? […] Where you understand what the different exercises [rehabilitative exercises] do, and if you don't' do them, what should you do instead, and what are the consequences? […] It could also be fun to record my weight, repetitions, how much, and then see if there is any progress.* [Male, profile 4, ID10]

Attitude Toward Physical Activity

The theme "attitude toward physical activity" is related to what the participants think about physical activity and how the participants prefer to be physically active (ie, incidental or structured physical activity, in a social context or individually).

Incidental and Structured Physical Activity

Across all profiles, 8 interviewees talked about their incidental physical activity in relation to both transportation (eg, bicycling) and leisure time (eg, gardening, using the stairs). They considered these activities as a supplement or a substitute for exercise (structured physical activity). In profiles 1 and 2, (low self-management, low to medium eHealth literacy), 3 of 4 interviewees were reluctant to participate in organized physical

activity, whereas the fourth was accustomed to organized team-based exercise (fitness center, yoga), emphasizing the social aspect of structured physical activity. In profiles 3 and 4 (high self-management, medium to high eHealth literacy), 5 of 7 interviewees were accustomed to participating in structured physical activity. All 3 interviewees in profile 4 had performed structured physical activity regularly before the cancer and complementary to the services offered by the rehabilitation center. In addition, two of the profile 4 interviewees mentioned that the time of year affected their activity level; they did not use structured physical activity facilities in the summer because they were generally more active in their everyday lives.

Social Relations

This context refers to social relationships with other cancer survivors or exercise participants and health professionals or trainers in connection with performing physical activity. There was a tendency among the individuals in profiles 1 and 2 to mention social relations as important. For example, 3 of 4 interviewees in profiles 1 and 2 (low self-management, low to medium eHealth literacy), and one interviewee in profile 3, specifically mentioned social relations as important (either social relationships with other exercise participants, social relationships with other cancer survivors, or coaching from staff and being accountable to others). None of the other interviewees mentioned the social aspect of physical activity as important to them.

Attitude Toward Technology-Assisted Physical Activity

The theme "attitude toward technology-assisted physical activity" is related to the interviewees' motivation to use technology for physical activity and the prerequisites they have if they were to use technology for physical activity. One interviewee (ID1, low self-management, low eHealth literacy) was not open to the idea of exercising and, as such, could not relate to rehabilitation assisted by digital technology. Among the remaining 10 interviewees, none of them clearly rejected digital technology–assisted physical activity.

Motivation

The participants' motivation for using digital technology (eg, an app or website) in connection with physical activity ranged from being open-minded to being reluctant. For example, 2 interviewees (profile 1, low self-management and eHealth literacy; profile 4, high self-management and eHealth literacy) were nonusers of technology in connection with physical activity and had never thought about using it but were open to the idea. They saw the potential benefits of using it, such as being aware of the miles walked and not having to run around with a piece of paper describing the exercises. However, 5 interviewees (profiles 2 and 3, medium self-management and eHealth literacy; profile 4, high self-management and eHealth literacy) were reluctant to use digital technology in connection with structured physical activity because they would miss the social aspect of exercising, were active already, and could not see the purpose of an app or website; they did not like to be told what to do by, for example, an app, and held the assumption that the role of technology is to provide data about activity rather than to contribute to a good experience. However, 3 individuals that were reluctant to use digital technology already used or had used digital technology for physical activity (eg, to count steps using their phone's built-in pedometer or to watch online exercise or rehabilitative exercise instruction videos).

Prerequisites

The subcategory "prerequisites" includes the interviewees' preferences for instruction, information, and setting in relation to the use of technology for physical activity. Among participants, 7 interviewees, mainly from profiles 3 and 4 (high self-management, medium to high eHealth literacy), and mainly users of digital technology for physical activity (both structured and incidental) had preferences about what they would like digital technology to do, if they were to use it during rehabilitation. Some interviewees focused on technology for their rehabilitative exercises, with the aim of improving or maintaining a specific function (eg, being reminded of swallowing exercises or memory exercises, and instructions on how to perform rehabilitative exercises to regain function after an operation). Others talked about technology as a support for structured physical activity, like yoga or cardiovascular exercises, either to be used in their own home or in a fitness center. In general, profiles 3 and 4 (high self-management, medium to high eHealth literacy) expressed a need for support to make decisions in connection with training using technology. They wanted to know why they should do physical activity and what the benefits of exercise are. For example, 2 interviewees said that digital technology could provide information about why they should do exercise, not just how to exercise, and one interviewee believed this might increase her enthusiasm for exercising. One interviewee mentioned the relevance of a yoga app because yoga demands instruction and knowledge of how to combine postures. In addition, 2 interviewees mentioned the availability of visual instructions to be able to see how the various exercises are performed. One interviewee thought that technology may contribute to devising an individualized plan based on an assessment of his condition and this plan could be adapted to day-to-day changes in activity levels. The convenience of being able to exercise on demand when using technology and monitoring progression was also mentioned. One interviewee mentioned reminders of when to do rehabilitative exercises and the possibility of getting in contact with the center contact person/health professional directly through an app.

Discussion

Principal Findings

In this study, we give cancer survivors a voice in a rehabilitation setting to understand their assumptions and reasonings about digital technology–assisted physical activity in general, and its possible implications for the future design and introduction of technology. Recruiting individuals from each of the four health technology readiness profiles ensured that perspectives from a wide range of cancer survivors were included. We identified two important themes in terms of cancer survivors' receptiveness to using digital technology in relation to their physical activity during rehabilitation: (1) attitude toward physical activity, with the subcategories "incidental and structured physical activity" and "social relations," and (2) attitude toward

technology-assisted physical activity, with the subcategories "motivation" and "prerequisites." The two themes and related subcategories describe the thinking of the interviewees and contribute to an understanding of how interventions can be designed to address both attitude toward exercise and attitude toward technology. Our previous quantitative data [34] suggests that not everyone is receptive to and able to use digital technologies to increase and maintain physical activity in connection with rehabilitation. This is supported by our interviewees. The results indicate that it is important to address the cancer survivors' motivation (ie, the degree to which they are open-minded about using technology for physical activity). Another aspect of motivation is how to include the cancer survivors' particular needs, such as their preferences, in relation to physical activity, in terms of the following: (1) incidental or structured physical activity (where structured activity includes both cardiovascular and strength exercise or disease-specific rehabilitative exercises); (2) social or individual context; and (3) instruction (know-how) or information (know-why). Notably, some of the cancer survivors assumed that technology would reduce their ability to fulfill their needs (eg, enough social interaction or having a good experience). In relation to preferences, the context is important (eg, time and place; fitness center, rehabilitation center, or at home; summer or winter). Furthermore, the results indicate that those scoring lower on health technology readiness prefer activities that have a social component, whereas those scoring higher prefer to participate in individual activities (fitness center, use apps) and seek an understanding of why they should exercise and receive personalized support. We were able to identify information that can help providers offer stratified services that take the individual's different needs and preferences into consideration. The identified preferences provide new insight that complements the cancer survivors' readiness level and may also help designers, service providers, and caregivers to provide solutions that increase receptiveness toward technology-assisted physical activity.

Comparison With Other Studies

Recently, 3 studies explored health technology in relation to physical activity in cancer survivors [24,25,27]. Nielsen et al [27] found that women with breast cancer undergoing chemotherapy were enthusiastic about a smartphone application designed to help them self-monitor activity between coaching sessions, and that they wanted to include family and friends and have access to a personalized tailored application. In alignment with this, Puszkiewicz et al [25] found that colorectal, breast, and prostate cancer survivors who used a generic app to support physical activity during a six-week intervention felt that an application for cancer survivors should be tailored to the individual's lifestyle and enable social support from friends and family. Robertson et al [24] added another dimension to these findings with a focus on tools for personal goal setting to support personally held priorities and values. In all 3 studies, recruitment appeared to favor a selection of interviewees familiar with technology and who appeared to be more homogenous than our population. Our approach of also including cancer survivors who were less digitally ready may provide us with broader insight regarding the various needs and preferences of cancer survivors. Although all our interviewees would benefit from stratification, non–digital solutions, varying degrees of social interaction, and incidental physical activity also play a role, just as it must be kept in mind that not everyone wanted to know why they should be physically active.

Implications

A recent synthesis of the acceptability and engagement of web-based interventions for cancer survivors suggests that future work should also involve identifying the optimal stage of cancer survivorship to facilitate intervention delivery [45]. We suggest, in contrast, that technology be introduced to cancer survivors at any stage of the cancer trajectory but that health professionals must take into account the identified patient preferences together with their readiness level. This could be done by either assisting cancer survivors in selecting applications from a library based on a declaration of specific content as described by Short et al [26] or by designing applications [25] that can adapt to the cancer survivor's preferences and readiness level. By hosting a variety of existing applications [46] selected to cover the various readiness levels and preferences, more advanced digital platforms could also serve to aid adoption of technology at any stage of the cancer trajectory. Some platform applications should support social interaction, while others should aim to provide an underlying understanding of the activities and how they contribute to the cancer survivors' health, increasing their ability to manage their condition. For those high in self-management, more individualized applications should be available and potentially incorporate personal coaching. For those not interested in structured programs, incentives to increase incidental activity should be provided to enable cancer survivors to create a palette of services better suited to their individual needs.

The data presented here and in our previous studies relates to a broad spectrum of cancer survivors, including cancers related to a high sociodemographic status (eg, breast and prostate cancer) and to a low status (eg, lung cancer and some head and neck cancers) [47]. The present population also consisted of individuals with one or more long-term conditions. Although cancer survivors may have specific needs in relation to treatment, such as prevention of lymphedema, the proposed model for digital technology–assisted physical activity during rehabilitation is likely to meet their personal needs and competences in other settings. Currently, we are exploring the READHY profile of other groups (eg, patients with type 2 diabetes mellitus and cardiovascular diseases). Qualitative studies with additional samples and other patient groups may contribute to identifying whether or not our findings are generalizable or specific to our sample.

Strengths and Limitations

A strength of this study is that recruitment was based on stratification into READHY profiles, ensuring a broad representation of individuals regarding eHealth literacy and self-management. In addition, participants represented several different cancer types, which increases the generalizability of the findings. This study also has limitations. For example, analysis of the interview content originally applied deductive coding, but in the iterative process of creating the subcategories,

we came closer to an abductive [40] workflow less influenced by the initial READHY structure. This may introduce bias but may also be a strength as the abductive approach offers an opportunity to interpret our findings in a new way in the process of analyzing the data. We fully acknowledge that other factors also need to be addressed when discussing the likelihood of adoption, such as perceived ease of use and usefulness, usability, and the user experience. We did not recruit many male participants, and we did not recruit as many participants from each profile as planned (3 per profile). This may have caused important perspectives on incidental and structured technology-assisted physical activity to be overlooked. In our previous quantitative study, we showed that age and education are associated with readiness for technology, while sex is not. The age and educational level of the interviewees may therefore influence perspectives. By discussing our results in relation to the READHY profile, we hope to have adjusted for this to a certain extent. Although the interviewer was not blinded to the profile of the interviewees, keeping an open mind was a strict priority.

Conclusion

Combining digital technology based on the cancer survivors' needs, preferences, and readiness with capacity building of the workforce will aid in tailoring digital solutions to suit individuals able and receptive to using them but also to those reluctant to do so. This may contribute to expanding the number of cancer survivors able to take advantage of technology-assisted services and to allocating more in-person resources to those unable to take advantage of technology. For individuals who are unreceptive to using technology, the awareness of the variations in preferences regarding incidental and structured physical activity, social interaction, and the need to receive an explanation about why exercises are important may also help motivate more cancer survivors to participate in physical activity during rehabilitation without digital support.

Acknowledgments

The authors would like to thank the interviewees who shared their views on physical activity and digital technology–assisted exercise. We also thank Emilia Lüders for transcribing the interviews. This work was supported by the Danish foundation TrygFonden. The Centre for Physical Activity Research is supported by a grant from TrygFonden.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 23 KB - jmir_v22i8e15335_app1.docx]

Multimedia Appendix 2
Codebook.
[DOCX File , 23 KB - jmir_v22i8e15335_app2.docx]

Multimedia Appendix 3
Noninformant characteristics.
[DOCX File , 27 KB - jmir_v22i8e15335_app3.docx]

References

1. Ammitzbøll G, Søgaard K, Karlsen RV, Tjønneland A, Johansen C, Frederiksen K, et al. Physical activity and survival in breast cancer. Eur J Cancer 2016 Dec;66:67-74. [doi: 10.1016/j.ejca.2016.07.010] [Medline: 27529756]
2. Ballard-Barbash R, Friedenreich CM, Courneya KS, Siddiqi SM, McTiernan A, Alfano CM. Physical activity, biomarkers, and disease outcomes in cancer survivors: a systematic review. J Natl Cancer Inst 2012 Jun 06;104(11):815-840 [FREE Full text] [doi: 10.1093/jnci/djs207] [Medline: 22570317]
3. Ibrahim EM, Al-Homaidh A. Physical activity and survival after breast cancer diagnosis: meta-analysis of published studies. Med Oncol 2011 Sep;28(3):753-765. [doi: 10.1007/s12032-010-9536-x] [Medline: 20411366]
4. Chlebowski RT. Nutrition and physical activity influence on breast cancer incidence and outcome. Breast 2013 Aug;22 Suppl 2:S30-S37. [doi: 10.1016/j.breast.2013.07.006] [Medline: 24074789]
5. Cormie P, Zopf EM, Zhang X, Schmitz KH. The Impact of Exercise on Cancer Mortality, Recurrence, and Treatment-Related Adverse Effects. Epidemiol Rev 2017 Jan 01;39(1):71-92. [doi: 10.1093/epirev/mxx007] [Medline: 28453622]
6. Tomlinson D, Diorio C, Beyene J, Sung L. Effect of exercise on cancer-related fatigue: a meta-analysis. Am J Phys Med Rehabil 2014 Aug;93(8):675-686. [doi: 10.1097/PHM.0000000000000083] [Medline: 24743466]
7. Buffart LM, Kalter J, Sweegers MG, Courneya KS, Newton RU, Aaronson NK, et al. Effects and moderators of exercise on quality of life and physical function in patients with cancer: An individual patient data meta-analysis of 34 RCTs. Cancer Treat Rev 2017 Jan;52:91-104 [FREE Full text] [doi: 10.1016/j.ctrv.2016.11.010] [Medline: 28006694]

8. Fairman CM, Focht BC, Lucas AR, Lustberg MB. Effects of exercise interventions during different treatments in breast cancer. J Community Support Oncol 2016 May;14(5):200-209 [FREE Full text] [doi: 10.12788/jcso.0225] [Medline: 27258052]
9. World Health Organization. Global recommendations on physical activity for health. Geneva, Switzerland: World Health Organization; 2010.
10. Schmitz KH, Courneya KS, Matthews C, Demark-Wahnefried W, Galvão DA, Pinto BM, American College of Sports Medicine. American College of Sports Medicine roundtable on exercise guidelines for cancer survivors. Med Sci Sports Exerc 2010 Jul;42(7):1409-1426. [doi: 10.1249/MSS.0b013e3181e0c112] [Medline: 20559064]
11. Danish Health Authority. Forløbsprogram for rehabilitering og palliation i forbindelse med kræft - del afsamlet forløbsprogram for kræft. Copenhagen, Denmark; 2012. URL: https://stevns.dk/sites/default/files/hoeringsversionfp_kraeft_09022018.pdf [accessed 2020-07-04]
12. Troeschel AN, Leach CR, Shuval K, Stein KD, Patel AV. Physical Activity in Cancer Survivors During. Prev Chronic Dis 2018 May 24;15:E65 [FREE Full text] [doi: 10.5888/pcd15.170277] [Medline: 29806579]
13. Blaney JM, Lowe-Strong A, Rankin-Watt J, Campbell A, Gracey JH. Cancer survivors' exercise barriers, facilitators and preferences in the context of fatigue, quality of life and physical activity participation: a questionnaire-survey. Psychooncology 2013 Jan;22(1):186-194. [doi: 10.1002/pon.2072] [Medline: 23296635]
14. Wurz A, St-Aubin A, Brunet J. Breast cancer survivors' barriers and motives for participating in a group-based physical activity program offered in the community. Support Care Cancer 2015 Aug;23(8):2407-2416 [FREE Full text] [doi: 10.1007/s00520-014-2596-2] [Medline: 25605568]
15. Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C, ICBP Module 1 Working Group. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. Lancet 2011 Jan 08;377(9760):127-138 [FREE Full text] [doi: 10.1016/S0140-6736(10)62231-3] [Medline: 21183212]
16. Allemani C, Weir HK, Carreira H, Harewood R, Spika D, Wang X, et al. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). Lancet 2015 Mar 14;385(9972):977-1010 [FREE Full text] [doi: 10.1016/S0140-6736(14)62038-9] [Medline: 25467588]
17. Johansen C. Rehabilitation of cancer patients - research perspectives. Acta Oncol 2007 Jul;46(4):441-445. [doi: 10.1080/02841860701316057] [Medline: 17497310]
18. Dickinson R, Hall S, Sinclair JE, Bond C, Murchie P. Using technology to deliver cancer follow-up: a systematic review. BMC Cancer 2014 May 03;14(1):311 [FREE Full text] [doi: 10.1186/1471-2407-14-311] [Medline: 24885758]
19. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res 2013 Feb;15(2):e37 [FREE Full text] [doi: 10.2196/jmir.2281] [Medline: 23425685]
20. Alfano CM, Bluethmann SM, Tesauro G, Perna F, Agurs-Collins T, Elena JW, et al. NCI Funding Trends and Priorities in Physical Activity and Energy Balance Research Among Cancer Survivors. J Natl Cancer Inst 2016 Jan;108(1):1-7. [doi: 10.1093/jnci/djv285] [Medline: 26547926]
21. Åkerberg A, Söderlund A, Lindén M. Technologies for physical activity self-monitoring: a study of differences between users and non-users. OAJSM 2017 Feb;Volume 8:17-26. [doi: 10.2147/oajsm.s124542]
22. Timmerman J, Dekker-van Weering M, Stuiver M, Groen W, Wouters M, Tönis TM, et al. Ambulant monitoring and web-accessible home-based exercise program during outpatient follow-up for resected lung cancer survivors: actual use and feasibility in clinical practice. J Cancer Surviv 2017 Dec 10;11(6):720-731 [FREE Full text] [doi: 10.1007/s11764-017-0611-6] [Medline: 28396993]
23. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HWW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017 Dec;11(6):704-719 [FREE Full text] [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
24. Robertson MC, Tsai E, Lyons EJ, Srinivasan S, Swartz MC, Baum ML, et al. Mobile Health Physical Activity Intervention Preferences in Cancer Survivors: A Qualitative Study. JMIR mHealth uHealth 2017 Jan 24;5(1):e3 [FREE Full text] [doi: 10.2196/mhealth.6970] [Medline: 28119278]
25. Puszkiewicz P, Roberts AL, Smith L, Wardle J, Fisher A. Assessment of Cancer Survivors' Experiences of Using a Publicly Available Physical Activity Mobile Application. JMIR Cancer 2016 May 31;2(1):e7 [FREE Full text] [doi: 10.2196/cancer.5380] [Medline: 28410168]
26. Short CE, Finlay A, Sanders I, Maher C. Development and pilot evaluation of a clinic-based mHealth app referral service to support adult cancer survivors increase their participation in physical activity using publicly available mobile apps. BMC Health Serv Res 2018 Dec 16;18(1):27 [FREE Full text] [doi: 10.1186/s12913-017-2818-7] [Medline: 29338722]
27. Nielsen AM, Welch WA, Gavin KL, Cottrell AM, Solk P, Torre EA, et al. Preferences for mHealth physical activity interventions during chemotherapy for breast cancer: a qualitative evaluation. Support Care Cancer 2020 Apr 31;28(4):1919-1928. [doi: 10.1007/s00520-019-05002-w] [Medline: 31367917]

28. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the Effectiveness of Consumer-Focused Health Information Technology Systems Through eHealth Literacy: A Framework for Understanding Users' Needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
29. Kayser L, Rossen S, Karnoe A, Elsworth G, Vibe-Petersen J, Christensen JF, et al. Development of the Multidimensional Readiness and Enablement Index for Health Technology (READHY) Tool to Measure Individuals' Health Technology Readiness: Initial Testing in a Cancer Rehabilitation Setting. J Med Internet Res 2019 Feb 12;21(2):e10377 [FREE Full text] [doi: 10.2196/10377] [Medline: 30747717]
30. Norgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L, et al. The e-health literacy framework: A conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowl Manag E-Learn 2015;7(4):A. [doi: 10.34105/j.kmel.2015.07.035]
31. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A Multidimensional Tool Based on the eHealth Literacy Framework: Development and Initial Validity Testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
32. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013 Jul 16;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
33. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): An outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Education and Counseling 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002]
34. Rossen S, Kayser L, Vibe-Petersen J, Ried-Larsen M, Christensen JF. Technology in exercise-based cancer rehabilitation: a cross-sectional study of receptiveness and readiness for e-Health utilization in Danish cancer rehabilitation. Acta Oncol 2019 May;58(5):610-618. [doi: 10.1080/0284186X.2018.1562213] [Medline: 30698060]
35. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
36. Kvale S, Brinkmann S. Interviews: An Introduction to Qualitative Research Interviewing. Thousand Oaks: Sage; 1996:1-344.
37. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
38. World Health Organization. WHO Framework on integrated people-centred health services. 2016. URL: http://www.who.int/servicedeliverysafety/areas/people-centred-care/en/ [accessed 2016-08-17]
39. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
40. Graneheim UH, Lindgren B, Lundman B. Methodological challenges in qualitative content analysis: A discussion paper. Nurse Educ Today 2017 Sep;56:29-34. [doi: 10.1016/j.nedt.2017.06.002] [Medline: 28651100]
41. DeCuir-Gunby JT, Marshall PL, McCulloch AW. Developing and Using a Codebook for the Analysis of Interview Data: An Example from a Professional Development Research Project. Field Methods 2010 Dec 27;23(2):136-155. [doi: 10.1177/1525822X10388468]
42. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131 [FREE Full text] [Medline: 3920711]
43. Strath SJ, Kaminsky LA, Ainsworth BE, Ekelund U, Freedson PS, Gary RA, American Heart Association Physical Activity Committee of the Council on Lifestyle and Cardiometabolic Health and Cardiovascular, Exercise, Cardiac Rehabilitation and Prevention Committee of the Council on Clinical Cardiology, and Council. Guide to the assessment of physical activity: Clinical and research applications: a scientific statement from the American Heart Association. Circulation 2013 Nov 12;128(20):2259-2279. [doi: 10.1161/01.cir.0000435708.67487.da] [Medline: 24126387]
44. Stamatakis E, Johnson NA, Powell L, Hamer M, Rangul V, Holtermann A. Short and sporadic bouts in the 2018 US physical activity guidelines: is high-intensity incidental physical activity the new HIIT? Br J Sports Med 2019 Feb 20:1-3 [FREE Full text] [doi: 10.1136/bjsports-2018-100397] [Medline: 30786998]
45. Corbett T, Singh K, Payne L, Bradbury K, Foster C, Watson E, et al. Understanding acceptability of and engagement with Web-based interventions aiming to improve quality of life in cancer survivors: A synthesis of current research. Psychooncology 2018 Jan;27(1):22-33. [doi: 10.1002/pon.4566] [Medline: 29044832]
46. Maddison R, Rawstorn JC, Shariful Islam SM, Ball K, Tighe S, Gant N, et al. mHealth Interventions for Exercise and Risk Factor Modification in Cardiovascular Disease. Exerc Sport Sci Rev 2019 Apr;47(2):86-90 [FREE Full text] [doi: 10.1249/JES.0000000000000185] [Medline: 30883472]
47. Dalton SO, Schüz J, Engholm G, Johansen C, Kjaer SK, Steding-Jessen M, et al. Social inequality in incidence of and survival from cancer in a population-based study in Denmark, 1994-2003: Summary of findings. Eur J Cancer 2008 Sep;44(14):2074-2085. [doi: 10.1016/j.ejca.2008.06.018] [Medline: 18674895]

Abbreviations

READHY: Readiness and Enablement Index for Health Technology
HeiQ: Health Education Impact Questionnaire
HLQ: Health Literacy Questionnaire
eHLQ: eHealth Literacy Questionnaire

Edited by G Eysenbach; submitted 05.07.19; peer-reviewed by K Dieperink, S Hoffman Marchese, Z Niu, L Payne; comments to author 21.10.19; revised version received 06.03.20; accepted 03.06.20; published 05.08.20.

Please cite as:
Rossen S, Kayser L, Vibe-Petersen J, Christensen JF, Ried-Larsen M
Cancer Survivors' Receptiveness to Digital Technology–Supported Physical Rehabilitation and the Implications for Design: Qualitative Study
J Med Internet Res 2020;22(8):e15335
URL: https://www.jmir.org/2020/8/e15335
doi:10.2196/15335
PMID:32755892

Original Paper

Translatability of a Wearable Technology Intervention to Increase Adolescent Physical Activity: Mixed Methods Implementation Evaluation

Harriet Koorts[1], PhD; Jo Salmon[1], PhD; Anna Timperio[1], PhD; Kylie Ball[1], PhD; Susie Macfarlane[2], BSc; Samuel K Lai[1], MHP; Helen Brown[1], PhD; Stephanie E Chappel[1], BEx&SportSc (Hons); Marina Lewis[3], PhD; Nicola D Ridgers[1], PhD

[1]School of Exercise and Nutrition Sciences, Institute for Physical Activity and Nutrition, Deakin University, Burwood, Australia

[2]Learning Futures, Deakin University, Geelong, Australia

[3]School of Exercise and Nutrition Sciences, Deakin University, Geelong, Australia

Corresponding Author:
Harriet Koorts, PhD
School of Exercise and Nutrition Sciences
Institute for Physical Activity and Nutrition
Deakin University
221 Burwood Highway
Burwood, VIC 3125
Australia
Phone: 61 392446685
Email: h.koorts@deakin.edu.au

Abstract

Background: Wearable technology interventions combined with digital behavior change resources provide opportunities to increase physical activity in adolescents. The implementation of such interventions in real-world settings is unknown. The Raising Awareness of Physical Activity (RAW-PA) study was a 12-week cluster randomized controlled trial targeting inactive adolescents attending schools in socioeconomically disadvantaged areas of Melbourne, Australia. The aim was to increase moderate- to vigorous-intensity physical activity using (1) a wrist-worn Fitbit Flex and app, (2) weekly challenges, (3) digital behavior change resources, and (4) email or text message alerts.

Objective: This paper presents adolescents' and teachers' perceptions of RAW-PA in relation to program acceptability, feasibility and perceived impact, adolescent engagement and adherence, and the potential for future scale-up.

Methods: A mixed methods evaluation of the RAW-PA study assessed acceptability, engagement, feasibility, adherence, and perceived impact. A total of 9 intervention schools and 144 intervention adolescents were recruited. Only adolescents and teachers (n=17) in the intervention group were included in the analysis. Adolescents completed web-based surveys at baseline and surveys and focus groups postintervention. Teachers participated in interviews postintervention. Facebook data tracked engagement with web-based resources. Descriptive statistics were reported by sex. Qualitative data were analyzed thematically.

Results: Survey data were collected from 142 adolescents at baseline (mean age 13.7 years, SD 0.4 years; 51% males) and 132 adolescents postintervention. A total of 15 focus groups (n=124) and 9 interviews (n=17) were conducted. RAW-PA had good acceptability among adolescents and teachers. Adolescents perceived the intervention content as easy to understand (100/120, 83.3%) and the Fitbit easy to use (112/120; 93.3%). Half of the adolescents perceived the text messages to be useful (61/120; 50.8%), whereas 47.5% (57/120) liked the weekly challenges and 38.3% (46/120) liked the Facebook videos. Facebook engagement declined over time; only 18.6% (22/118) of adolescents self-reported wearing the Fitbit Flex daily postintervention. Adolescents perceived the Fitbit Flex to increase their physical activity motivation (85/120, 70.8%) and awareness (93/119, 78.2%). The web-based delivery facilitated implementation of the intervention, although school-level policies restricting phone use were perceived as potential inhibitors to program roll-out.

Conclusions: RAW-PA showed good acceptability among adolescents attending schools in socioeconomically disadvantaged areas and their teachers. Low levels of teacher burden enhanced their perceptions concerning the feasibility of intervention delivery. Although adolescents perceived that RAW-PA had short-term positive effects on their motivation to be physically active, adolescent adherence and engagement were low. Future research exploring the feasibility of different strategies to engage

adolescents with wearable technology interventions and ways of maximizing system-level embeddedness of interventions in practice would greatly advance the field.

(J Med Internet Res 2020;22(8):e13573) doi:10.2196/13573

KEYWORDS

wearable technology; social media; implementation science; adolescent; physical activity; awareness

Introduction

Physical activity is an important component of a healthy lifestyle, yet physical inactivity is a global pandemic with far-reaching consequences for health and well-being both now and in the future [1]. The benefits of physical activity in childhood and adolescence include reduced cardio-metabolic risk factors, improved body composition, and higher fitness [2]. However, global estimates suggest that over 80% of adolescents do not engage in the recommended 60 min of moderate- to vigorous-intensity physical activity (MVPA) every day [3], and steep declines in physical activity levels during adolescence are common [4]. This is particularly evident for adolescents living in areas of socioeconomic disadvantage, who are less likely to meet physical activity guidelines and are at greater risk of declining activity levels [5]. There is clearly a need to identify strategies for maintaining - if not increasing - adolescent physical activity levels. However, compared with interventions targeting primary school children, few have been conducted, particularly with those living in socioeconomically disadvantaged areas [6].

One approach that may have the potential to promote physical activity levels in adolescents is the use of wearable activity trackers, which are electronic devices that are designed to be worn on the body that use sensors (eg, accelerometers) to track movement and/or biometric data [7]. These technologies enable constant self-monitoring of physical activity through the provision of data and feedback via a visual display and/or an accompanying app [8]. There is a dearth of data about wearable activity tracker ownership in adolescents, although one study found that approximately 25% of adolescents owned such devices, with more males than females reporting ownership [9]. Of note, school-based research has shown that wearable activity monitors have moderate acceptability among school-aged children in low-income communities [10]. Furthermore, there is some initial evidence that adolescents are generally positive about the use of wearable technology for tracking physical activity, and use device features to set goals and undertake challenges against friends [11]. From a school's perspective, wearable technology interventions have the potential to be highly implementable, given the low burden placed on teachers to deliver and potential minimal interference with existing school practices [10].

Wearable technology has recently been combined with social media as a platform to disseminate health-promoting messages and effect behavior change. Evidence from adults in a clinical setting has shown that wearable technology combined with a social media–based health education intervention increased daily light-intensity physical activity and MVPA [12]. Adolescents are known to be high users of social media [13], and those from low-income families report higher social media use than those from high-income homes [14]. The potential reach of such combined interventions among youth living in socioeconomically disadvantaged areas may comprise an opportunity to increase activity levels among these groups. Although it is known that adolescent males and females differ in their preferences for types of physical activities [15], less is known about how youth respond to technologies promoting physical activity, such as wearable activity trackers and digital behavior change resources, and if different social media strategies are more effective for engaging males or females. Physical activity apps, including web-based platforms, have promising reach and a low burden; however, there is limited evidence of their efficacy in adolescent populations [16].

To shift population-level physical activity, not only must interventions demonstrate effectiveness, but they must be sustainably implemented over time and under real-world conditions [17]. Physical activity interventions that are designed with real-world implementation and scale-up in mind are recommended [17]. However, most wearable device physical activity intervention studies, regardless of age group, are delivered in small samples (eg, <100 participants), and there is a dearth of studies in youth populations [18,19]. In a number of cases, particularly among clinical populations, participants are prescribed how to use the devices (eg, via a counseling study approach) [20,21], and implementation of wearable technology interventions in nonclinical settings and among nonclinical populations is, therefore, less well understood. “Furthermore, the majority of school-based studies investigating the impact and feasibility of wearable activity trackers on physical activity in youth have not been tested under real-world conditions and over longer periods (eg, >3 months), which may subsequently impede implementation and effectiveness [22].

This study aimed to assess the implementation of a wearable technology intervention to increase physical activity among adolescents: The Raising Awareness of Physical Activity (RAW-PA) study.

The specific aims of this study were to evaluate adolescent (individual level) and teacher (individual and school level) perceptions of intervention acceptability, feasibility and perceived impact, and adolescent engagement in and adherence to the intervention. Findings from this evaluation will provide important first evidence for the feasible real-world implementation of wearable technology interventions among inactive adolescents and provide recommendations to potentially enhance future implementation of such interventions if delivered on a larger scale.

Methods

Overview of Raising Awareness of Physical Activity

A detailed description of the program and study protocol has been published elsewhere (ANZCTR: ACTRN12616000899448) [22]. In brief, RAW-PA was a 12-week cluster randomized controlled trial conducted in 2016-2018, which targeted adolescents (year 8, ie, second year of secondary school) attending schools in socioeconomically disadvantaged areas of Melbourne, Australia. After the 12-week intervention period, the intervention ceased as intended. RAW-PA combined both wearable technology and digital behavior change resources accessible via social media. These types of combined interventions can be known as *digital behavior change interventions*; however, given that this term can include a broad range of intervention types, for the purposes of this paper, we refer to RAW-PA as a *wearable technology intervention*. RAW-PA was co-designed (eg, style and frequency of delivery) with the target users (adolescents) [22], and it incorporated low-cost strategies to facilitate real-world implementation and the potential for wider scale-up [22]. Based on the social cognitive theory [23] and behavioral choice theory [24], the intervention promoted awareness of physical activity via a wearable physical activity tracker and accompanying app, and focused on increasing activity levels using digital behavior change resources.

RAW-PA aimed to increase adolescent MVPA by targeting the accumulation of activity across the day, which included out-of-school hours and weekends. Core components of the intervention included: (1) a wrist-worn Fitbit Flex and accompanying Fitbit app; (2) interactive weekly individual and/or team *missions* or *challenges* (delivered via email and/or text message approximately 2-3 times/week); (3) digital behavior change resources (eg, motivational videos and social forums accessible via a private Facebook group); and (4) email and/or text message alerts to new content, missions, or challenges (approximately 2-3 times/week). One of the weekly challenges (*Mark it Up!*) focused on increasing awareness of activity opportunities in schools where students and teachers identified and shared strategies for increasing activity at school and competed in a step challenge against each other. To facilitate this challenge, 2 teachers from each intervention school were provided with a Fitbit Flex. The remaining challenges targeted behavior change outside of school, and content was delivered during out-of-school hours [22]. The trial adhered to the consolidated standards of reporting trials guidelines, and ethical approval was obtained from the Deakin University human research ethics committee (2016-179) and the Victorian Department of Education and Training. Participating schools provided written informed consent, and parents provided signed consent, which included student assent.

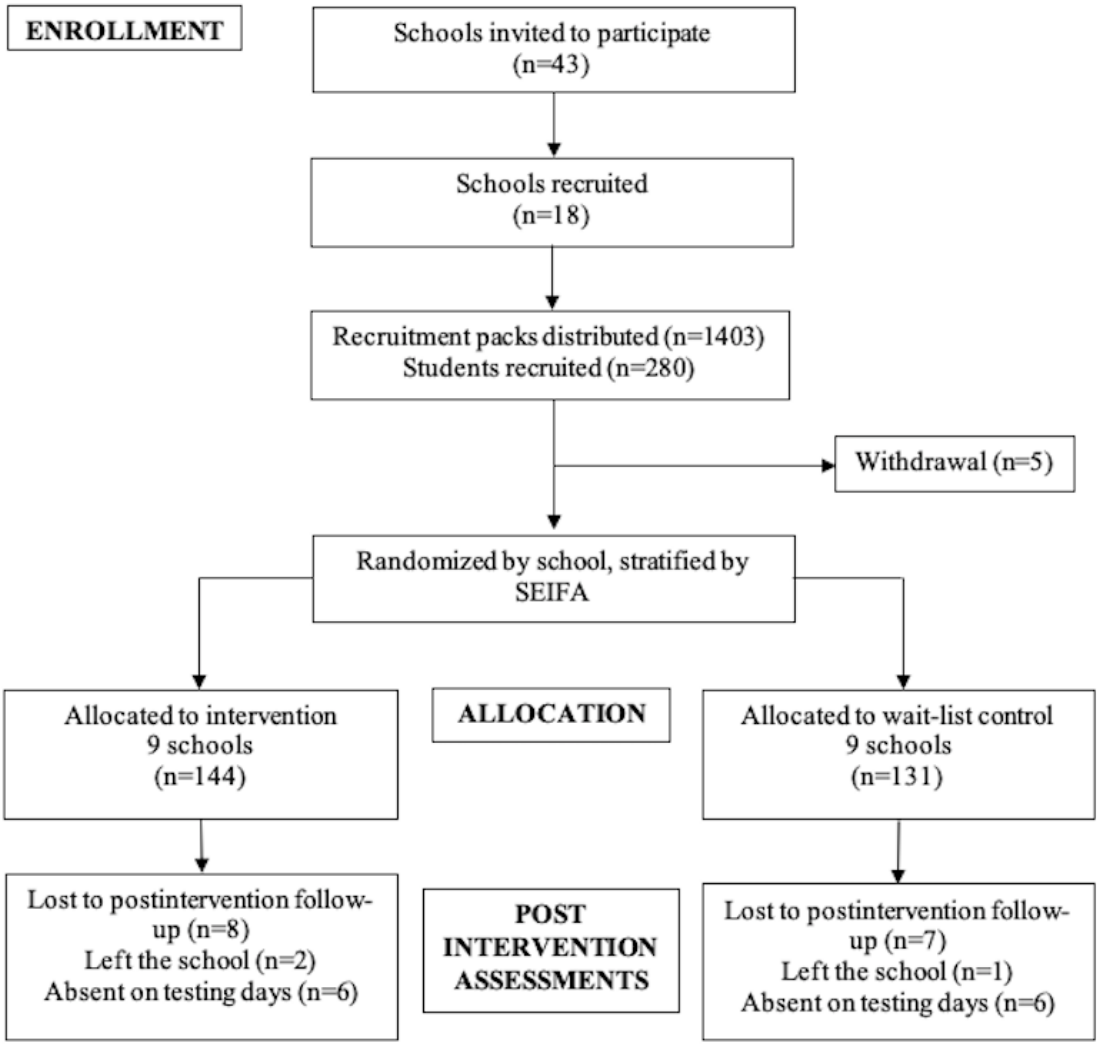
Evaluation Design

This study used a mixed methods evaluation design based on the UK Medical Research Council recommendations [25]. A total of 5 evaluation indicators were identified based on recommended outcomes for process evaluation and implementation-related research [25,26]: (1) *acceptability* (eg, adolescent enjoyment, ease of understanding and Fitbit use, and teacher-perceived barriers to uptake), (2) *engagement* (eg, frequency of adolescent interaction with Facebook group/posts), (3) *feasibility* (eg, adolescent barriers to Fitbit wear and teacher-perceived suitability of delivery in the school setting), (4) *adherence* (eg, adolescent self-reported completion of weekly challenges), and (5) *perceived impact* (eg, perceived changes in motivation, awareness, and encouragement for physical activity).

Participants and Recruitment

Adolescents were recruited through the school setting for ease of recruitment. In total, 18 schools (intervention: n=9 and wait-list control: n=9) and 280 students were recruited. A total of 5 participants withdrew before baseline data collection; therefore, 275 students (intervention=144 and wait-list control=131) took part in the study (Figure 1). Schools were eligible to participate in the study if they were located in areas that had a score of ≤5 (lowest 50%) on the socioeconomic indexes for areas (SEIFA [27]) within 60 km of Deakin University's Burwood Campus. Eligible schools were randomly selected to receive an invitation to participate in the study. The year eight coordinators were invited to be the school liaison and help with data collection. Year coordinators, as opposed to classroom teachers, were recruited as they typically had contact with all students in the year group. Schools that provided written informed consent from their school principal to participate in the study were matched based on SEIFA score and size and randomly assigned to either the intervention or wait-list control group by a computer-based random number generator [22]. All participating schools were located in urban areas. Recruitment and baseline data collection were conducted before school randomization.

Figure 1. Flow of participants. SEIFA: socioeconomic index for areas.



For eligibility, adolescents were required to (1) be at least 13 years old (minimum age required to have a Fitbit and a Facebook account); (2) not engage in regular organized physical activity sports outside of school; (3) not meet national physical activity guidelines of at least 60 min of MVPA daily [28]; (4) not be a current or past owner of an activity tracker; (5) have, or be willing to create, a Facebook account; and (6) have access to the internet outside of school (eg a mobile device with data or Wi-Fi at home). A self-report checklist containing these eligibility criteria was included with each parental consent form, which was completed by parents and students. Each student who returned a completed parental consent form (which included student assent) and met the eligibility criteria was recruited into the study. Participants in the wait-list control group were provided access to the intervention materials on completion of the 6-month follow-up assessments, but no implementation/process data were collected.

Procedure

All adolescents participating in the study were invited to complete a web-based survey at baseline and postintervention (12 weeks after baseline). Adolescents completed the web-based survey using iPads within school hours. Hard copies of the survey were provided to schools for those adolescents absent on the day of data collection at baseline (n=13) and postintervention (n=29) and were completed in the presence of teachers in the intervention schools. Due to modem failure, 23 adolescents completed hard copies of the survey at postintervention. Adolescents attending schools randomized to the intervention group were additionally invited to participate in a focus group (ranging between 4 and 13 students per group) at postintervention to explore their perspectives on RAW-PA [22]. Teachers from intervention schools were also invited to participate in an interview postintervention to provide insights about the approach from an organizational perspective. All interviews and focus groups were conducted on-site during school hours. Discussions followed a semistructured format and were audio recorded. The 15 focus groups (mean duration 26 min, SD 5 min) and 9 interviews (mean duration 21, min SD 7 min) were then transcribed verbatim for further analysis.

Measures

Table 1 summarizes the relevant data for the 5 implementation evaluation indicators included in this study, collected postintervention. Baseline adolescent surveys captured participant reach and sociodemographic variables. A complete

and detailed list of measures has been previously published [22].

Quantitative survey data were used to assess adolescent-perceived acceptability (11 questions), feasibility (3 questions), adherence (3 questions), and perceived impact (6 questions) of RAW-PA. Items were rated on a 5-point Likert scale from strongly disagree (1) to strongly agree (5), with 1 question requiring reverse coding (*Please indicate the extent to which you agree with the following statements: Wearing the Fitbit was uncomfortable and I felt embarrassed wearing the Fitbit*). Qualitative free-text survey responses contributed to assessing acceptability (3 questions: what adolescents liked most and least, and suggested improvements) and feasibility (1 question: reasons for suggesting a different intervention duration; Table 1) and adherence (1 question: reasons for Fitbit nonwear). These items were included in the intervention group surveys only.

Qualitative focus group and interview data provided information about adolescent- and teacher-perceived program acceptability, feasibility, adherence, and impact. Exemplar questions are shown in Table 1.

Web-based Facebook data of *views* and *likes* were recorded weekly during the 12-week intervention to assess the frequency of adolescent engagement with the Facebook group. As the research team posted comments within the Facebook group in addition to participants, only *views* (ie, posts clicked on and viewed) and *likes* (ie, participant *liked* a post) of the web-based program material were used to assess engagement.

Table 1. Implementation evaluation indicators and assessment criteria postintervention.

Evaluation indicator and assessment criteria	Data source	Exemplar question/reporting criteria
Acceptability		
Adolescent-perceived intervention enjoyment, ease of understanding/use, and comfort wearing a Fitbit	Web-based survey (14 questions)	"The information was easy to understand." "Wearing the Fitbit was uncomfortable."
Adolescent levels of intervention enjoyment and challenges faced	Focus group (6 questions)	"Did you experience any issues when using/accessing features of the program?"
Teacher-perceived barriers and facilitators to intervention uptake and perceptions of acceptability	Interview (5 questions)	"How acceptable would such a program be to schools?"
Engagement		
Frequency of *views* and *likes* of intervention strategy Facebook posts[a]	Facebook data	Total number of *views* and changes over time (ie, the post was clicked on and viewed) and *likes* of each Facebook postadolescents received
Feasibility		
Adolescent-perceived appropriateness of the intervention duration	Web-based survey (4 questions)	"The length of the RAW-PA[b] program was just right."
Barriers and facilitators to accessing the program	Focus group (3 questions)	"Did anything help you to use/access any feature(s) of the program?"
Teacher-perceived appropriateness of intervention delivery in the school setting	Interview (2 questions)	"What considerations would schools make before participating in such a school-based challenge?"
Adherence		
Adolescent self-reported adherence to wearing a Fitbit and completion of weekly challenges	Web-based survey (4 questions)	"On how many days did you wear the Fitbit in the last week?"
Barriers and facilitators to wearing the Fitbit and adhering to the program	Focus group (1 question)	"Did anything stop you using/accessing any feature(s) of the program?"
Perceived impact		
Adolescent-perceived impact of intervention on motivation, awareness, and encouragement for physical activity	Web-based survey (6 questions)	"The Fitbit motivated me to be more active." "The Fitbit made me think about how much activity I do."
Adolescent-perceived impact of intervention and change in awareness regarding physical activity	Focus group (2 questions)	"Did the program change your awareness of your activity levels?"
Teacher-perceived impact of the intervention on students and school and impact on teacher awareness of own physical activity	Interviews (3 questions)	"Did the program change your awareness of your own physical activity levels?"

[a]Web-based Facebook engagement was captured weekly during the 12-week intervention period.

[b]RAW-PA: Raising Awareness of Physical Activity Study.

Analyses

Descriptive statistics from adolescent survey data were calculated for 4 of the 5 evaluation indicators by sex (acceptability, adherence, feasibility, and perceived impact). Survey data were combined into 3 groups, classified as *agree* (sum of responses *strongly agree* and *agree*), *neither* (sum of responses *neither agree* nor *disagree*) and *disagree* (sum of responses *strongly disagree* and *disagree*). This approach is appropriate for analyzing ordinal data and provides insights into the adolescents' perspectives about the intervention [29]. Continuous sample characteristics are presented as means and SDs, and categorical data are presented as counts and percentages. The total *potential* number of weekly views/likes possible on Facebook was calculated based on the number of participants registered for the Facebook group multiplied by the number of posts provided to participants each week. This score was based on the assumption that every registered participant had the opportunity to view/like each post on at least one occasion per week. Restrictions on the information available for download via Facebook meant that Facebook data could not be stratified by sex. Sex differences in adolescent perceptions of intervention acceptability, adherence, feasibility, and perceived impact were calculated using a Mann-Whitney test. All analyses were conducted using Stata SE 15 (StataCorp LP).

Qualitative free-text survey responses were coded thematically. Transcribed qualitative data were imported into NVivo 12 (QRS International) for coding and thematic analysis. Thematic analysis requires initial data familiarization, coding, and tabulation of raw themes, which are then grouped based on patterns of emergence and overlapping relevance [30]. Coding and theme development were first deductive (theory driven), guided by the study aims, process evaluation framework [25], and project team's previous research and conceptualization [11], followed by an inductive approach (data driven) directed by the content of the data [31]. Specifically, the process evaluation provided a metathematic structure to guide initial coding, and subthemes relating to the adolescents' and teachers' perceptions and experiences were then identified. Consistent with the recommended approaches [30], 2 researchers independent of the project team (ML and SC) were engaged to analyze the qualitative data. Selections of raw data were independently coded by the lead author (HK) by means of cooperative triangulation. Presented themes were critically questioned and interpretations of these data were challenged. Instances of divergence were discussed until consensus was reached. Illustrative quotes were extracted from the coded data to reflect and support the themes identified from these data.

Results

Overview

In the 9 intervention schools, 144 and 136 adolescents (Figure 1) were invited to complete the surveys at baseline and postintervention, respectively. Of the 8 students not invited at postintervention, 2 had left the school and 6 were absent. A total of 142 (99%) adolescents completed baseline surveys (mean age 13.7 years, SD 0.4 years; 51% males), and 132 (98%) adolescents completed surveys postintervention (mean age 14.0 years, SD 0.4 years; 52% males). Approximately 85% (n=122) of adolescents (52% males) registered for the Facebook group. The 15 focus groups conducted comprised 124 (86%) adolescents (51% males), and the 9 interviews involved 17 (81%) teachers (47% males).

Acceptability

Table 2 presents the percentage of adolescents as well as the proportion of males and females who stated *agree* or *strongly agree* to questions relating to acceptability, feasibility, adherence, and perceived impact. The majority of adolescents reported that the program was easy to understand, enjoyable, and they would recommend it to their friends. Half of all adolescents thought the text messages were useful, and less than half liked the weekly challenges and Facebook videos. Males were significantly more likely than females to agree that they liked the Facebook pages and videos. With regard to the acceptability of the Fitbit Flex, almost all adolescents perceived that the Fitbit Flex was easy to use, and the majority agreed that they got used to wearing it and they did not feel embarrassed.

Table 2. Descriptive statistics of evaluation indicators postintervention.

Implementation evaluation indicators[a]	Overall n (%)	Male n (%)	Female n (%)	*P* value
Acceptability				
Fitbit				
The Fitbit was easy to use (N[b]=120)	112 (93.3)	57 (91.9)	55 (94.8)	.54
I got used to wearing the Fitbit (N[b]=119)	85 (71.4)	42 (68.9)	43 (74.1)	.49
The Fitbit was comfortable to wear (N[b]=120)	70 (58.3)	37 (59.7)	33 (56.9)	.94
I was not embarrassed wearing the Fitbit (N[b]=120)	94 (78.3)	48 (77.4)	46 (79.3)	.69
RAW-PA[c] program				
The text messages were useful (N[b]=120)	61 (50.8)	34 (54.8)	27 (46.6)	.50
I liked the Facebook page *(N[b]=120)*	*71 (59.2)*	*44 (71.0)*	*27 (46.6)*	*.03*
Information was easy to understand (N[b]=120)	100 (83.3)	52 (83.9)	48 (82.8)	.94
I liked the weekly challenges/missions (N[b]=120)	57 (47.5)	33 (53.2)	24 (41.4)	.22
I liked the videos *(N[b]=120)*	*46 (38.3)*	*29 (46.8)*	*17 (29.3)*	*.03*
I enjoyed the program (N[b]=120)	89 (74.2)	48 (77.4)	41 (70.7)	.32
I would recommend the program to friends (N[b]=120)	85 (70.8)	48 (77.4)	37 (63.8)	.07
Feasibility				
The program length was appropriate (N[b]=118)	74 (62.7)	36 (60.0)	38 (65.5)	.65
Adherence				
I completed the weekly challenges/missions (N[b]=119)	43 (36.1)	26 (42.6)	17 (29.3)	.13
Perceived impact				
Fitbit				
Motivated me to be more active (N=120)	85 (70.8)	45 (72.6)	40 (69.0)	.68
Made me think about how much activity I do (N[b]=119)	93 (78.2)	48 (77.4)	45 (79.0)	.80
RAW-PA program				
Challenges motivated me to be more active (N[b]=119)	41 (34.5)	24 (38.7)	17 (29.8)	.66
Encouraged increased activity on own (N[b]=120)	74 (61.7)	40 (64.5)	34 (58.6)	.42
Encouraged increased activity with family (*N[b]=119*)	*48 (40.3)*	*31 (50.0)*	*17 (29.8)*	*.05*
Encouraged increased activity with friends (N[b]=120)	65 (54.2)	38 (61.3)	27 (46.6)	.16

[a]Data (%) reported is a combined score relating to those who stated *agree* and *strongly agree*. *P* value for sex differences significant at less than or equal to .05 (italics); calculated by Mann-Whitney tests.

[b]The N values differ due to questions not being completed in the surveys.

[c]RAW-PA: Raising Awareness of Physical Activity.

Responses to open-ended survey questions revealed that the aspects of the program participants liked most were receiving a free Fitbit (n=40) and the increased motivation to be active (n=28). The least preferred aspects included the Fitbit Flex design (eg, discomfort wearing and frequent need to charge the device, n=21) and the weekly challenges being too hard or demotivating (n=10). The frequency and volume of program notifications were also perceived negatively by a small number of participants (n=7), including receipt of text messages during class time.

During focus groups, a number of subthemes emerged as influencing perceptions of acceptability. Adolescents referred positively to specific features of the Fitbit Flex, such as the monitoring of sleep and physical activity:

> *I like how it records your steps because then at lunch you look at it and you can slowly start to improve your steps and then yep, it works out, it evolves and yeah, it's good. I really liked that.* [adolescent, school A]

Adolescents also described the visual feedback on physical activity performance from the Fitbit Flex and Fitbit app as motivating:

> *Sometimes at the end of the day I want to get two dots and then next day I tried to like, get five dots.* [adolescent, school J]

> *And it [Fitbit app] also said...it also said like, which day you are the most active and which day is the worst so you can just like try to be more active like, every day.* [adolescent, school A]

Adolescents also referred positively to the goal-setting component of RAW-PA via the Fitbit Flex and Facebook posts. The focus groups highlighted that the social aspects of the intervention, such as challenging and competing with peers, influenced their goal setting:

> *It shows like, how many steps I've taken in a day and it also encouraged me since like, there's challenges where I can do it with my friends and try to beat them.* [adolescent, school M]

> *I like [RAW-PA] challenges to be honest. When I first got it [the challenge] I tried to be active.* [adolescent, school J]

Less favorable aspects of the program were those associated with the Fitbit Flex and app. This included technical difficulties connecting to the app and inaccuracies of device monitoring and feedback:

> *When you shake your hand a lot because—and I was playing drums and then it was really hard because every second it would vibrate. It was annoying so I had to take it off and playing the drums counted as a step.* [adolescent, school A]

A small proportion of adolescents referred to a lack of awareness of the Facebook notifications due to the amount of competing web-based information. This included negative perceptions about the number of notifications sent:

> *Well, I didn't like how many notifications it sent, because if I want to check my - your email for something, all the notifications would come up instead.* [adolescent, school O]

Interviews with teachers revealed that RAW-PA was perceived as highly acceptable at the school level. Perceived enjoyment of the program by students was central to teachers' ongoing support, as well as the small amount of teacher time and investment required to support program delivery. This ease of use increased the acceptability and support for RAW-PA in schools:

> *But once it was set up with the small group that did end up participating, then it was minimal amount of my time, which was fantastic...So that made it easy for me to have it running in the school, and easy to support it.* [teacher, school K]

One teacher referred to the personal benefits of increased awareness of their physical activity as a result of the Fitbit challenges:

> [For] *me personally, it probably got me off my bum a little bit. I enjoyed the challenge with the kids, and I still wear it daily, and I've been monitoring, that I am getting my minimum 10,000. I'm getting beyond that each day, but I'm more conscious of that. So from a personal level, it was good for me.* [teacher, school N]

The teachers perceived that role modeling physical activity was positively associated with student engagement:

> *It motivated the students a little bit more. I did see that there was a little bit more engagement with students with the teachers involved.* [teacher, school I]

> *It was really good to actually get the teachers involved into the program, because the students are also affected by what we do, and if we sort of role model it to them they're more likely to get engaged and involved in the activity as well.* [teacher, school K]

Teachers acknowledged the advantages of their reduced involvement in the implementation of RAW-PA in minimizing any potential increase in their existing workloads. Limited teacher involvement was also perceived to increase adolescent autonomy and responsibility:

> *They had to actually commit to something, which they'd never been asked to before...it gave them a bit of responsibility that they've not had before.* [teacher, school N]

Nonetheless, teachers equally acknowledged that their lack of involvement in the Facebook group meant they had no awareness of what the program was promoting via Facebook and what interactions were taking place. The 4 teachers described that the lack of feedback impacted both their own participation and their knowledge of what was happening among the students:

> *So it [the program] didn't maybe like, get us, like, had a major impact on me...But, yeah, if we had like, kids talking to us, or we at least like, knowing who won, and yeah, but no one really talked to us, and we're like OK.* [teacher, school M]

Although RAW-PA was not designed as a school-based physical activity intervention, teachers consistently described wanting greater information regarding the content of the program material and awareness of their students' involvement and performance:

> *I wish that we had a bit more like information of how the kids were going...like the steps, about our challenge, so like, if we could see that, let's say, I don't know, like a newsletter...* [teacher, school M]

> *So maybe some more teacher involvement just to, sort of checking in and getting feedback, and being a little bit more active in the kids who are in the programme is probably needed I'd say.* [teacher, school K]

If implemented at scale, challenges highlighted by teachers included the need for increased responsibility at the school level, including how the program would be funded. Sustainable

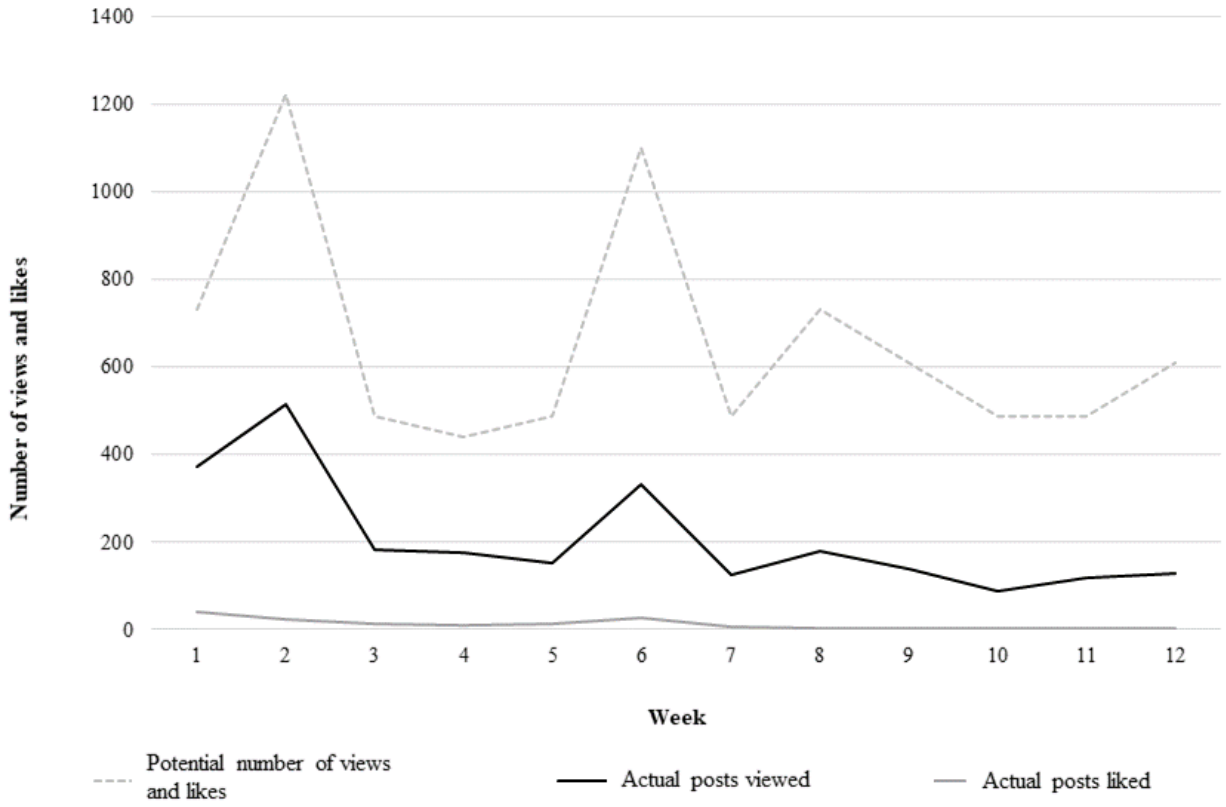
implementation was suggested to require stronger links with the community to achieve broader program acceptability:

> *Schools are very big on community involvement, so whether it's linked to the school, I don't know if you can use the school Facebook page as well and give updates through that.* [teacher, school I]

Engagement

Engagement in RAW-PA was based on the frequency that the Facebook posts were *liked* and/or *viewed* by participants registered for the Facebook group. In general, more posts were viewed than liked. Engagement in the Facebook group declined over time (Figure 2). In week 1, there was the potential for at least 732 views and/or likes of 6 different Facebook posts that were made (Figure 2); however, adolescents viewed and liked the posts only 370 and 40 times in week 1, respectively.

Figure 2. Number of Facebook posts viewed and liked by adolescents. The potential number of views and likes was based on the total number of participants registered for the Facebook group (n=122) multiplied by the number of posts provided on Facebook each week during the intervention. Actual posts liked and viewed tracked by Facebook.



Feasibility

Irrespective of sex, the majority of adolescents agreed that the 12-week intervention duration was appropriate (Table 2). A total of 23 (20%) adolescents did not think the 12-week duration was an appropriate length, and of these, the majority preferred a 6-week program (Table 3). A preference for intervention duration did not differ between males and females. The most frequent reason for preferring a shorter duration was to retain participant motivation in the program and subsequent Fitbit use (n=9).

Table 3. Preferred intervention duration for adolescents who did not agree with the 12-week duration (N=23).

Preferred intervention duration	Overall (n=23), n (%)	Males (n=9), n (%)	Females (n=14), n (%)
4 weeks	2 (8)	0 (0)	2 (14)
6 weeks	7 (30)	5 (55)	2 (14)
8 weeks	5 (21)	3 (33)	2 (14)
10 weeks	5 (21)	1 (11)	4 (28)
16 weeks	2 (8)	0 (0)	2 (14)
20 weeks	2 (8)	0 (0)	2 (14)

During the focus groups, however, adolescents most frequently stated that the 12-week duration was too long and that they would prefer a program duration of 5 to 10 weeks. This was often justified based on the declining motivation to participate over time:

> *I feel like the program was a bit long. I feel like a lot of people started to get less motivated around week ten than week eight.* [adolescent, school N]

A number of subthemes emerged as influencing perceptions of intervention feasibility. These were characteristics of the intervention design and factors within the home and school environment. In relation to intervention design, program reminders via Facebook and the Fitbit app meant RAW-PA was freely accessible by adolescents in their leisure-time outside of school hours. One adolescent referred to family involvement as a key facilitator of participation:

> *My brother would take me training...he said oh he'd really like to, so he started taking me out nearly every single day playing basketball, just doing drills on the oval and stuff.* [adolescent, school J]

The potential feasibility of RAW-PA was hindered by technical issues such as Fitbit Flex syncing and charging, challenges accessing the app, and the requirement for internet access. Some adolescents reported having limited access to Wi-Fi and limited mobile phone data plans, which made accessing intervention content challenging. Adolescents described these as potentially having a negative impact on their sustained engagement and program impact:

> *Oh, sometimes when I had like, five dots on my Fitbit, at around nine o'clock at night it would just reset back to zero steps, and that was quite annoying.* [adolescent, school O]

> *My Fitbit wouldn't connect to the app so when I just like went on the app it wouldn't tell me how many steps I took and stuff.* [adolescent, school I]

> *I have a limited amount of WiFi.* [adolescent, school F]

Despite technical difficulties experienced with the technology, teachers described this *hands off* intervention design as a major facilitator of implementation and program feasibility:

> *So if we've got the freedom to manipulate the program around what we're doing, it's a whole lot easier than us having to stop everything to fit the program in. That would have been almost impossible, and I think it would have fallen over.* [teacher, school N]

Nonetheless, there was mixed support for the participants' use of phones during the school day. During the intervention period, teachers allowed students to access social media (ie, to receive RAW-PA notifications); however, most teachers indicated that the use of phones was not permitted during class time. Although phone use during recess and lunchtime was mostly unrestricted, one teacher described their school attempting to extend restrictions on screen use to include recess:

> *Well at recess we're trying to limit the screen time on iPads, and so we're not encouraging them to use it [during recess] either.* [teacher, school M]

One teacher described the potential negative implications for future uptake and implementation of the program as a result:

> *I think that if you came to a school and said, "We'd be requiring the kids to be checking on their Facebook during the school day," you'd find the school a lot less enthusiastic about it.* [teacher, school J]

To increase potential feasible implementation, several teachers suggested that RAW-PA could be introduced during a dedicated physical education lesson or integrated as part of the existing health and physical education (HPE) curriculum.

Adherence

Adherence to RAW-PA was evaluated based on self-reported adherence to wearing the Fitbit Flex and the completion of weekly challenges, including self-reported barriers and facilitators. In the surveys, 18.6% (22/118) of adolescents reported wearing the Fitbit Flex daily in the last week of the program and 35.5% (42/118) reported not wearing it at all. A total 22.4% (13/58) of females and 15.0% (9/60) of males wore the Fitbit on all 7 days. There were no significant sex differences in Fitbit wear. The main reasons for failing to wear the Fitbit included forgetting to wear it (50/97, 51.5%), forgetting to charge it (39/97, 40.2%), and having a flat battery (36/97, 37.1%). The most common *other* reason for nonwear was losing the charger (n=6). Approximately one-third of adolescents reported completing the challenges during the 12 weeks (Table 2).

Consistent with survey data, the most commonly reported barrier to Fitbit wear in the focus groups was the loss of the device and/or charger. For example, an adolescent referred to a subsequent lack of interest in continuing to use the Fitbit after a period of not wearing it:

> *I wore it [the Fitbit] for like, the first nine weeks and then I lost it. Then I found it again then I wasn't interested in wearing it again so I just didn't wear it.* [adolescent, school O]

One of the main reasons for declining interest in adhering to the program over time was adolescents' short-term motivation to achieve daily step goals:

> *I started off with 10,000 and then it was like, every day if I reach my goal, I'll increase it by 1,000...But that was only like, during the first week or two and then eventually like, I just don't care for it anymore. It's just sort of, the interesting bit's worn off.* [adolescent, school N]

Perceived Impact

Surveys showed that the Fitbit Flex was perceived to have had a greater impact on the adolescents' physical activity than other RAW-PA components (Table 2). Consistent for males and females, the majority of adolescents agreed that wearing the Fitbit Flex motivated them to be more active and increased their awareness of their own physical activity levels. The program encouraged the majority of adolescents to be more active on

their own; however, only around one-third agreed that the weekly challenges motivated them to be more active. Males were significantly more likely than females to report that the program encouraged them to be more active with their families. During the focus groups, adolescents spoke positively regarding the perceived impact of RAW-PA and understood the aims of the intervention. Increased motivation to be active due to competing with peers was often described as having an impact on their physical activity:

> *I went shopping with my family, which I never do. We were going furniture shopping. I was like nope, got to get my steps up...I was walking around [name of shop] over and over and over, just to get the steps up, because I wanted to win.* [adolescent, school N]

However, for some adolescents, the competition aspect was in fact demotivating and had a negative impact:

> *At the start, I was like, okay I'm going to beat everyone...I wake up the next morning...and everyone's already got like a million steps...how am I meant to overtake everyone? So I was like, well, there's no hope.* [adolescent, school N]

When asked if they thought the program was successful, the majority of adolescents indicated that they did. Overall, adolescents reported a greater awareness of their physical activity levels as a result of RAW-PA:

> *I kind of noticed how much like, I actually moved around during the school day, because I'm in a lot of classes in lunch time and recess that I actually have to move around to cross the school.* [adolescent, school O]

Some adolescents reported that RAW-PA made them want to be more active, leading to potential increases in their physical activity levels:

> *It like, made you want to do more, like physical things.* [adolescent, school F]

> *I don't really like to exercise, because I don't think I'm very good at it. But since I got the Fitbit I'm actually bothering to exercise now...* [adolescent, school O]

However, there was some evidence that any changes in behavior were unlikely to be sustained:

> *The start when we got the Fitbit, the app round the start I had, I was very inactive. I slowly started getting more steps but then I started falling back down.* [adolescent, school J]

> *I don't think it affected me that much. I mean, it - it increased it a little bit, but most of the time it just became normal and I just continued whatever my normal week would be.* [adolescent, school O]

Consistent with adolescents' feedback, teachers agreed that the intervention was positive; they perceived it had raised adolescents' awareness of physical activity and was therefore likely to be of benefit. However, the sustainability of any changes leading to long-term impact was questioned:

> *Yeah, two girls who normally weren't very active had decided that they would start to do a gym program...It was a short term thing, but they did actually start to work together and try and push each other a little bit. So for those two girls, that's extraordinary because they normally don't do much activity.* [teacher, school N]

At the individual teacher level, 2 teachers reported increased awareness of their own physical activity and a subsequent change in their physical activity behavior as a result of participating:

> *The teacher challenge bit, you know, had an impact on me. Like, I'm still wearing it, and I'm still using it, and it's not something I've done before.* [teacher, school J]

> *Well the impacts on myself and I could say my colleague as well were just that we were really aware of our activity to make sure that we were quite active.* [teacher, school A]

Discussion

Principal Findings

The RAW-PA program had good acceptability among inactive adolescents and was highly acceptable at the school level. Based on the qualitative and quantitative data, text messaging and weekly Facebook challenge components of the program were less popular, although the social aspects of peer challenges and competition were popular with some but not others. Technical difficulties with the Fitbit Flex and app hindered the adolescents' experiences of the program, and adherence and engagement were generally low as a result of device loss and declining use of Facebook over time. Adolescents perceived that the Fitbit Flex increased their physical activity awareness and likelihood of being more active on their own in the short term, although competition with peers was associated with both increased and decreased motivation to be active. The web-based program delivery was perceived by teachers as a key facilitator of implementation and program feasibility in schools, although teachers wanted greater access to and awareness of the program content and student involvement. Teachers considered the program highly acceptable and feasible for the school setting; however, they highlighted that any future implementation may be significantly limited by school policies restricting mobile phone use within schools.

Comparison With Prior Work

The findings from this study support previous research that shows that wearable technology interventions are both acceptable among adolescents [11] and feasible for implementation within the school setting [10]. Overall, there were few sex differences in adolescent males' and females' experiences and perceptions of RAW-PA, apart from males being significantly more likely than females to like the Facebook page and weekly videos. Co-designing digital health interventions with those affected by the issues of interest is associated with increased engagement in both research and real-world settings [32]. Despite adolescents having input into

the RAW-PA program design before study implementation [22], including input into the format and content of the digital resources and program length, engagement with the RAW-PA Facebook group declined rapidly over time. Although adolescents reported a perceived short-term increase in their motivation for and awareness of physical activity resulting from the device, in line with previous research [33], adolescents' interest in the Fitbit Flex and motivation to achieve daily step goals was also often short-lived and use was reported to be low at the end of the program. This is consistent with a study in the United Kingdom involving 100 adolescents (13- to 14-years old), which showed positive increases in physical activity motivation in response to wearing a Fitbit Charge for 8 weeks, although the effects were not maintained [33]. Although the majority of adolescents reported that the 12-week program was appropriate, these results suggest that shorter programs (eg, 6-8 weeks) may be needed to sustain engagement and interest, particularly for males.

It was also shown that approximately 60% of adolescents found that the Fitbit was comfortable to wear. Design and esthetics are important considerations for wearable activity trackers [34,35], as these can promote engagement and use of devices [18]. Although this device was trialed with adolescents before use [11], and contrasting findings were found in relation to comfort, this was mentioned less frequently than other potential issues such as knowing how to use the device. It is possible that the greater study length (12 weeks vs 6 weeks) may have impacted perceptions of comfort. Overall, this reinforces that comfort is an important consideration for wearable activity tracker interventions.

Overall, social aspects related to peer competition in RAW-PA were viewed positively by adolescents. Nonetheless, in focus groups, adolescents described the competitive goal-setting aspect of RAW-PA as having both a positive and negative effect on their motivation to be active. Peer social influences and support for physical activity are linked to adolescent physical activity [36,37], which may explain the positive experiences among some participants. However, the strength of this relationship via wearable technology and digital behavior change/social media interventions in youth is less clear. In adults, tracking of goals via the Fitbit Flex has been linked to increased motivation for changing physical activity behaviors [38], and web-based social networks have also been effective at increasing physical activity due to the promotion of social comparison (eg, competitive relationships) and support to motivate behavior change [39,40].

Although web-based networks can improve physical activity through social support for healthy behaviors [41], low levels of participant engagement and a lack of overall program adherence in this study make it difficult to draw conclusions regarding the role of peer social support on adolescent physical activity in this context. Nonetheless, the *forced* competitive elements of wearable technology interventions (eg, step count challenges via the Fitbit app) have been linked to a loss of autonomy in adolescents and reduced self-determined motivation to be active [33]. A study exploring the impact of an 8-week Fitbit intervention on 13- to 14-year old adolescents' motivation to be physically active showed that adolescents' autonomous motivation to be active was significantly reduced postintervention [33]. The competition resulting from ongoing self-monitoring via the Fitbit and app in RAW-PA may have negatively impacted some adolescents' autonomy to be active and thus explains the mixed experiences relating to the competitive goal-setting aspects.

From a school-level perspective, however, low levels of burden experienced by teachers in RAW-PA increased program acceptability and feasibility. This was expected given that common barriers to physical activity intervention implementation include, for example, timetabling and staffing constraints, and a lack of integration into the school curriculum [42]. Although RAW-PA was designed to be implemented outside of school and independent of the school curriculum, teachers identified that integrating elements of the program (eg, use of the Fitbit to track activity) within the existing HPE curriculum has the potential to provide leverage for organizational support and thus potentially increase sustainability. Institutionalization of physical activity interventions within the school system has previously been demonstrated for a health education intervention [43], although in-service teacher training was a key factor in the sustainability of such an intervention. Although RAW-PA teachers acknowledged the advantages of fewer implementation demands, teachers wanted greater involvement with and awareness of the program interactions with their students. Teachers also perceived RAW-PA to have positively influenced their own and the adolescents' awareness of physical activity behaviors but recognized that community links may be required for broad program acceptability and thus sustainability in schools. If RAW-PA were to be integrated as part of the existing school curriculum, a degree of staff training may be required. This would also enable teachers to evaluate the advantages and/or disadvantages of increased knowledge/awareness of the intervention and involvement in overall implementation, on their existing time demands. Not only may this enhance ongoing implementation, but it may further embed physical activity promotion within school settings. Future research could explore the benefits of shared involvement between teachers, adolescents, and members of the school community in interventions such as RAW-PA on program effectiveness and long-term institutionalization in this context.

At an organizational level, factors known to enhance the implementation of physical activity interventions include the structure or policies within schools, resources available to support interventions, and the school climate [44]. For example, teachers questioned how such a program could be funded, given the costs associated with purchasing wearable activity trackers. Lower cost wearable activity trackers are currently available that may be purchased for use, either by students individually or as part of entry into a step challenge, for example, but this may still limit uptake into such programs. Teachers also flagged that there was a potential conflict between the RAW-PA delivery format and existing school policies restricting mobile phone use in schools. Adolescents also perceived the receipt of mobile phone–related program content (eg, email and/or text messaging alerts) during class time less favorably. Intermittent participant notifications about RAW-PA content were a central feature of

the intervention, which were delivered predominantly outside of school hours to minimize any potential disruption during the school day. Nonetheless, where and when adolescents downloaded such notifications (eg, open Facebook to retrieve notifications) could not be controlled. As a result, sustainable implementation or system-wide institutionalization of RAW-PA may be substantially limited. Internationally, several countries and jurisdictions have implemented bans on mobile phones in schools, and such measures are increasingly being considered elsewhere [45-47]. Without modifications to the program design, such as changes to the volume and timing of program-related notifications, in its current format, RAW-PA may be a less feasible strategy for further implementation in schools or at scale.

Strengths and Limitations

The strengths of this study are the larger sample size in comparison with previous wearable activity tracker interventions in adolescents, intervention duration, and collection of mixed methods data at multiple levels to understand implementation outcomes. The study was co-designed with the end users (adolescents); therefore, the potential translatability of the findings into practice at a larger scale was considered from the outset. However, the study is not without limitations. Although we considered sex differences during data collection, we were unable to consider sex differences in both the Facebook engagement data and focus group data due to restrictions with data identification. The Facebook engagement data were not split by sex; thus, we cannot comment on any differential use of Facebook or web-based resources by males and females. Restrictions on the Facebook data available for download also meant that it was not possible to determine if the same participant had viewed a Facebook post on more than one occasion. As such, the degree of engagement is based on the sum of all registered participants' *views*, which may underestimate the level of individual engagement for some and overestimate it for others. Focus group data did not contain any identifying data; therefore, sex differences in the perspectives of the adolescents could not be examined further. Given that males and females engage differently with social media [14], it is recommended that future studies investigate any potential disparities and aim to capture data of this detail. The reported technical difficulties associated with the Fitbit Flex, app, and web-based platform may also mean that some adolescents who used their Fitbits were unable to, or chose not to, sync their data or access Facebook. As a result, this may have led to an underestimation of active adolescent engagement in the study. Second, 3 of the focus groups conducted included more than 10 participants, which may have impacted the extent to which each participant could put forth their views and experiences.

Conclusions

RAW-PA showed good acceptability, but engagement and adherence were low among inactive adolescents living in socioeconomically disadvantaged areas. The intervention had high acceptability among teachers at the school level. There was evidence for self-reported perceived short-term positive effects on physical activity motivation and awareness, although these behavior changes were unlikely to be sustained. Low levels of teacher burden enhanced their perceptions concerning the feasibility of intervention delivery. However, sustainable implementation and institutionalization of digital behavior change interventions in schools may be limited by policies restricting the use of mobile phones in schools. Future research exploring the feasibility of differential strategies to engage young people with wearable technology interventions and ways of maximizing system-level embeddedness of interventions such as RAW-PA in practice would greatly advance the field.

Acknowledgments

This project was funded by a Victorian Health Promotion Foundation (VicHealth) Innovation Research Grant. AT was supported by a Future Leader Fellowship from the National Heart Foundation of Australia (award ID 100046); KB and JS were supported by the National Health and Medical Research Council Principal Research Fellowships during this study (APP1042442 and APP1026216). NR is supported by a Future Leader Fellowship from the National Heart Foundation of Australia (award ID 101895). The content of this paper is the responsibility of the authors and does not necessarily reflect the views of the funding bodies. The authors wish to thank all schools and adolescents who participated in the study and acknowledge the contribution of all project staff, including Kara Richards, Stephanie Renehan, Kate Dullaghan, Tom Steele, Dr Felicity J Pendergast, Sarah Tighe, Sophie Aarts, and Holly Beswick.

Conflicts of Interest

NR and JS declare involvement in a start-up technological company. The remaining authors declare no conflicts of interest.

References

1. Kohl HW, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, Lancet Physical Activity Series Working Group. The pandemic of physical inactivity: global action for public health. Lancet 2012 Jul 21;380(9838):294-305. [doi: 10.1016/S0140-6736(12)60898-8] [Medline: 22818941]
2. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010 May 11;7:40 [FREE Full text] [doi: 10.1186/1479-5868-7-40] [Medline: 20459784]

3. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]
4. Nader PR, Bradley RH, Houts RM, McRitchie SL, O'Brien M. Moderate-to-vigorous physical activity from ages 9 to 15 years. J Am Med Assoc 2008 Jul 16;300(3):295-305. [doi: 10.1001/jama.300.3.295] [Medline: 18632544]
5. Drummond M, Drummond C, Dollman J, Abery L. Physical activity from early childhood to adolescence: a literature review of issues and interventions in disadvantaged populations. J Stud Wellbeing 2011 Jan 31;4(2):17. [doi: 10.21913/jsw.v4i2.722]
6. Ball K. Traversing myths and mountains: addressing socioeconomic inequities in the promotion of nutrition and physical activity behaviours. Int J Behav Nutr Phys Act 2015 Nov 14;12:142 [FREE Full text] [doi: 10.1186/s12966-015-0303-4] [Medline: 26572225]
7. Lewis ZH, Lyons EJ, Jarvis JM, Baillargeon J. Using an electronic activity monitor system as an intervention modality: a systematic review. BMC Public Health 2015 Jun 24;15:585 [FREE Full text] [doi: 10.1186/s12889-015-1947-3] [Medline: 26104189]
8. Hoy MB. Personal activity trackers and the quantified self. Med Ref Serv Q 2016;35(1):94-100. [doi: 10.1080/02763869.2016.1117300] [Medline: 26794199]
9. Ng K, Tynjälä J, Kokko S. Ownership and use of commercial physical activity trackers among Finnish adolescents: cross-sectional study. JMIR Mhealth Uhealth 2017 May 4;5(5):e61 [FREE Full text] [doi: 10.2196/mhealth.6940] [Medline: 28473304]
10. Evans EW, Abrantes AM, Chen E, Jelalian E. Using novel technology within a school-based setting to increase physical activity: a pilot study in school-age children from a low-income, urban community. Biomed Res Int 2017;2017:4271483 [FREE Full text] [doi: 10.1155/2017/4271483] [Medline: 29670894]
11. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. Wearable activity tracker use among Australian adolescents: usability and acceptability study. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e86 [FREE Full text] [doi: 10.2196/mhealth.9199] [Medline: 29643054]
12. Pope Z, Zeng N, Zhang R, Lee H, Gao Z. Effectiveness of combined smartwatch and social media intervention on breast cancer survivor health outcomes: a 10-week pilot randomized trial. J Clin Med 2018 Jun 7;7(6):140 [FREE Full text] [doi: 10.3390/jcm7060140] [Medline: 29880779]
13. Shaw JM, Mitchell CA, Welch AJ, Williamson MJ. Social media used as a health intervention in adolescent health: a systematic review of the literature. Digit Health 2015;1:2055207615588395 [FREE Full text] [doi: 10.1177/2055207615588395] [Medline: 29942540]
14. Lenhart A, Purcell K, Smith A, Zickuhr K. Pew Internet Research. 2010. Social Media and Young Adults URL: http://pewinternet.org/Reports/2010/Social-Media-and-Young-Adults.aspx [accessed 2018-09-27]
15. Koorts H, Mattocks C, Ness A, Deere K, Blair S, Pate R, et al. The association between the type, context, and levels of physical activity amongst adolescents. J Phys Act Health 2011 Nov;8(8):1057-1065. [doi: 10.1123/jpah.8.8.1057] [Medline: 22039136]
16. Dute DJ, Bemelmans WJ, Breda J. Using mobile apps to promote a healthy lifestyle among adolescents and students: a review of the theoretical basis and lessons learned. JMIR Mhealth Uhealth 2016 May 5;4(2):e39 [FREE Full text] [doi: 10.2196/mhealth.3559] [Medline: 27150850]
17. Koorts H, Eakin E, Estabrooks P, Timperio A, Salmon J, Bauman A. Implementation and scale up of population physical activity interventions for clinical and community settings: the PRACTIS guide. Int J Behav Nutr Phys Act 2018 Jun 8;15:51 [FREE Full text] [doi: 10.1186/s12966-018-0678-0] [Medline: 29884236]
18. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and effectiveness of using wearable activity trackers in youth: a systematic review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]
19. Coughlin SS, Stewart J. Use of consumer wearable devices to promote physical activity: a review of health intervention studies. J Environ Health Sci 2016 Nov;2(6):1 [FREE Full text] [doi: 10.15436/2378-6841.16.1123] [Medline: 28428979]
20. Hooke MC, Gilchrist L, Tanner L, Hart N, Withycombe JS. Use of a fitness tracker to promote physical activity in children with acute lymphoblastic leukemia. Pediatr Blood Cancer 2016 Apr;63(4):684-689. [doi: 10.1002/pbc.25860] [Medline: 26756736]
21. Mendoza JA, Baker KS, Moreno MA, Whitlock K, Abbey-Lambertz M, Waite A, et al. A Fitbit and Facebook mhealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: a pilot study. Pediatr Blood Cancer 2017 Dec;64(12):e26660. [doi: 10.1002/pbc.26660] [Medline: 28618158]
22. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. A cluster-randomised controlled trial to promote physical activity in adolescents: the raising awareness of physical activity (RAW-PA) study. BMC Public Health 2017 Jan 4;17:6 [FREE Full text] [doi: 10.1186/s12889-016-3945-5] [Medline: 28052773]
23. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice Hall; 1986.

24. Epstein LH. Integrating theoretical approaches to promote physical activity. Am J Prev Med 1998;15(4):257-265. [doi: 10.1016/s0749-3797(98)00083-x] [Medline: 9838972]
25. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: medical research council guidance. Br Med J 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
26. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
27. Australian Bureau of Statistics. 2011. Socio-Economic Indexed for Areas (SEIFA) URL: http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa2011 [accessed 2016-08-20]
28. Australian Government Department of Health. 2014. Does Your Child Get 60 Minutes of Physical Activity Every Day? URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/health-pubhlth-strateg-phys-act-guidelines/$File/Brochures_PAG_5-12yrs.pdf [accessed 2014-09-09]
29. Jakobsson U. Statistical presentation and analysis of ordinal data in nursing research. Scand J Caring Sci 2004;18(4):437-440. [doi: 10.1111/j.1471-6712.2004.00305.x] [Medline: 15598252]
30. Burnard P. A method of analysing interview transcripts in qualitative research. Nurse Educ Today 1991;11(6):461-466. [doi: 10.1016/0260-6917(91)90009-y] [Medline: 1775125]
31. Joffe H, Yardley L. Content and thematic analysis. In: Marks DF, Yardley L, editors. Research Methods for Clinical and Health Psychology. California, US: Sage Publications; 2004:56-66.
32. Partridge S, Redfern J. Strategies to engage adolescents in digital health interventions for obesity prevention and management. Healthcare (Basel) 2018 Jun 21;6(3):70 [FREE Full text] [doi: 10.3390/healthcare6030070] [Medline: 29933550]
33. Kerner C, Goodyear VA. The motivational impact of wearable healthy lifestyle technologies: a self-determination perspective on FitBits with adolescents. Am J Health Educ 2017 Jul 28;48(5):287-297. [doi: 10.1080/19325037.2017.1343161]
34. Schaefer SE, Ching CC, Breen H, German JB. Wearing, thinking, and moving: testing the feasibility of fitness tracking with urban youth. Am J Health Educ 2016 Jan 8;47(1):8-16. [doi: 10.1080/19325037.2015.1111174]
35. Schaefer SE, van Loan M, German JB. A feasibility study of wearable activity monitors for pre-adolescent school-age children. Prev Chronic Dis 2014 May 22;11:130262 [FREE Full text] [doi: 10.5888/pcd11.130262] [Medline: 24854236]
36. Cheng LA, Mendonça G, Júnior JC. Physical activity in adolescents: analysis of the social influence of parents and friends. J Pediatr (Rio J) 2014;90(1):35-41 [FREE Full text] [doi: 10.1016/j.jped.2013.05.006] [Medline: 24156835]
37. Biddle S, Mutrie N. Psychology of Physical Activity: Determinants, Well-Being, and Interventions. Oxfordshire, UK: Routledge; 2008.
38. Asimakopoulos S, Asimakopoulos G, Spillers F. Motivation and user engagement in fitness tracking: heuristics for mobile healthcare wearables. Informatics 2017 Jan 22;4(1):5. [doi: 10.3390/informatics4010005]
39. Zhang J, Brackbill D, Yang S, Becker J, Herbert N, Centola D. Support or competition? How online social networks increase physical activity: a randomized controlled trial. Prev Med Rep 2016;4:453-458 [FREE Full text] [doi: 10.1016/j.pmedr.2016.08.008] [Medline: 27617191]
40. Foster D, Linehan C, Lawson S. Motivating Physical Activity at Work: Using Persuasive Social Media for Competitive Step Counting. In: Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments. 2010 Presented at: MindTrek'10; October 6-8, 2010; Tampere, Finland. [doi: 10.1145/1930488.1930510]
41. Centola D. Social media and the science of health behavior. Circulation 2013;127(21):2135-2144. [doi: 10.1161/CIRCULATIONAHA.112.101816] [Medline: 23716382]
42. Naylor P, Nettlefold L, Race D, Hoy C, Ashe MC, Higgins JW, et al. Implementation of school based physical activity interventions: a systematic review. Prev Med 2015;72:95-115. [doi: 10.1016/j.ypmed.2014.12.034] [Medline: 25575800]
43. Hoelscher D, Feldman HA, Johnson CC, Lytle LA, Osganian SK, Parcel GS, et al. School-based health education programs can be maintained over time: results from the CATCH institutionalization study. Prev Med 2004;38(5):594-606. [doi: 10.1016/j.ypmed.2003.11.017] [Medline: 15066362]
44. Domitrovich CE, Bradshaw CP, Poduska JM, Hoagwood K, Buckley JA, Olin S, et al. Maximizing the implementation quality of evidence-based preventive interventions in schools: a conceptual framework. Adv Sch Ment Health Promot 2008 Jul;1(3):6-28 [FREE Full text] [doi: 10.1080/1754730x.2008.9715730] [Medline: 27182282]
45. Légifrance. 2018. Law N 2018-698 of 3 August 2018 Relating to the Supervision of the Use of the Mobile Telephone in Educational Establishments URL: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000037284333&dateTexte=&categorieLien=id [accessed 2019-08-25]
46. Hymas C. The Telegraph. 2018. Secondary Schools Are Introducing Strict New Bans on Mobile Phones URL: https://www.telegraph.co.uk/news/2018/06/24/secondary-schools-introducing-strict-new-bans-mobile-phones/ [accessed 2019-10-31]
47. Department of Education - NSW Government. 2020. Review Into the Non-educational Use of Mobile Devices in NSW Schools URL: https://education.nsw.gov.au/about-us/strategies-and-reports/our-reports-and-reviews/mobile-devices-in-schools [accessed 2020-02-14]

Abbreviations

HPE: health and physical education
MVPA: moderate- to vigorous-intensity physical activity
RAW-PA: Raising Awareness of Physical Activity
SEIFA: socioeconomic index for areas

Edited by G Eysenbach; submitted 04.02.19; peer-reviewed by J Sirard, K Ng; comments to author 24.05.19; revised version received 30.08.19; accepted 15.04.20; published 07.08.20.

Please cite as:
Koorts H, Salmon J, Timperio A, Ball K, Macfarlane S, Lai SK, Brown H, Chappel SE, Lewis M, Ridgers ND
Translatability of a Wearable Technology Intervention to Increase Adolescent Physical Activity: Mixed Methods Implementation Evaluation
J Med Internet Res 2020;22(8):e13573
URL: https://www.jmir.org/2020/8/e13573
doi:10.2196/13573
PMID:32763872

Original Paper

Information and Communications Technology–Based Interventions Targeting Patient Empowerment: Framework Development

Liran Karni[1], MSc; Koustuv Dalal[1,2], PhD; Mevludin Memedi[1], PhD; Dipak Kalra[3], MD, PhD; Gunnar Oskar Klein[1], MD, PhD

[1]Centre for Empirical Research on Information Systems, Örebro University School of Business, Örebro, Sweden

[2]School of Health Sciences, Mid Sweden University, Sundsvall, Sweden

[3]European Institute for Innovation through Health Data, Gent, Belgium

Corresponding Author:
Liran Karni, MSc
Centre for Empirical Research on Information Systems
Örebro University School of Business
Fakultetsgatan 1
Örebro, 70281
Sweden
Phone: 46 19 301218
Email: liran.karni@oru.se

Abstract

Background: Empowerment of patients is often an explicit goal of various information and communications technology (ICT) (electronic, digital) interventions where the patients themselves use ICT tools via the internet. Although several models of empowerment exist, a comprehensive and pragmatic framework is lacking for the development of such interventions.

Objective: This study proposes a framework for digital interventions aiming to empower patients that includes a methodology that links objectives, strategies, and evaluation.

Methods: This study is based on a literature review and iterated expert discussions including a focus group to formulate the proposed model. Our model is based on a review of various models of empowerment and models of technology intervention.

Results: Our framework includes the core characteristics of the empowerment concept (control, psychological coping, self-efficacy, understanding, legitimacy, and support) as well as a set of empowerment consequences: expressed patient perceptions, behavior, clinical outcomes, and health systems effects. The framework for designing interventions includes strategies to achieve empowerment goals using different ICT services. Finally, the intervention model can be used to define project evaluations where the aim is to demonstrate empowerment. The study also included example indicators and associated measurement instruments.

Conclusions: This framework, which includes definitions, can be useful for the design and evaluation of digital interventions targeting patient empowerment and assist in the development of methods to measure results in this dimension. Further evaluation in the form of interventional studies will be needed to assess the generalizability of the model.

(J Med Internet Res 2020;22(8):e17459) doi:10.2196/17459

KEYWORDS

empowerment; ICT intervention; digital health; eHealth; framework model; ICT patient empowerment model (ICT4PEM)

Introduction

Information and communication technologies (ICT) have already transformed all aspects of health care and, along with technological developments and innovation, ICT will continue to strengthen its role in the future. ICT development initially targeted health professionals as primary users through the provision of electronic health record systems [1]. Recently, however, patients have emerged as additional primary users and targets of ICT interventions. The most abundant examples of this strategy are ICT interventions targeting patient communication (unidirectional or bidirectional), such as messaging, chat services, and real-time video meetings that replace personal contact with physicians or nurses. The social media revolution has led to online patient communities and networks with empowering effects on patients through the exchange of information and group representation [2]. Providing patients with access to their own personal health records and

various online patient education programs has been widely viewed by stakeholders as a tool for patient empowerment [3-5]. ICT-based strategies have also grown in popularity and proved to be efficient in the self-administered management of various conditions such as online cognitive behavioral therapies for depression and anxiety [4,6,7]. Recent efforts include ICT interventions in the form of home-based sensors, wearable devices, and mobile apps for empowering both patients and their physicians by providing accurate and detailed information about disease symptoms and management [8]. The evolving concept designated as quantify-self, along with artificial intelligence and machine learning, harbors as of yet largely unexplored potential in health care.

The concept of patient empowerment is closely associated with the ongoing paradigm shift from a paternalistic to a patient-centered model of engagement. In recent decades, health care systems have undergone a major paradigm shift from a patient-doctor relationship towards a patient-health care relationship where patients, as consumers, are enjoying more equality in the provision of, and access to, their own health care [9,10]. Decision makers, including the World Health Organization, have long considered patient empowerment as a priority, based on the anticipation that it can translate into widespread improvements of disease control, optimized health care utilization, and patient satisfaction. Expectations are high for ICT being capable of improving patient empowerment, self-efficacy, and self-assessment. Accordingly, numerous ICT interventions [5] have claimed to enhance patient empowerment. It is therefore surprising that scientific evidence supporting this notion is rather limited [11]. Although there is general agreement that patient empowerment conceptualizes the enablement of self-control and self-efficacy, its exact boundaries, content, and operationalization have remained elusive and are variably described by published frameworks of the concept [12,13]. Through a lack of unifying definitions, studies that include assessments of patient empowerment often apply arbitrary definitions or simply leave the concept undefined. Studies of patient empowerment usually include components of other potentially associated parameters, such as disease control, quality of life, and health care utilization, as descriptors or integrated parts of patient empowerment. The methodological and conceptual controversies, including patient empowerment assessment, as recently reviewed [14,15], explain and are paralleled by the current inability to reliably measure patient empowerment.

There is a clear demand for further investigations to understand how ICT interventions may affect patient empowerment and alter other important related parameters. However, there is no model that links a conceptual analysis of empowerment to strategies for ICT interventions and evaluation.

Studies that aim to intervene and assess patient empowerment often arbitrarily conceptualize, or simply refrain from providing, a specific definition for empowerment. This makes the interpretation and comparability of these studies difficult and further contributes to conceptualization-related obscurities.

The objective of this study was to develop a framework called ICT for Patient Empowerment Model (ICT4PEM) for ICT interventions that aim to empower patients. The framework provides a methodology to link objectives, strategies, ICT services, and evaluation. Precise definitions of the terms used in the model of empowerment were developed, and examples of indicators to be used for its evaluation are provided.

Methods

This study applies multiple methodologies including a scoping literature review for laying down the theoretical foundation for the ICT4PEM framework, followed by various forms of iterative expert discussions to finalize the concepts of the framework elements. A focus group discussion (FGD) was conducted to review, modify, and finalize the ICT4PEM framework. The utility of the framework was further demonstrated in case studies where patient empowerment had been an explicit objective.

Scoping Review

We started by conducting a scoping review in accordance to the recommendations by Levac et al [16], to identify the relevant literature and knowledge gaps and to establish a theoretical foundation with respect to our research purpose, which was to identify conceptual and methodological issues and necessary requirements for conducting meaningful ICT interventions targeting patient empowerment. A scoping review is effective for identifying a knowledge gap, scoping the body of literature, and clarifying concepts [17].

To identify the questions for researching the concepts of ICT intervention and patient empowerment, we searched for review articles in PUBMED, using the following combinations of search terms: “ICT” OR “eHealth” AND “patient empowerment” AND “Intervention“ AND “Evaluation.” A group of 4 of the authors identified and selected relevant studies by first reading the article titles and then reading abstracts. Every researcher individually read and proposed relevant articles based on our scope and the objective of the study. Then, in a decision-making meeting, all 4 researchers presented their proposed articles. At the meeting, the articles relevant to the objective and research questions were selected through consensus.

We also used the snowball technique to retrieve additional relevant articles [18]. The inclusion of relevant studies with respect to the research question and final purpose was based on the assessment of methodological and contextual relevance through regular meetings. One junior and one senior researcher mapped, summarized, and classified the information in these articles. Potential disagreements were resolved by a third, senior researcher. We charted the information based on the starting points of initial study identification with subsequent iterative subgrouping, in accordance with the accumulating knowledge regarding the research question and purpose.

The analyses did not attempt to include all of the existing literature on this concept, which has been used for many different types of discourses. Even if definitions for various characteristics and measurement instruments of empowerment were searched broadly, the main scope that guided the review and analyses was the use of ICT to empower individual patients. Therefore, aspects of empowerment of all of the patients as a

political group vs. society or health professionals were excluded. Results of the scoping review were summarized and applied as the input and theoretical foundation for the authors' subsequent iterative step, as well as for expert consultations as described in the following sections.

Expert Discussions

The initial draft of the new framework with the inclusion of the initial variables of interest was subsequently discussed in a series of consultations with domain experts, in order to refine the elements of the framework. This involved a series of iterative feedback rounds with experts in electronic health (eHealth), health and implementation science, and patient empowerment. Input from the experts was then integrated into the framework through parallel and iterative rounds of consensus-seeking consultations between the authors, followed by presentations of our draft model at various professional conferences including the eHealth conference Vitalis in Gothenburg, Sweden, May 2019 and Swedish DOME consortium meeting in Karlstad, Sweden, June 2019 and was published as "research in progress" at the AMCIS conference [19] allowing the model to be refined through multiple steps.

Focus Group Discussion

The emerging version of ICT4PEM was further evaluated and modified based on an FGD. The FGD involved 8 participants from Germany, Spain, the United Kingdom, and Sweden. The participants were health care professionals, health systems researchers, and informatics specialists. Due to geographical constraints, the FGD was conducted online using the ZOOM platform and lasted 120 minutes. All the FGD participants received a brief written description of the ICT4PEM a week prior to the meeting. The FGD started with a short presentation of the framework. The FGD objectives were to discuss the design choices, content, and perceived usefulness of the framework. A small number of prompts were used to elicit discussion. The FGD was recorded with due permission from the participants. One senior researcher also took notes during the FGD.

The entire recording of the FGD was transcribed. Qualitative content analysis was conducted, following the guidelines by Graneheim et al [20]. The participants' words were analyzed as the actual content; hereafter, the interpretation and judgment of participants' responses were analyzed as latent content [21]. We analyzed the data with a repeated look over the written transcription by identifying each of the units of meaning and listening to the audio recording [20]. After the analysis and judgments, we modified and finalized the ICT4PEM framework.

Case Study Examples (Demonstration)

The resulting framework was demonstrated by applying it to 2 of our recent projects in eHealth as examples where empowerment had been an explicit objective: the EU-project C3-Cloud and Swedish project EMPARK.

Results

Theoretical Foundation for Developing the ICT4PEM Framework

ICT Intervention-Specific Requirements for the Conceptualization of Patient Empowerment

The conceptual obscurity of the boundaries of this concept, as well as the lack of widely accepted definitions for its included conceptual elements, were identified during the scoping review process as major obstacles for designing, implementing, and evaluating ICT interventions for patient empowerment.

It is important to recall that empowerment as a concept, in general, emerged as a descriptor of a mental state or the process leading to such a mental state in a group of individuals [22]. The approach to try to understand and describe patient empowerment as a mental state by conducting qualitative patient interviews [23] is especially compelling in this regard and may also help to circumvent the conceptual uncertainties [12,24] related to the wide variety of different scholarly definitions. Moreover, patient-derived descriptions of patient empowerment reveal a remarkable pool of internal (patient-perceived) dimensions of patient empowerment, which are consistent among the different studies despite their contextual (disease-specificity, health care, and social background) discrepancies. Behavioral (ie, patient engagement) or other health or health care parameters (ie, quality of life [QoL]) that are in presumed consequential relationships with the perception of internal empowerment fall outside of such a patient-centric patient empowerment paradigm, despite the fact that these parameters are often part of a wider patient empowerment paradigm in other models. The foremost advantages of such a patient-centric empowerment conceptualization become clear when ICT interventions on patient empowerment target the patients. This conceptualization enables the empowerment specificity of the intervention from the patient perspective, while also allowing for an analysis of the possible links between the targeted empowerment characteristics and secondary changes in other health care parameters.

Based on these considerations, 3 major conceptual requirements were identified, which were used as foundational grounds for developing the framework, including (1) patient-centeredness in the definition of boundaries and content for the concept of patient empowerment, (2) providing a clear distinction between patient empowerment defined as patient-derived perceptions and the consequential domains, and (3) clear definitions for each conceptual element.

Specific Requirements for ICT Interventions and Evaluation Targeting Empowerment

Most ICT interventions in health care are multifaceted with respect to their composition of incorporated ICT tools and methodologies. This multimethod approach of ICT interventions and the use of additional methods of interacting with participants have been shown to increase the effect of the ICT intervention. The literature also describes that a lack of conceptual constructs underlying the interventional target confers negative effects on intervention efficacy. Indeed, specific targets for the

interventional steps may remain ill-defined across studies [25,26]. Specifically, ICT interventions for achieving patient empowerment are typical examples of this problem [26].

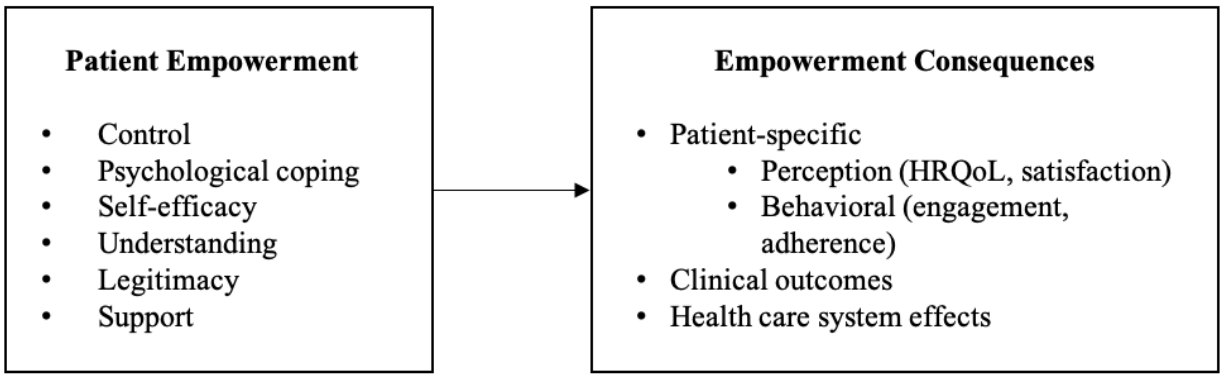
Therefore, the following requirements were identified and subsequently utilized during the framework design: (1) ICT interventions with the primary target of patient empowerment should have their theoretical foundation in an integrated framework that includes a clear conceptualization of patient empowerment, (2) ICT interventions should define empowerment as the primary target of intervention, and (3) the linkage of individual ICT services should correspond with target elements among the patient empowerment characteristics for the scientific quality, evaluability, reproducibility, and comparability of the intervention.

The ICT4PEM Framework for ICT-Based Interventions Targeting Patient Empowerment

The Patient Empowerment Model (PEM)

Figure 1 presents the patient empowerment model (PEM) and depicts the core characteristics of patient empowerment in the left box, while the right box depicts the possible consequences of empowerment that may result from the change in characteristics. The link between the two boxes indicates an indirect effect of empowerment on the consequences, since the latter can be affected by other types of interventions. For a specific ICT intervention, the project management selects the characteristics and consequences that are considered appropriate to study and be influenced. One should also view the attributes depicted here as the most important aspects of empowerment and its consequences, and it may be relevant to add further aspects in a specific case.

Figure 1. Patient empowerment model (PEM). HRQoL: health-related quality of life.



Although the composition of the patient empowerment conceptualization is unique and adapted to the necessities of interventional trial design, the individual conceptual elements of the model and their association with patient empowerment are derived from the literature. Based on the previously described conceptual considerations (see the section Theoretical Foundation for Developing the ICT4PEM Framework), constructs of PEM were selected based on qualitative studies aiming to understand the patient perspective in patient empowerment. The review performed by Agner et al [23] was used as a basis to identify conceptual elements that are common across the incorporated studies, irrespective of a specific disease or health care background. This selection consequently represents the core concept of perceived patient empowerment. Based on the findings by Agner et al [23], knowledge, control, and coping related to the disease and the process of health provision were selected along with the feeling of support and legitimacy, as explained in detail in Table 1.

Table 1. Characteristics of empowerment.

Characteristic	Definition
Control	ability by which an individual can decide about his or her level of engagement in the health care process and participate in decisions regarding alternative treatment options, also when these are performed by professionals
	Note: Patient control is described by [27] as "having the opportunity to use power, namely, making choices, implementing intentions, taking action, and affecting the actions of others". Our definition of control is similar, though we consider control to be an ability, rather than an opportunity, as the latter suggests an imbalanced power situation between the provider and the patient.
Psychological coping	state of a process in which one psychologically tries to adapt to the challenges associated with the negative changes of health status
	Note: Coping strategies are traditionally divided into problem-focused coping (managing or altering the problem) and emotional-focused coping (regulating the emotional response to the problem) [28]. Our definition encompasses both these perspectives.
Self-efficacy	sum of cognitive and physical capabilities possessed by the patient that can be used for self-care
	Note: Köhler et al [29] stated that self-efficacy influences a patient on his/her thinking, feeling, motivation and his/her action towards an attempt of new health behavior.
Understanding	potential use of the information a patient has regarding his or her own *health status*, the *diseases*, and the function of the actual and possibly available *health care processes*
	Note: Understanding, in our definition, represents the patient´s capacity to apply *knowledge* in the specific and individual context of the disease and healthcare provision. *Information* and its availability to the patient serve as a base for understanding. Consequently, neither knowledge, nor information are sufficient as characteristics of *empowerment*, though both should be considered as pre-requisites for understanding.
Legitimacy	perception that the care from a professional health care system is fair with regard to issues of being lawful in the jurisdiction and available with a sufficient degree of equity
	Note: Legitimacy refers to the patient´s perception of fairness and trust in the healthcare system in general or in a specific situation. This is particularly important when the care situation may include aspects that are beyond the direct control of the patient, such as when a person is unconscious. The second aspect is associated with the perception of the right to receive required services to the same degree as other persons.
Support	quantity and quality of support as assessed by an individual of support that is being offered or received from the care provider or the non-medical supporting environment
	Note: Support can both enhance and decrease autonomy. Elderly, sick individuals often receive high amounts of unsolicited help that, although normally well-intentioned, may further reduce their possibility of making choices, autonomy, self-esteem and their longer-term competence or coping abilities [30].

We followed the general principles of the International Organization for Standardization (ISO) when constructing the definitions [31]. Thus, a definition is a single phrase that can replace the term wherever used and does not start with an article (eg, "a", "the") or end with a full stop.

Possible consequences of patient empowerment are various. The literature identifies several of these concepts either in connection to or as part of patient empowerment. Although their connections to the other conceptual dimensions of patient empowerment were mapped using the patient empowerment model of Bravo et al [12], PEM (Figure 1) rather considers them as part of an extended empowerment concept. Patient empowerment consequences are divided into 4 major groups: patient perceptions with close conceptual relationship to empowerment perception, (health-related quality of life [HRQoL] and patient satisfaction); behavioral elements (patient engagement and adherence); and finally, clinical outcome and health care system effects as their own categories, with definitions given in Table 2.

Table 2. Characteristics of empowerment consequences.

Characteristic	Definition
Health-related quality of life (HRQoL)	impact of health status on a person's quality of life assessed with a multidimensional instrument Note 1: It is measured in general by EQ-5D [32] or SF-36[a] [33]. Note 2: There are also disease-specific HRQoLs available (eg, the PDQ8[b] for Parkinson's disease [34].
Patient satisfaction	degree of fulfilment of the patient's expectation of the health care services received Note 1: Patient satisfaction has been the subject of research and a fundamental driving force behind the health care policy development for decades. Satisfaction as a concept shows a tight linkage to the concept of expectations [35]. Note 2: Indicators may measure timeliness, effectiveness, and patient-centeredness as well as quality of health care staff including their interpersonal communication skills. In general, it may also include patient views on accessibility and the state of health care facilities. Note 3: There are several validated instruments for measuring patient satisfaction (eg, PSQ-18[c] [36], GS-PEQ[d] [37] and SAPS[e] [38]). However, these instruments are non-generic and context-dependent.
Patient engagement	degree to which the patient is an active agent for managing their own health Note 1: This involves actions and collaborative partnerships at complex levels including the individual, familial, organizational, and health care policy levels and often in response to the recommendations of the health care professional system. Note 2: Studies assessing health care performance on patient engagement are few in number and limited by the lack of instrument able to assess it [39]. Recently, Graffigna et al [40] made efforts towards a conceptualization followed by the development and validation of a construct (Patient Health Engagement Scale) for patient engagement. More recently, researchers introduced another validated, 20-item construct, the Patient Engagement Index (PEI) for assessment [39].
Adherence	degree to which the patient's behavior follows a care plan agreed to by the health professional and the patient Note 1: The care plan may not be an explicit document but a mutual understanding of a professional recommendation. Note 2: Patient adherence as a concept has practically replaced *compliance* with the evolution of the patient's role in health care from being a passive, obedient recipient of a physician's authority to an active partnership. Our definition is in line with the WHO's[f] definition, which describes patient adherence as "the extent to which a person's behavior — taking medication, following a diet, and/or executing lifestyle changes — corresponds with agreed recommendations from a health care provider."
Clinical outcomes	professional measurable health state of a patient allowing for a comparison before and after a health care intervention Note: The measurable indicator is often dependent on laboratory analysis or chemical, microbiological, physiological, or medical devices used or at least controlled by the health professional organization. However, clinical outcome can also be measured using some form of professional standardized clinical assessment following a defined process of observation. Further, patient-reported outcome measures can be regarded as clinical outcomes, provided they are collected in a way that is professionally validated.
Health care system effects	health care system resource utilization before, during, and after a specific new procedure such as an ICT[g]-based patient empowerment intervention Note: Here, we consider both productivity and quality, which can also be described as the degree to which the goals of the organization are fulfilled in relation to resource utilization. Productivity in general is the total production divided by the total resources used, in this case to produce some specific health care service. The analysis of health system effects can include changes measured economically but the effects can also be analyzed and described qualitatively such as a new process being established or organizational change.

[a]SF-36: Short-Form 36.

[b]PDQ-8: Parkinson's Disease Questionnaire-8.

[c]PSQ-18: Patient Satisfaction Questionnaire Short Form.

[d]GS-PEQ: Generic Short Patient Experiences Questionnaire.

[e]SAPS: Short Assessment of Patient Satisfaction.

[f]WHO: World Health Organization.

[g]ICT: information and communications technology.

ICT Interventional Strategies for ICT4PEM

Using the ICT4PEM framework (Figure 2) requires that an interventional strategy be explicitly selected for influencing the empowerment characteristics that the specific intervention addresses. For the ICT design strategy, our model inherits some elements of the Behavioral Interventional Technology (BIT) model [41]. "Education, Feedback, and Monitoring" are presented as conceptual approaches to achieve behavioral change in the BIT model. The "Engagement" strategy in the ICT4PEM framework corresponds well and combines what is described as "aim setting" and "motivation" in the BIT model to achieve behavioral change. Engagement as an ICT interventional strategy may be coupled to multiple core empowerment characteristics of PEM, such as "Control, Legitimacy, Support, and Coping." "Communication and Analysis" were added as additional and frequently applied strategies to those already mentioned in the BIT model. "Communication and Analysis" could lead to an improvement in "Knowledge, Legitimacy, and Control" as core PEM characteristics. "Monitoring" disease activities and providing patients with "Feedback" are additional strategies with the potential to improve the PEM characteristics of "Understanding, Control, and Coping." Taken together, ICT4PEM includes "Education, Feedback, Monitoring, Communication, Analysis, and Engagement" as interventional strategies. The application and proportional contribution of these strategies in the intervention may differ between studies based on the specific context of the patient empowerment intervention. Importantly, the ICT4PEM framework provides precise definitions for each of these strategies in order to enhance conceptual clarity and interoperability. These definitions, along with some important considerations, are summarized in Table 3.

Figure 2. Information and Communications Technology for Patient Empowerment Model (ICT4PEM) framwork.

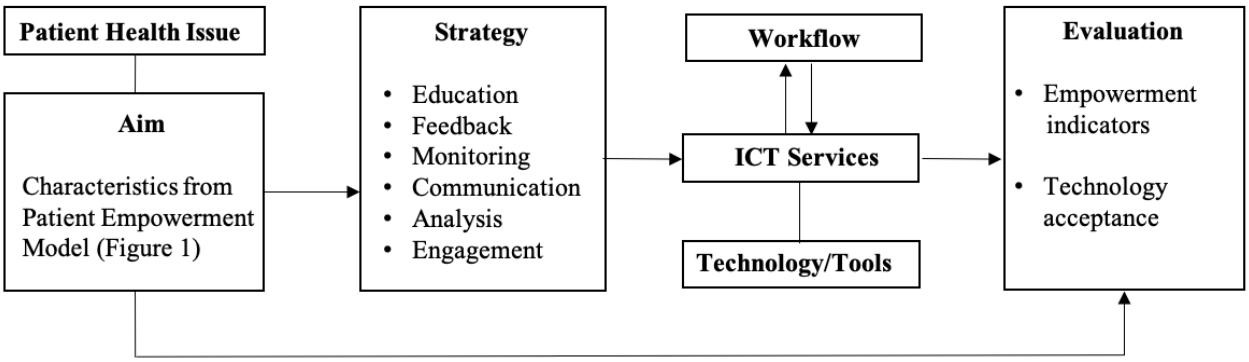


Table 3. Definitions of the information and communications technology (ICT) intervention strategy.

Component	Definition
Education	process where an intervention aims to provide the patient with increased knowledge through a defined learning process to empower patients for increasing their understanding of their health situation, characteristics of their disease(s), and participation in shared decision-making concerning the professional health care utilization and for self-care Note 1: Educational strategies may include making knowledge information in text, audio, videos, or interactive learning media available or, alternatively, educational programs where a certain course material is prescribed or required for an individual patient in a certain situation.
Feedback	
Feedback to the patient	information from the health care system to the patient regarding the present measured health state of the subject that is based on monitoring and presented to the patient at certain intervals or made available for retrieval at the discretion of the patient
Feedback to the health care system	patient-reported assessment of opinions on the care given at a particular health care organization and possibly an identified health professional
Monitoring	planned and systematic process of observation of a patient's health status or health care activities, which closely follows a planned course of activities that may include medical technologies and reports to perform the observations Note 1: Health care professionals monitor patients from health care perspectives for optimizing organizational benefits subject to patients' health status and compares between what is "happening," what is "expected to happen" and what is intended to happen. Note 2: Monitoring in this context can often serve to allow giving feedback to the patient.
Communication	process where information is flowing between a patient and a health care professional in a bidirectional manner
Analysis	process where collected data are analyzed with a specific health issue in mind and providing the resulting information to a health professional or the patient Note 1: The techniques used for the analyses may depend on evidence-based clinical guidelines where patient specific data is analyzed algorithmically to provide diagnosis or recommendations for treatment, including behavioral changes to be undertaken by the patient. Note 2: Analyses may also be based on machine learning.
Engagement	technique for stimulating the patient to be an active agent for managing their own health

Structural Components and Operational Characteristics of the ICT4PEM Framework

The individual components and structure of the ICT4PEM framework are shown in Figure 2. Most of the conceptual elements of the framework are described in detail including operational definitions in the prior two subchapters. The "Workflow, ICT Services, Technologies/Tools, and Evaluation" components of the framework emerged via the expert discussions and later the FGDs. This structure of ICT4PEM represents the integration of the PEM empowerment conceptualization and empowerment characteristics (Figure 1) into the interventional framework and was developed based on the identified specific requirements for an ICT intervention and evaluation specific to patient empowerment targeting. The core patient empowerment characteristics described by PEM correspond to the conceptual aims for the intervention regarding both its targeted design and evaluation. At this stage, at least one or possibly several internal empowerment characteristics should be selected as the primary aims of an intervention whenever an explicit goal is the improvement of empowerment. ICT4PEM also ensures the context specificity of the interventional design by incorporating specific health issues for the patient population that will be the target of the intervention. Health issues are not predefined in the framework. Examples of health issues include, but are not limited to, obesity, diabetes, or Parkinson's disease.

In the next step, the framework requires a chain-like coupling of each targeted internal PEM empowerment characteristic with one or more "Strategies." These "Strategies" are described in detail with operational definitions in the previous section. Each selected "Strategy" element should be coupled to specific "ICT Services" with some functionality to support these. Obviously, a given "ICT Service" may be coupled to several "Strategies," and one "Strategy" may be addressed by several independent "ICT Services". These services will need to be selected for each specific case and may include a reminder function, set of educational videos, comprehensive display of symptom development to current date, and a messaging service. These services should be delivered by specific "Technologies/Tools" (eg, internet, apps, sensors). ICT4PEM purposely refrains from

providing any list and definitions for elements regarding the "ICT Services" and "Technologies/Tools" based on consideration of the rapid technological development that would deem any list soon outdated and the fact that it lies outside the primary goals of ICT4PEM. The framework emphasizes the importance of understanding how the "Workflow" functions in a specific field of application in health care. It includes a consideration of the care plan or how the services will be used in a temporal relation to a symptom or other developments in the course of the patient's life with the health issue (eg, a service may be designed to mitigate a rare event that may occur for some patients and will only be relevant in such a workflow).

This indicates that an ICT intervention design using ICT4PEM must be considered as a composite of elementary interventions that can be grouped based on the applied ICT tools at the basic level, ICT services at the intermediate level, and ICT strategies at the highest level. Importantly, the framework provides correspondence between intervention and patient empowerment at the basic level, which means that each elementary component of the intervention should target one or several components of patient empowerment with an associated strategy for evaluation. This approach ensures precise, detailed, and conceptually rooted planning and evaluation, along with the possibility for direct comparison between interventional trials.

The framework emphasizes that the intervention project should develop a strategy for "Evaluation" that considers the defined "Conceptual Aims" as primary targets for evaluation. This ensures that the selected PEM empowerment characteristics are in the epicenter of both the interventional and evaluation design. The level of evaluation should reflect and correspond to the level of the interventions. The main rule is that any conceptual component of PEM that is specifically targeted by an intervention should be evaluated. If it is not possible, attempts must be made to explain this deviation and set up the possible link between the intervention and observed changes in the indirect parameters, such as "Empowerment Consequences." Finally, we included "Technology Acceptance" as an additional indicator for the "Evaluation" domain based on the considerations that no ICT intervention can achieve the desired empowerment unless the technology is used and accepted.

In many cases, an ICT intervention after completion of the first design, implementation, and evaluation would benefit from iterated design to address the possible weaknesses discovered during evaluation. The new round may include changes to the "Strategies," ICT Services, or various technologies used by these followed by a new evaluation study. ICT4PEM is an overall conceptual framework for the core design process rather than an iterative process.

As described previously, the PEM model features a narrow empowerment conceptualization and, correspondingly, primary targets for the intervention. On the other hand, ICT4PEM does not exclude secondary targets for intervention and evaluation, mostly corresponding to the consequential domain of PEM. As previously explained (in the section ICT Intervention-Specific Requirements for Conceptualization of Patient Empowerment), this distinction between core (perceived) empowerment and empowerment consequences within PEM and its correspondence to the ICT targets are those key components of the framework that help to avoid otherwise inevitable misinterpretation and confusion regarding the interpretation of the results of interventions.

ICT4PEM sets stricter requirements regarding the target of an intervention than the evaluation. For instance, the application of the framework can be applied to the intervention design even if that exclusively evaluates the consequences of empowerment (ie, patient engagement, QoL) with the condition that the primary intervention occurs on the core empowerment as defined by PEM. The theoretical foundation underlying the framework excludes studies from its scope where the sole target of interventions falls within the consequential domain of PEM, even if the evaluation would include the perceptive empowerment in agreement with its definition of PEM. This remains true, even though it is theoretically possible that the intervention on the consequential domain of PEM (ie, messaging patients with encouragement for engagement in the curing process) may indirectly enhance the core (perceptional) patient empowerment as well. However, the framework does not consider such interventions as "empowerment targeting," although it acknowledges that the wide variety of "side interventions" may affect patient empowerment.

Case Study Examples

Case Study C3Cloud

The C3-Cloud project focuses on elderly patients with diabetes, heart failure, renal failure, and depression in different comorbidity combinations [42]. Three European pilot sites are involved in the study: Osakidetza (Basque Country, Spain), Region Jämtland Härjedalen (Sweden), and South Warwickshire NHS Foundation Trust (United Kingdom). The C3-Cloud system consists of a variety of components: the Coordinated Care and Cure Delivery Platform; Patient Empowerment Platform (PEP); and Clinical Decision Support Module.

The C3-Cloud services include services for both health professionals and patients with multiple conditions. Empowerment was an important explicit goal of the application and the project planning. The PEP provides an internet-based interface for the patients. They can study various educational materials in tailored homework or a possible task program that can entail going through the educational resources selected by the professional for the coming period. The patients are also able and encouraged to review the documented care plan from home, not only once after a scheduled visit but repeatedly as needed. Finally, in the care plan, patients have various tasks to perform such as lifestyle changes regarding exercise, diet, or smoking cessation. These activities should be performed in addition to taking the medication as prescribed during the interaction with the physician and available through the PEP. As stated previously, some patients also have scheduled monitoring tasks that may include daily measurements to be registered into the system and reviewed periodically in another workflow in the life of the patients during the C3-Cloud project.

Here, we illustrate how ICT4PEM could be used to make explicit the various aspects of empowerment in relation to this project.

First, we detail the goals as related to the "Empowerment Characteristics," noting that the project also has other goals that can be described as empowerment consequences.

Improving patients' and, when relevant, their informal caregivers' "Understanding" is a key important objective. The ICT strategy that was selected to improve understanding was "Education" through a set of resources in the form of informative documents and short videos provided through the "ICT Service" PEP, to which the patients had access from home through the internet and a tablet PC as a "Technology." These educational resources were individually selected by a professional to meet the current stage of disease and the patient's presumed ability. The "ICT Service" had a large set of educational resources for each of the four disorders we targeted, and patients had two or more of these disorders.

Another strategy selected to improve understanding was "Feedback." In this case, this means that the professional care plan with results such as lab tests and planned medication and other actions is available to the patients through the "ICT Service" PEP. For some patients, we also selected "Monitoring" as a strategy, where the patients could automatically register certain measurements such as weight or blood pressure using special devices as "Technology" at home, into the system through the PEP interface to the system. The results then became available to the professionals as well to the patients.

Finally, "Communication" was a strategy to increase feelings of "Support" as another major characteristic and to some extent to exert "Control" as another important empowerment characteristic. This was implemented through a "Messaging" ICT service, being part of the PEP.

The different ICT services for the patients to improve empowerment has a special relationship to the "Workflow" of the patient care processes. The start of the use of the system comes after a scheduled visit to the primary care physician where the current state of the chronic diseases was evaluated and a care plan was drawn up using automatic decision support for the health professionals based on the patient data and available clinical guidelines. This aspect targets the professional user, but the patient is of course also benefitting from hopefully improved quality of care through this support service.

There are a number of "Empowerment Consequences" that are possible results of the C3-Cloud project and that will be evaluated during the project: HRQoL, "Patient satisfaction," "Clinical Outcomes," and "Health Care System Effects."

HRQoL, as measured with the EQ-5D, could potentially be affected, although the chronic nature of the selected diseases and the relatively small number of patients in our study means we may not detect it in this study.

"Patient Satisfaction," defined as "the degree of fulfilment of the patient's expectation of the health care services received" will be assessed using our own instrument rather than using tools such as the Patient Satisfaction Questionnaire Short Form (PSQ-18), Short Assessment of Patient Satisfaction (SAPS), or Generic Short Patient Experiences Questionnaire (GS-PEQ).

"Clinical Outcomes" will be measured and compared to a control group from the same regions receiving the same care in principle but without the C3-Cloud intervention.

"Health Care System Effects," defined as "the health care system resource utilization before, during, and after a specific new procedure such as an ICT-based patient empowerment intervention" will be partly measured through interviews with professionals and decision makers and also by a calculation of potential economic consequences if the project should be launched on a large scale either throughout the regions or nationally.

An assessment program has been planned where patients' use and views on the various components are evaluated using questionnaires. These are both related to "Technology Acceptance" but can also be seen as addressing empowerment even if the explicit goals as stated here are not mentioned in the design of these questionnaires.

Case Study EMPARK

To explain the ICT4PEM framework, we provide another example of a system (EMPARK) aimed at improving the treatment of Parkinson's disease by empowering patients with data from their own measurements [43]. EMPARK is an Internet of Things–based system designed to help patients improve their self-management and increase their self-awareness. The system consists of sensors for logging motor function and sleep information, an electronic dosing device for logging and delivering dose intakes, and a tablet-based app for logging daily activities, meal timing, and self-assessments from the patients' home environments. A separate tablet-based app feeds the gathered information back to the patients, providing them access to their individual symptoms and activity records. Additionally, treating clinicians gain access to the detailed data of their own patients through a web application. The overall objective of EMPARK is to promote patient empowerment and thus improve patients' HRQoL.

Using the ICT4PEM framework, the primary "Aims" are first defined from the "Empowerment characteristics," which in this case will be to provide "Understanding" and "Control" to the patients. The selected "Strategies" consist of "Monitoring" and "Feedback" for collecting and visualizing the data, respectively. The "ICT Services" include the collection of sensor-based and patient-reported data as well as the patient interface where the data is presented in an easy-to-understand and comprehensive manner. These services are intended to be used in routine clinical practice to allow patients to take an active role in the decision-making process and improve the patient-clinician interaction ("Workflow"). The services are supported by "Technologies/Tools" such as the sensors, custom applications, and communication architecture. The "Evaluation" of the intervention using the EMPARK system focuses on the extent of achieving the empowerment characteristics, which are "Understanding" and "Control" and "Technology Acceptance," which has been studied separately [44]. Possible "Empowerment" consequences of the EMPARK system include: "HRQoL Outcomes," as measured by EQ-5D and Parkinson's disease-specific instruments (eg, the 39-item Parkinson's Disease Questionnaire); "Patient Satisfaction," as measured

through customized interviews; "Patient Engagement," as measured by analyzing systems data (eg, compliance with measurements using the sensors); "Patient Adherence," as measured by analyzing systems data (eg, adherence to the treatment plan); and "Clinical Outcomes," as measured by Parkinson's disease–specific clinical rating scales (eg, Unified Parkinson's Disease Rating Scale).

Discussion

The objective of our study was to develop a framework for ICT interventions that aim to empower patients. To the best of our knowledge, the ICT4PEM is the first framework model that pursues this goal and was developed to facilitate and provide guidance for designing, implementing, and evaluating ICT interventions on patient empowerment. The ICT4PEM systematically addresses an array of challenges that investigators commonly face during the design of the ICT interventions.

Improving patient empowerment via ICT-based interventions is a commonly accepted strategy, but the evaluability and real scientific value of these efforts have remained elusive largely due to conceptual and methodological limitations. Consequently, it frequently remains unclear what exactly was targeted and how the effect was evaluated in case of specific ICT interventions on patient empowerment. Previous conceptual models of patient empowerment have limited value to be utilized as a framework or patient empowerment conceptualization model for ICT interventions. Early patient empowerment models were characterized by narrow contextuality that limited their applicability outside a specific clinical scenario or disease setting [12]. Comprehensive and synthesizing conceptual models of patient empowerment were recently published [12,45,46]. The model by Bravo et al [12] is the latest and most comprehensive that positions patient empowerment in a wider contextual perspective by elaborating psychological, medical, or social constructs that conceptually surround patient empowerment [12]. In our view, the possible contribution of this model to ICT intervention design targeting patient empowerment is in its ability to reveal potential associations and secondary targets for the intervention. Although the model by Bravo et al [12] may help to identify potential conceptual elements of patient empowerment for an ICT intervention design, it lacks definitions for the included conceptual elements of empowerment and allows speculations regarding the described parameters being part or falling outside the scope of the patient empowerment conceptualization. Although ICT4PEM utilizes some elements of the patient empowerment model of Bravo et al [12] in the identifications of potential interventional targets, ICT4PEM contrasts that model by several means. Our framework, and specifically its empowerment conceptualization model (PEM), provides clarity with ISO-style definitions and clear separation of core characteristics of patient empowerment and consequential domains. In contrast to prior patient empowerment models, which largely inflated or contextually limited the conceptual boundaries and components of patient empowerment, the ICT4PEM framework purposely narrows the conceptual boundaries of patient empowerment to the core, perceived characteristics from the patient-centric approach. This is based on both theoretical and practical considerations. The theoretical considerations include the deep patient centeredness of empowerment and the inherited difficulties to establishing a commonly accepted definition and conceptual boundaries. In contrast to prior synthesizing approaches, we turned our attention to studies that provided pure patient-derived conceptualization by conducting patient interviews. It helped us to elaborate PEM, a patient-derived empowerment self-conceptualization with minimal modifications. Consequently, PEM exclusively includes patient perceptions as core empowerment, and behavioral domains of prior empowerment models are instead grouped together with other parameters as patient empowerment consequences. Providing definitions in the ISO style for each conceptual element both within the core empowerment and consequential domains of the PEM further facilitates the precise targeting and evaluation of the intervention. We argue that these features of the ICT4PEM are key to promoting future interoperability and comparison of the results between ICT interventions on patient empowerment.

It is generally assumed that improving patient empowerment translates into wider benefits for both the patients and health provider, such as better quality of life, disease control, or more effective health care utilization. However, there is subtle evidence to support this notion. ICT interventional trials using the ICT4PEM framework can now provide the possibility to quantitatively analyze the association between core patient empowerment and possible consequences, such as changed patient behavior, disease management and control, quality of life, and health care utilization.

The ICT4PEM framework provides ICT interventions with the methodology to link interventional objectives, strategies, ICT services, and evaluation centered on a clarified conceptualization of patient empowerment by PEM. The ICT4PEM framework explicitly sets patient empowerment and its consequences as the objectives for the elaboration of the selected strategies and ICT services. Technically, it means the requirement for allocating one or several specific ICT interventional steps to one or several PEM core characteristics for their modification. This design also ensures a detailed inventory of ICT interventions, where the elementary ICT components are defined at the level of the technology and ICT tool used and grouped further based on the ICT service type, ICT strategy, and specific intervention, respectively. With respect to the included list of ICT strategies and services, ICT4PEM benefits from prior eHealth-related ICT frameworks. The framework follows the logic of the BIT [41] in two specific ways. First, BIT, similarly to the core concept of ICT4PEM, makes a distinction between conceptual and instantiation domains. Second, it also describes ICT developments as a chain of developmental phases, with each having a well-defined specific target at the conceptual level. However, in contrast to ICT-PEM, BIT defines broadly all ICTs that support users in changing behaviors and cognitions related to health, mental health, and wellness, while not addressing the conceptual controversies of patient empowerment. Another difference between ICT4PEM and BIT is that ICT4PEM does not intend to provide a detailed recipe for the technological instantiation of each ICT intervention step,

as it is unnecessary for the overall goal and would pose unnecessary limitations to its applicability.

Although ICT4PEM demands the inclusion of the targeted PEM core characteristics into the evaluation, it refrains from providing methodological recommendations for the evaluation. Instead, it extends the wider applicability of the framework by suggesting that the exact methodology of evaluation should follow the specific interests of the study. However, the framework specifically addresses two important issues. One, as it was previously explained, is the strong recommendation for the inclusion of the targeted PEM core characteristics into the evaluation. The other consideration is technology acceptance. The role of technology acceptance in the efficacy of ICT interventions is well known and described in the literature [47,48]. More specifically, it is one of the major sources of possible bias in the observed results based on the characteristic of the study population. We warrant that ICT4PEM should be applied in conjunction with well-established frameworks, such as the Technology Acceptance Model [49], during the evaluation phase of the ICT4PEM, in order to mitigate such biases.

The application of ICT4PEM was demonstrated by showing how it could be functional in the design and evaluation of two ongoing, complex ICT interventions with different health status backgrounds: C3Cloud and Empark. What is common in these ICT interventions is their incorporation of patient empowerment as a major target, besides other goals. We emphasize the real-life aspect of our examples, as ICT interventions on empowerment are rarely limited to empowerment, but rather are complex regarding their background (ie, health status, workflow) and desired effects. Although our example projects were not designed by ICT4PEM, our analyses clearly identify the applicability of the framework, as it was possible to identify an array of individual ICT interventional steps with clear targets within the PEM core or consequential characteristics. During the utilization of ICT4PEM in their later, evaluation phase, these projects could benefit most from (1) clear definitions for empowerment characteristics by ICT4PEM, (2) establishing the proportional contribution of the single interventions to change in a given parameter, and (3) establishing possible causative relationships between the observed changes of the parameters regardless of their original relative importance as a target. The first point is a straightforward advantage in light of the current confusion with respect to the patient empowerment conceptualization. The latter two points need explanation. ICT4PEM anchors the interventions to the workflow of health provision and consequently enable the adjustment and harmonization between the temporality of the evaluation and interventional steps. For example, investigators may identify the improvement of understanding and control (as core PEM characteristics) to be a cause of observed improvement in patient engagement and HRQoL (as consequential PEM characteristics) and not the other way around, by establishing the temporality of these changes. This could be done by sequential interventional and evaluation steps designed to fit the specific workflows of the health care environment in which the study is implemented.

Limitations

The ICT4PEM framework has some limitations. Most importantly, it has not yet been used to guide complex ICT interventions starting from the design to the evaluation. This warrants the real-life validation of ICT4PEM. Although we demonstrated that ICT4PEM can be used to guide empowerment-targeted strategies within very complex interventions, the full complexity of ICT intervention projects may reveal some specific design and evaluation issues that ICT4PEM does not cover. There are ample other conditions for the overall success of any ICT intervention in health care that are beyond the focus of our framework. We are aware that successful ICT project implementation in health care depends on several other crucial conceptual elements and practical considerations as well and that all should be taken into account. ICT4PEM provides a vital conceptual background and requirements design for patient empowerment but we suggest our model be used together with broader models of ICT implementation within health care. For instance, the CeHres Roadmap by van Gemert-Pijnen et al [50] is an established and well-cited model with a holistic approach to eHealth project implementation in general.

Applicability of ICT4PEM for interventions that only marginally address patient empowerment and set other parameters as their main target may instead benefit from other ICT models, such as the BIT [41]. Finally, we acknowledge that ICT4PEM does not address the problems of patient empowerment operationalization. Specifically, the framework could not include instruments on how to measure core empowerment as well as consequences for specific situations due to the limitations of the quantity and quality of the published literature on this issue. Further research is needed to validate our framework and to provide examples on its applicability in different diseases, health care and ICT intervention settings.

Conclusions

The new framework, ICT4PEM, with newly proposed definitions of core characteristics of empowerment and empowerment consequences, can be useful for the design of ICT interventions targeting empowerment and can assist the development of methods to measure the results in this dimension. However, further evaluation in future interventional studies are required to assess the generalizability of the model.

Acknowledgments

This project was supported by the European Union Horizon2020 C3-Cloud project and the Swedish Knowledge Foundation project Empark. The authors would also like to thank all experts that kindly provided feedback on early drafts of this study.

Conflicts of Interest

None declared.

References

1. Yi MY, Jackson JD, Park JS, Probst JC. Understanding information technology acceptance by individual professionals: Toward an integrative view. Information & Management 2006 Apr;43(3):350-363. [doi: 10.1016/j.im.2005.08.006]
2. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16(1):442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
3. Bouayad L, Ialynytchev A, Padmanabhan B. Patient Health Record Systems Scope and Functionalities: Literature Review and Future Directions. J Med Internet Res 2017 Nov 15;19(11):e388 [FREE Full text] [doi: 10.2196/jmir.8073] [Medline: 29141839]
4. Samoocha D, Bruinvels DJ, Elbers NA, Anema JR, van der Beek AJ. Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. J Med Internet Res 2010 Jun 24;12(2):e23 [FREE Full text] [doi: 10.2196/jmir.1286] [Medline: 20581001]
5. Calvillo J, Román I, Roa LM. How technology is empowering patients? A literature review. Health Expect 2015 Oct;18(5):643-652 [FREE Full text] [doi: 10.1111/hex.12089] [Medline: 23711169]
6. Lau Y, Htun TP, Wong SN, Tam WSW, Klainin-Yobas P. Therapist-Supported Internet-Based Cognitive Behavior Therapy for Stress, Anxiety, and Depressive Symptoms Among Postpartum Women: A Systematic Review and Meta-Analysis. J Med Internet Res 2017 Apr 28;19(4):e138 [FREE Full text] [doi: 10.2196/jmir.6712] [Medline: 28455276]
7. Rost T, Stein J, Löbner M, Kersting A, Luck-Sikorski C, Riedel-Heller SG. User Acceptance of Computerized Cognitive Behavioral Therapy for Depression: Systematic Review. J Med Internet Res 2017 Sep 13;19(9):e309 [FREE Full text] [doi: 10.2196/jmir.7662] [Medline: 28903893]
8. Domb M. Wearable Devices and their Implementation in Various Domains. In: Nasiri N, editor. Wearable Devices-the Big Wave of Innovation. London: IntechOpen; Dec 04, 2019.
9. Alexander JA, Hearld LR, Mittler JN, Harvey J. Patient-physician role relationships and patient activation among individuals with chronic illness. Health Serv Res 2012 Jun 18;47(3 Pt 1):1201-1223 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01354.x] [Medline: 22098418]
10. Gerber BS, Eiser AR. The patient physician relationship in the Internet age: future prospects and the research agenda. J Med Internet Res 2001 Apr 9;3(2):E15 [FREE Full text] [doi: 10.2196/jmir.3.2.e15] [Medline: 11720957]
11. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating Patient Empowerment in Association With eHealth Technology: Scoping Review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
12. Bravo P, Edwards A, Barr PJ, Scholl I, Elwyn G, McAllister M, Cochrane Healthcare Quality Research Group, Cardiff University. Conceptualising patient empowerment: a mixed methods study. BMC Health Serv Res 2015 Jul 01;15(1):252 [FREE Full text] [doi: 10.1186/s12913-015-0907-z] [Medline: 26126998]
13. Castro EM. Patient Participation and Empowerment. The involvement of experts by experience in hospitals. 2018. URL: https://lirias.kuleuven.be/2320741?limo=0 [accessed 2020-08-06]
14. Castro EM, Van Regenmortel T, Vanhaecht K, Sermeus W, Van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: A concept analysis based on a literature review. Patient Educ Couns 2016 Dec;99(12):1923-1939. [doi: 10.1016/j.pec.2016.07.026] [Medline: 27450481]
15. Cerezo PG, Juvé-Udina M, Delgado-Hito P. Concepts and measures of patient empowerment: a comprehensive review. Rev. esc. enferm. USP 2016 Aug;50(4):667-674. [doi: 10.1590/s0080-623420160000500018]
16. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
17. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
18. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. 2014 Presented at: 18th International Conference on Evaluation and Assessment in Software Engineering; May 13-14, 2014; London, England p. 1-10.
19. Karni L, Memedi M, Klein G. Targeting patient empowerment via ICT Interventions: An ICT-specific analytical framework. 2019 Presented at: 25th Americas Conference on Information Systems; August 15-17, 2019; Cancun, Mexico. [doi: 10.2196/17459]
20. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
21. Priest H, Roberts P, Woods L. An overview of three different approaches to the interpretation of qualitative data. Part 1: Theoretical issues. Nurse Res 2002 Oct;10(1):30-42. [doi: 10.7748/nr2002.10.10.1.30.c5877] [Medline: 12405004]

22. World Health Organization. User empowerment in mental health – a statement by the WHO Regional Office for Europe. 2010. URL: https://apps.who.int/iris/bitstream/handle/10665/107275/E93430.pdf [accessed 2020-08-06]
23. Agner J, Braun KL. Patient empowerment: A critique of individualism and systematic review of patient perspectives. Patient Educ Couns 2018 Dec;101(12):2054-2064. [doi: 10.1016/j.pec.2018.07.026] [Medline: 30143254]
24. Bailo L, Guiddi P, Vergani L, Marton G, Pravettoni G. The patient perspective: investigating patient empowerment enablers and barriers within the oncological care process. Ecancermedicalscience 2019 Mar 28;13:912 [FREE Full text] [doi: 10.3332/ecancer.2019.912] [Medline: 31123495]
25. De Angelis G, Davies B, King J, McEwan J, Cavallo S, Loew L, et al. Information and Communication Technologies for the Dissemination of Clinical Practice Guidelines to Health Professionals: A Systematic Review. JMIR Med Educ 2016 Nov 30;2(2):e16 [FREE Full text] [doi: 10.2196/mededu.6288] [Medline: 27903488]
26. World Health Organization. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening: Web Supplement 2: Summary of Findings and GRADE Tables. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1 [accessed 2020-08-06]
27. Asbring P, Närvänen AL. Patient power and control: a study of women with uncertain illness trajectories. Qual Health Res 2004 Feb;14(2):226-240. [doi: 10.1177/1049732303260682] [Medline: 14768459]
28. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York, NY: Springer Publishing Company; 1984.
29. Kärner Köhler A, Tingström P, Jaarsma T, Nilsson S. Patient empowerment and general self-efficacy in patients with coronary heart disease: a cross-sectional study. BMC Fam Pract 2018 May 30;19(1):76 [FREE Full text] [doi: 10.1186/s12875-018-0749-y] [Medline: 29843619]
30. Brady TJ, Murphy L, O'Colmain BJ, Beauchesne D, Daniels B, Greenberg M, et al. A meta-analysis of health status, health behaviors, and health care utilization outcomes of the Chronic Disease Self-Management Program. Prev Chronic Dis 2013 Jan 17;10:120112 [FREE Full text] [doi: 10.5888/pcd10.120112] [Medline: 23327828]
31. International Organization for Standardization. ISO 704:2009 Terminology work — Principles and methods. 2009. URL: https://www.iso.org/obp/ui/#iso:std:iso:704:ed-3:v1:en [accessed 2020-08-10]
32. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec 9;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
33. Ware JE. SF-36 health survey update. Spine (Phila Pa 1976) 2000 Dec 15;25(24):3130-3139. [doi: 10.1097/00007632-200012150-00008] [Medline: 11124729]
34. Jenkinson C, Fitzpatrick R, Peto V, Greenhall R, Hyman N. The PDQ-8: Development and validation of a short-form parkinson's disease questionnaire. Psychology & Health 1997 Dec;12(6):805-814. [doi: 10.1080/08870449708406741]
35. Thompson AG, Suñol R. Expectations as determinants of patient satisfaction: concepts, theory and evidence. Int J Qual Health Care 1995 Jun 01;7(2):127-141. [doi: 10.1093/intqhc/7.2.127] [Medline: 7655809]
36. Marshall GN, Hays RD. The patient satisfaction questionnaire short-form (PSQ-18). Santa Monica, CA: RAND; 1994. URL: https://www.rand.org/content/dam/rand/pubs/papers/2006/P7865.pdf [accessed 2020-08-06]
37. Sjetne IS, Bjertnaes OA, Olsen RV, Iversen HH, Bukholm G. The Generic Short Patient Experiences Questionnaire (GS-PEQ): identification of core items from a survey in Norway. BMC Health Serv Res 2011 Apr 21;11(1):88 [FREE Full text] [doi: 10.1186/1472-6963-11-88] [Medline: 21510871]
38. Hawthorne G, Sansoni J, Hayes L, Marosszeky N, Sansoni E. Measuring patient satisfaction with health care treatment using the Short Assessment of Patient Satisfaction measure delivered superior and robust satisfaction estimates. J Clin Epidemiol 2014 May;67(5):527-537. [doi: 10.1016/j.jclinepi.2013.12.010] [Medline: 24698296]
39. Xu RH, Cheung AW, Wong EL. Development and validation of an instrument to measure patient engagement in Hong Kong Special Administrative Region, China. PPA 2018 Sep;Volume 12:1667-1675. [doi: 10.2147/ppa.s171026]
40. Graffigna G, Barello S, Bonanomi A, Lozza E. Measuring patient engagement: development and psychometric properties of the Patient Health Engagement (PHE) Scale. Front Psychol 2015 Mar 27;6:274 [FREE Full text] [doi: 10.3389/fpsyg.2015.00274] [Medline: 25870566]
41. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
42. Traore L, Assele-Kama A, Keung S, Karni L, Klein GO, Lilja M, Jaulent. User-Centered Design of the C3-Cloud Platform for Elderly with Multiple Diseases - Functional Requirements and Application Testing. Stud Health Technol Inform 2019 Aug 21;264:843-847. [doi: 10.3233/SHTI190342] [Medline: 31438043]
43. Memedi M, Tshering G, Fogelberg M, Jusufi I, Kolkowska E, Klein G. An Interface for IoT: Feeding Back Health-Related Data to Parkinson's Disease Patients. JSAN 2018 Mar 12;7(1):14. [doi: 10.3390/jsan7010014]
44. Merendi M, Lindqvist J, Tunedal T, Duvayker A. A study on pre-adoption of a self-management application by Parkinson's disease patients. In: IT Implementation, Adoption, and Use. 2018 Presented at: International Conference on Information Systems (ICIS); December 13-16, 2018; San Francisco, CA. [doi: 10.17705/4icis]
45. Kosciulek JF. Structural Equation Model of the Consumer-Directed Theory of Empowerment in a Vocational Rehabilitation Context. Rehabilitation Counseling Bulletin 2016 Aug 13;49(1):40-49. [doi: 10.1177/0034355205049001050]

46. Prigge J, Dietz B, Homburg C, Hoyer WD, Burton JL. Patient empowerment: A cross-disease exploration of antecedents and consequences. International Journal of Research in Marketing 2015 Dec;32(4):375-386. [doi: 10.1016/j.ijresmar.2015.05.009]
47. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
48. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
49. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
50. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]

Abbreviations

BIT: Behavioral Interventional Technology
eHealth: electronic health
FGD: focus group discussion
GS-PEQ: Generic Short Patient Experiences Questionnaire
HR-QoL: health-related quality of life
ICT: information and communications technology
ISO: International Organization for Standardization
ICT4PEM: ICT for Patient Empowerment Model
PDQ-8: Parkinson's Disease Questionnaire-8
PEM: patient empowerment model
PEP: Patient Empowerment Platform
PSQ-SF: Patient Satisfaction Questionnaire Short Form.
QoL: quality of life
SAPS: Short Assessment of Patient Satisfaction
SF-36: Short-Form 36
WHO: World Health Organization

Edited by G Eysenbach; submitted 14.01.20; peer-reviewed by L Alpay, A Muñoz, C Schüttler; comments to author 10.03.20; revised version received 04.05.20; accepted 03.06.20; published 26.08.20.

Please cite as:
Karni L, Dalal K, Memedi M, Kalra D, Klein GO
Information and Communications Technology–Based Interventions Targeting Patient Empowerment: Framework Development
J Med Internet Res 2020;22(8):e17459
URL: http://www.jmir.org/2020/8/e17459/
doi:10.2196/17459
PMID:32845245

Original Paper

Capturing Relevant Patient Data in Clinical Encounters Through Integration of an Electronic Patient-Reported Outcome System Into Routine Primary Care in a Boston Community Health Center: Development and Implementation Study

Stephanie Loo[1], MSc; Chris Grasso[1], MPH; Jessica Glushkina[1], MPH; Justin McReynolds[2], MSc; William Lober[2], MD, MS; Heidi Crane[3], MD, MPH; Kenneth H Mayer[1,4,5], MD

[1]The Fenway Institute, Boston, MA, United States

[2]Clinical Informatics Research Group, Biobehavioral Nursing and Health Informatics, University of Washington, Seattle, WA, United States

[3]Division of Allergy and Infectious Diseases, Department of Medicine, University of Washington, Seattle, WA, United States

[4]Harvard Medical School, Boston, MA, United States

[5]HIV Prevention Research, Beth Israel Deaconess Hospital, Boston, MA, United States

Corresponding Author:
Stephanie Loo, MSc
The Fenway Institute
1340 Boylston St
Boston, MA, 02215
United States
Phone: 1 857 313 6582
Email: sloo@fenwayhealth.org

Abstract

Background: Electronic patient-reported outcome (ePRO) systems can improve health outcomes by detecting health issues or risk behaviors that may be missed when relying on provider elicitation.

Objective: This study aimed to implement an ePRO system that administers key health questionnaires in an urban community health center in Boston, Massachusetts.

Methods: An ePRO system that administers key health questionnaires was implemented in an urban community health center in Boston, Massachusetts. The system was integrated with the electronic health record so that medical providers could review and adjudicate patient responses in real-time during the course of the patient visit. This implementation project was accomplished through careful examination of clinical workflows and a graduated rollout process that was mindful of patient and clinical staff time and burden. Patients responded to questionnaires using a tablet at the beginning of their visit.

Results: Our program demonstrates that implementation of an ePRO system in a primary care setting is feasible, allowing for facilitation of patient-provider communication and care. Other community health centers can learn from our model in terms of applying technological innovation to streamline clinical processes and improve patient care.

Conclusions: Our program demonstrates that implementation of an ePRO system in a primary care setting is feasible, allowing for facilitation of patient-provider communication and care. Other community health centers can learn from our model for application of technological innovation to streamline clinical processes and improve patient care.

(J Med Internet Res 2020;22(8):e16778) doi:10.2196/16778

KEYWORDS

information technology in health; primary care; technology adoption; technology diffusion

Introduction

Electronic patient-reported outcome (ePRO) systems can improve health outcomes by detecting health issues or risk behaviors that may be missed when relying on provider elicitation [1]. The use of computerized assessments has the potential to further improve patient-provider communication and overall satisfaction though systematic data collection [2].

The ePRO system was developed for research in 2007, through the Center for AIDS Research Network of Integrated Clinical Systems' collection of HIV-specific PROs [3,4]. Fenway Health (hereafter, referred to as "Fenway") is a national leader in HIV care, research, and culturally responsive care to lesbian, gay, bisexual, transgender, and queer (LGBTQ) patients [5]. Fenway serves a diverse group of around 30,000 patients, more than 17,000 of whom identify as LGBTQ, over 2000 of whom are persons living with HIV, and 30% comprise racial and ethnic minorities. The ePRO integration project was implemented across all primary care clinic sites of Fenway, an urban federally qualified community health center. The pilot project began in 2013, with full implementation and rollout occurring across three primary care clinic sites at Fenway from 2014 to 2015.

An average primary care patient is due for 25 different services at the time of the visit [6], resulting in increased paperwork and data entry burden for staff [7]. Innovations such as the ePRO system at Fenway increase clinical efficiency and reduce both patient and provider burden during clinic visits, allowing providers to focus on the most salient aspects of the patient's reason of visit. An ePRO system that administers key health questionnaires was implemented at Fenway and integrated with the electronic health record (EHR), allowing medical providers to review and adjudicate patient responses in real-time during medical visits. The goal of this ePRO project was to implement and evaluate the ePRO interface for all patients accessing primary care at a Boston community health center.

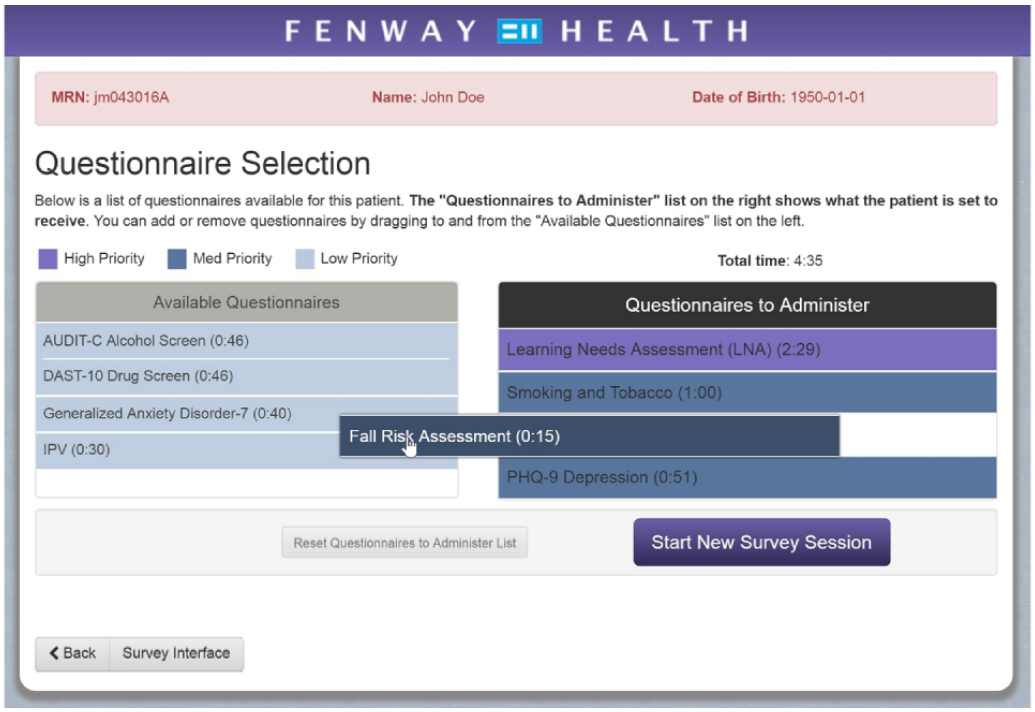
Methods

Development of Technology and Clinical Processes

We developed a graphical user interface and underlying algorithm to populate a set of assessments based on how recently the previous assessment was completed, if ever; clinical priority (eg, repeat depression screen for patient with a depression diagnosis); active diagnoses; and estimated time to complete the assessment.

Patients received a tablet device containing validated surveys focusing on key health domains such as Patient Health Questionnaire-9 (PHQ-9) depression scale and the Alcohol Use Disorders Identification Test (AUDIT) during their medical visit. Fenway customized the ePRO platform to alert providers if patients reported suicidal ideation (a positive response to Question 9 of the PHQ-9 depression screen [8]). Fenway began with 4 assessments and eventually expanded to 10, addressing behavioral health, substance use, and fall risk. As more assessments were added to the ePRO interface, the underlying algorithm for determining priority and frequency of each assessment was refined. Features of the selection menu included automated color coding and ranking of prioritized assessments, limiting total assessments to a 5-minute timeframe. In addition, clinical staff could select any or all available assessments for administration, depending on clinical need and available time for the patient to take the assessments. Estimated time to complete designations was generated for each survey based upon the total number of potential questions within each survey and running mock-survey sessions by the program team. The design of the assessments included multiple-choice, check box, fill-in-the-blank, and drop-down questions based on prior responses (Figure 1). The available surveys were, namely, learning needs assessment, PHQ-9/PHQ-9 modified for Adolescents (PHQ-A), smoking and tobacco, fall risk assessment, intimate partner violence, AUDIT-C alcohol screen, Drug Abuse Screening Test-10 (DAST-10), Generalized Anxiety Disorder-7 (GAD-7), Edinburgh postpartum screen, patient portal sign-up.

Figure 1. Fenway's electronic patient-reported outcome (ePRO) survey interface.

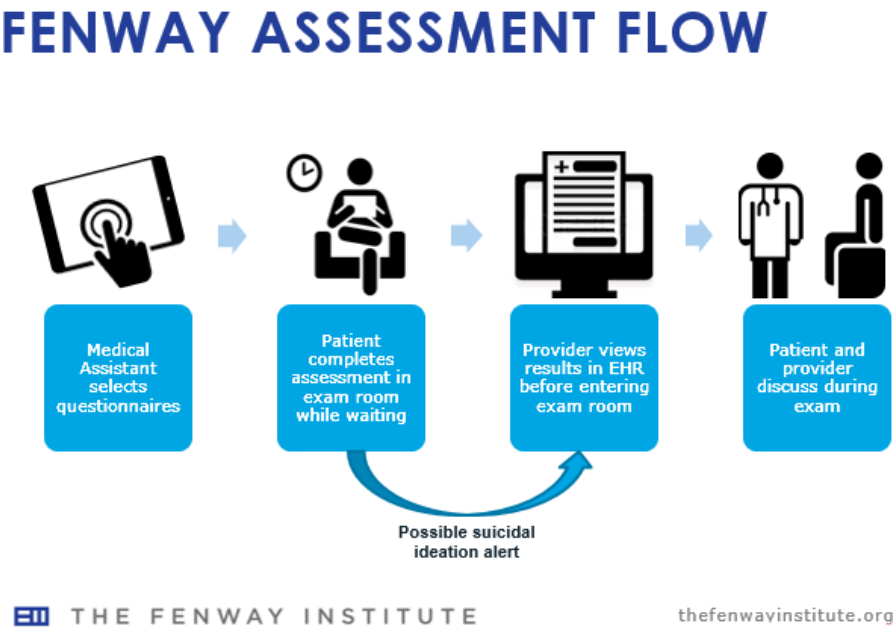


The platform was configured to determine which surveys were due for each patient, including logic to create skip patterns depending on a patient's response. Using Health Level 7 standards, the platform interfaced with Fenway's EHR (Centricity Practice Solutions). As such, patient responses automatically uploaded as structured data into their medical record for provider review, decreasing data entry errors and improving workflow efficiency and accuracy. The data were also sent to the EHR as a summary report in PDF format with score interpretations, making review of the data simple. Time from patient completion of assessments to results appearing in their medical record was approximately 2 minutes.

Patients received a tablet device containing validated surveys at the beginning of their medical visit prior to their medical provider entering the room. Patients filled out the assessments in the exam room after receiving the tablet from a medical assistant. The exam room was chosen over the waiting room area due to patients' shorter wait times as well as to allow for privacy should a patient have any questions for the medical assistant or provider. The platform was configured to determine which assessments were due for each patient (Figure 2), allowing for a dynamic system that would customize order and prioritize assessments to individual patient diagnoses and needs.

The suggested time limit of 5 minutes was generated based on the discussion with clinical leadership during the preproject implementation. However, the interface allowed medical staff to select and choose preferred assessments tailored to individual patient need and available waiting time. For example, if a patient arrived and was roomed early for their appointment, the medical assistant could select more assessments beyond the 5-minute time limit for the patient to complete. If the provider could see the patient earlier, the medical assistant could select one or two prioritized surveys that would take less time for the patient to complete and not impede upon their clinical visit time.

Figure 2. Visualization of electronic patient-reported outcome (ePRO) workflow in Fenway's medical department.



Evaluation of ePRO Implementation

We conducted two brief mixed methods evaluations during the implementation to ascertain provider and patient feedback on the ePRO system. A paper-based survey was sent out to providers and medical assistants in early 2016 to ask how the program was working for them. A focus group with Fenway primary care patients was held in July 2016 to generate patient feedback on the ePRO program. As these activities were part of clinical quality improvement activities, Institutional Review Board approval was not necessary.

Costs

This work was made possible through a one-time 2-year grant provided by Neighborhood Health Plan and the Partnership for Community Health's *Excellence and Innovation* program. This funding primarily covered costs of a full-time program manager to design; implement; and oversee the ePRO program as well as those of equipment needs, including tablets, storage lockers, protective cases, and sanitation wipes for cleaning.

Results

Development of Technology and Clinical Processes

Piloting began in 2013, with implementation and rollout occurring across three primary care clinical sites from 2014 to 2015. The ePRO system was initially piloted with 5 medical assistants serving 3 primary care providers, and eventually expanded to 18 medical assistants and 25 primary care providers across three clinical sites in the first year of expansion within the Fenway medical department. Completion rates of ePRO sessions grew significantly over time, from 2428 completed sessions in 2014 to 19,650 completed sessions in 2018. We noted an increase in the total number of ePRO sessions as well as the number of ePROs taken by the same patient, demonstrating increased and repetitive use of the interface as patients return for follow-up care. From 2016 to 2017, the

percentage of patients who took and completed at least one assessment using the ePRO system increased from 66% to 74%. In 2018, via ePRO, 41% of patients reported mild-to-severe depression, 35% reported mild-to-severe anxiety, 35% reported problem alcohol use, 4.4% reported a positive DAST score, and 10% reported current or some-day tobacco use (Table 1). In 2018, there were 300 suicide ideations alerts from the ePRO platform, wherein a suicide ideation response of "Nearly Every Day" comprised 1.5% of total ePRO sessions in the same year.

Table 1. Electronic patient-reported outcome (ePRO) prevalence data for 2016-2018.

Condition	Instrument	Prevalence, n/N (%)		
		2016	2017	2018
Mild-to-severe depression	PHQ-9[a]	6644/13899 (47.80)	8469/16237 (52.16)	9022/15857 (56.90)
Mild-to-severe anxiety	GAD-7[b]	2748/6813 (40.33)	4870/10542 (46.20)	7664/14456 (53.02)
Problem alcohol use (high risk - likely addiction)	AUDIT[c]	115/5365 (2.14)	195/8785 (2.22)	340/14345 (2.37)
Moderate-to-severe drug use	DAST-10[d]	178/5438 (3.27)	174/8988 (1.94)	641/14470 (4.43)
Current or some-day smoker	Smoking & Tobacco	994/8736 (11.38)	1333/14164 (9.41)	1527/14992 (10.19)

[a]PHQ-9: Patient Health Questionnaire-9.

[b]GAD-7: Generalized Anxiety Disorder-7.

[c]AUDIT: Alcohol Use Disorders Identification Test.

[d]DAST-10: Drug Abuse Screening Test-10.

Evaluation of ePRO Implementation

Primary care providers, medical assistants, and patients were generally satisfied with the ePRO program. A survey of primary care providers was conducted in early 2016, wherein 75% (27/36) providers responded. A total of 70.4% (19/20) of medical providers agreed/strongly agreed that use of the ePRO system improved access to real-time data in patient medical charts. Similarly, 66.7% (18/27) indicated that using the ePRO system reduced paperwork burden. In a patient focus group (n=8) conducted in July 2016, patients appreciated how the technology facilitated quick responses from providers and saved time during their medical appointments. Patients reported that the program allowed them to answer challenging questions in a nonstigmatized manner, which they found valuable, particularly when meeting a provider with whom they had no prior relationship. Patients noted that the ePRO system made them feel like direct participants in their medical care.

Costs

The grant used to fund staff and equipment purchases assisted with program start-up costs. As the ePRO program expanded and became highly accepted and endorsed by medical staff, associated costs with maintaining the equipment used were absorbed into health center operation budgets. These associated costs were minimal, as they primarily covered purchases of sanitation wipes for tablet cleaning between patients. The program manager who initially oversaw the project was transferred. The overall supervision of the ePRO program remained in place for troubleshooting any issues but the project manager moved on to other projects due to the reduction of needed full-time equivalent (FTE) units following successful implementation of the system.

Discussion

Fenway's ePRO program is highly accepted by both clinical staff and patients. Fenway recently developed other ePRO programs for the behavioral health and patient registration departments following the success of the program in primary care. The patient registration ePRO department uses interactive registration forms to collect PRO data from patients in the waiting room. While this program initially received a grant aimed at implementation of PROs that covered high early start-up costs, Fenway found that these costs were absorbed over time, as equipment needs did not change and initial staff management of ePRO became embedded in clinical workflow. Further, successful implementation of the program resulted in multiple benefits for the health center in saved paperwork and data entry time as well as helping the health center meet PRO-related quality measures and fee-for-service benchmarks. Achieving sustainability for this project required buy-in from health center and departmental leadership, in addition to establishing an ePRO process that did not disrupt clinical workflows.

The primary challenge in implementing the ePRO program centered on overcoming hesitations of primary care providers and medical assistants at implementing a new program within the existing clinical workflows. This reluctance dissipated as clinic teams observed how colleagues participating in the pilot program saved time, as the program eliminated the need to print and hand out health questionnaires before patient visits and to input patient responses in patient charts afterwards. The program manager also met one-on-one with primary care providers and medical assistants to train them on the new system and any changes to clinic forms. This hands-on approach to building rapport combined with frequent solicitation of clinical staff feedback assisted with gaining staff buy-in.

The routine integration of PROs into clinical care can have many potential advantages, including improving patient-provider communication, improving care, and facilitating research. Patient care can be improved by enabling clinicians to address functional problems, mental health problems, or symptomatic conditions that might otherwise be missed. The ePRO system at Fenway proves to be an acceptable, sustainable method of quickly assessing patients, although further work needs to understand the true effects of ePRO on clinical care and improvement in a primary care setting.

Acknowledgments

This work was supported by Neighborhood Health Plan and the Partnership for Community Health's *Excellence and Innovation* funding in addition to building on PRO work funded by AHRQ grant U18HS026154. The authors would like to thank the clinical staff and patients of Fenway Health. Their contributions and feedback were invaluable to the success of this implementation.

Authors' Contributions

SL led the implementation of the project, and drafted the article. CG, WL, HC, and KHM contributed to project conceptualization and design and obtained funding. JM led the technical development of the ePRO system for Fenway. All authors contributed to the interpretation of data, critical revision of the article for important intellectual content, and administrative or technical support.

Conflicts of Interest

None declared.

References

1. Intro to Person-Centered Assessment. Health Measures: Transforming How Health is Measured. 2017. URL: http://www.nihpromis.org/Patients/PROs [accessed 2019-09-23]
2. Rotenstein LS, Huckman RS, Wagle NW. Making Patients and Doctors Happier - The Potential of Patient-Reported Outcomes. N Engl J Med 2017 Oct 05;377(14):1309-1312. [doi: 10.1056/NEJMp1707537] [Medline: 28976860]
3. Crane H, Lober W, Webster E, Harrington R, Crane P, Davis T, et al. Routine collection of patient-reported outcomes in an HIV clinic setting: the first 100 patients. Curr HIV Res 2007 Jan 01;5(1):109-118. [doi: 10.2174/157016207779316369] [Medline: 17266562]
4. Kitahata MM, Rodriguez B, Haubrich R, Boswell S, Mathews WC, Lederman MM, et al. Cohort profile: the Centers for AIDS Research Network of Integrated Clinical Systems. Int J Epidemiol 2008 Oct 08;37(5):948-955 [FREE Full text] [doi: 10.1093/ije/dym231] [Medline: 18263650]
5. Mayer K, Appelbaum J, Rogers T, Lo W, Bradford J, Boswell S. The evolution of the Fenway Community Health model. Am J Public Health 2001 Jun;91(6):892-894. [doi: 10.2105/ajph.91.6.892] [Medline: 11392929]
6. Medder JD, Kahn NB, Susman JL. Risk factors and recommendations for 230 adult primary care patients, based on U.S. Preventive Services Task Force guidelines. Am J Prev Med 1992;8(3):150-153. [Medline: 1633001]
7. Woolhandler S, Himmelstein DU. Administrative work consumes one-sixth of U.S. physicians' working hours and lowers their career satisfaction. Int J Health Serv 2014;44(4):635-642. [doi: 10.2190/HS.44.4.a] [Medline: 25626223]
8. Kroenke K, Spitzer RL. The PHQ-9: A New Depression Diagnostic and Severity Measure. Psychiatric Annals 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
DAST: Drug Abuse Screening Test
EHR: electronic health record
ePRO: electronic patient-reported outcome
FTE: full-time equivalent
GAD-7: Generalized Anxiety Disorder-7
LGBTQ: lesbian, gay, bisexual, transgender, and queer
PHQ-9: Patient Health Questionnaire-9
PHQ-A: PHQ-9 modified for Adolescents

Edited by G Eysenbach; submitted 30.10.19; peer-reviewed by D Moores, C Cannon; comments to author 22.02.20; revised version received 07.06.20; accepted 13.06.20; published 19.08.20.

Please cite as:
Loo S, Grasso C, Glushkina J, McReynolds J, Lober W, Crane H, Mayer KH
Capturing Relevant Patient Data in Clinical Encounters Through Integration of an Electronic Patient-Reported Outcome System Into Routine Primary Care in a Boston Community Health Center: Development and Implementation Study
J Med Internet Res 2020;22(8):e16778
URL: http://www.jmir.org/2020/8/e16778/
doi:10.2196/16778
PMID:32554372

Original Paper

Communication Behavior Changes Between Patients With Diabetes and Healthcare Providers Over 9 Years: Retrospective Cohort Study

Arriel Benis[1,2], PhD; Refael Barak Barkan[3], MD, PhD; Tomer Sela[4], BA, MBA; Nissim Harel[3], PhD

[1]Faculty of Technology Management, Holon Institute of Technology, Holon, Israel
[2]Clalit Research Institute, Clalit Health Services, Tel-Aviv, Israel
[3]Faculty of Sciences, Holon Institute of Technology, Holon, Israel
[4]Online Division, Clalit Health Services, Tel-Aviv, Israel

Corresponding Author:
Arriel Benis, PhD
Faculty of Technology Management
Holon Institute of Technology
Golomb St 52
POB 305
Holon
Israel
Phone: 972 3 5026892
Email: arrielb@hit.ac.il

Abstract

Background: Health organizations and patients interact over different communication channels and are harnessing digital communications for this purpose. Assisting health organizations to improve, adapt, and introduce new patient–health care practitioner communication channels (such as patient portals, mobile apps, and text messaging) enhances health care services access.

Objective: This retrospective data study aims to assist health care administrators and policy makers to improve and personalize communication between patients and health care professionals by expanding the capabilities of current communication channels and introducing new ones. Our main hypothesis is that patient follow-up and clinical outcomes are influenced by their preferred communication channels with the health care organization.

Methods: This study analyzes data stored in electronic medical records and logs documenting access to various communication channels between patients and a health organization (Clalit Health Services, Israel). Data were collected between 2008 and 2016 from records of 311,168 patients diagnosed with diabetes, aged 21 years and over, members of Clalit at least since 2007, and still alive in 2016. The analysis consisted of characterizing the use profiles of communication channels over time and used clustering for discretization purposes and patient profile building and then a hierarchical clustering and heatmaps to visualize the different communication profiles.

Results: A total of 13 profiles of patients were identified and characterized. We have shown how the communication channels provided by the health organization influence the communication behavior of patients. We observed how different patients respond differently to technological means of communication and change or don't change their communication patterns with the health care organization based on the communication channels available to them.

Conclusions: Identifying the channels of communication within the health organization and which are preferred by each patient creates an opportunity to convey messages adapted to the patient in the most appropriate way. The greater the likelihood that the therapeutic message is received by the patient, the greater the patient's response and proactiveness to the treatment will be.

International Registered Report Identifier (IRRID): RR2-10.2196/10734

(J Med Internet Res 2020;22(8):e17186) doi:10.2196/17186

KEYWORDS

population characteristics; eHealth; mHealth; consumer health informatics; delivery of health care; machine learning; clustering; quality of health care; point-of-care systems; physician-patient relations

Introduction

Background

Communications between patients and health care professionals are based on a range of communication channels [1-10] and influenced by cultural factors [11-13]. Traditional channels supporting these interactions are face-to-face visits and phone calls. Health management organizations (HMOs) are capitalizing on the digital revolution [8,14] and innovating and providing patients with new digital tools [15]. Their goal is to provide patients with alternative ways for asking, getting, and sharing health-related information and knowledge [2-6,9,10,16].

Interactions between patients and health care professionals in an HMO must be analyzed over time to better understand the potential impacts of technological changes. Data mining and machine learning methodologies are used in the analysis of a large amount of data. Several techniques can be used to define or redefine clusters of patients based on sociodemographics and biological and clinical data [17,18]. We are not aware of an attempt to cluster patients based on communication, sociodemographic, and bioclinical characteristics, let alone at a large scale involving data from hundreds of thousands of patients collected for almost a decade. In this paper, we are disclosing the results of this kind of approach [19].

Aims and Objectives

This retrospective data study aims at assisting health care administrators in defining and developing new communication channels and policy makers in improving and personalizing communication between patients and health care professionals (eg, physicians and nurses). By expanding the capabilities of currently available communication channels and introducing new channels, we hope to help policy makers enhance the accessibility of health care professionals and organizations and improve the quality of patient follow-up and treatment adherence and the overall patient experience with HMO services [13,19-22].

This work characterizes the use profiles of chronic patients with the communication channels available at Clalit Health Services, a large HMO in Israel, between 2008 and 2016. The use profiles are then associated with sociodemographic and medical patient profiles.

The leading objective of this analysis is to propose new ways to promote the use of the most appropriate communication channels based on the patient profile. An additional objective is to recommend sociological and technological ways that should be developed for increasing the quality and effectiveness of the patient–health care professional communication and interaction.

Hypotheses

This retrospective research is led by 3 hypotheses:

- Preferred communication channels of a patient with the health care professional in an HMO influence follow-up and clinical outcomes.
- Adoption of a new communication channel by a patient is affected by their sociodemographic and clinical profile.
- Introducing a new communication channel impacts the use of existing communication channels.

This research focuses on quantifying these behaviors. The goal is to identify sociodemographic and bioclinical attributes affecting engagement with newly launched communication channels. This research characterizes changes in the use of existing communication channels once a new communication channel is introduced.

Methods

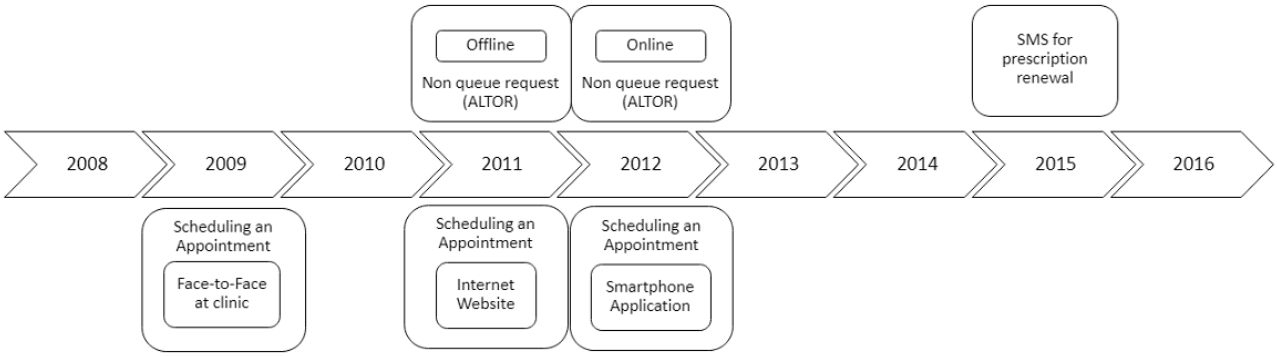
The study design including details about material and methods has been described and published elsewhere [19]. Ethical approval for the study was granted by the Clalit ethical committee.

Material

Data were extracted from Clalit electronic medical records (EMRs), which include documented access to various communication channels between patients and Clalit. Clalit is the largest Israeli HMO, with 4.53 million insured members (53% of the Israeli population) in 2016. Since 1998, Clalit's EMRs have been stored in a data warehouse [19,23,24].

The period of time investigated in this research allowed us to analyze the launch of new communication channels such as a website, mobile apps, and text messaging (short message service [SMS]) system (Figure 1) and identify communication behavior changes as functions of time and the introduction of new communication channels.

Figure 1. Communication channel introduction over time.



The cohort consisted of patients aged 21 years and older, diagnosed with diabetes, members of Clalit for at least 1 year before 2008, and still alive in 2016 [25-27].

Analysis Process

Overview

This research used the knowledge discovery in databases (KDD) framework [28-30]. Our study analyzes communication channel use over a 9-year period wherein Clalit introduced and changed the methods of interactions between patients, health care professionals, and administrative staff. We identified the sociodemographic and bioclinical characteristics for each communication profile and qualitatively evaluated the influence of the profile on patient engagement and follow-up quality.

We ran 1-dimensional and multidimensional statistical tools and different data mining algorithms, which were used during the data cleansing step [31,32]. The data extraction, preprocessing, data mining, and information visualization are briefly described below. Details have been published elsewhere [19].

Data Extraction

Data extracted from the Clalit data warehouse for each patient included sociodemographic [19] and bioclinical [25,26,33-37] data and contacts with the HMO using communication channels.

Data Preprocessing

Cleansing of extracted data reduced noise by detecting and removing or correcting outliers [38,39]. An outlier is a data measurement that is inconsistent with other historical measurement data of the same individual. For some measurements (eg, BMI), specific algorithms have been developed in-house by Clalit. In the absence of these algorithms, statistical approaches and machine learning algorithms were used [40-44].

Several machine learning algorithms require data reformulation to support data categorization or grouping numerical, categorical, or textual data [41,45-48]. For some attributes that don't have predefined scales, we used the k-means clustering algorithm to discretize the data into 6 groups: very small, small, small-to-moderate, moderate, moderate-to-large, and large. The cluster bounds were validated by a domain expert (Table 1).

Table 1. Gradient reformulation and ranges of values related to each resource consumption level.

Characteristic	Gradient reformulation						
	N/A[a]	No (very small)	Small	Small to moderate	Moderate	Moderate to large	Large
Contact with health care provider							
Physician consultation	N/A	0	1-7	8-19	20-28	29-51	52+
Nurse consultation	N/A	0	1-3	4-5	6-8	9-14	15+
Hospitalization	N/A	0	1	2	3	4-5	6+
ED[b] visit	N/A	0	1	2	3	4	5+
Scheduling an appointment							
Face-to-face at clinic	N/A	0	1	2	3-4	5-8	9+
Call to clinic or call center	N/A	0	1-2	3-5	6-9	10-15	16+
Smartphone app	N/A	0	1	2	3-4	5-8	9+
Internet website	N/A	0	1	2-3	4-6	7-10	11+
Nonqueue request							
Online	N/A	0	1-3	4-12	13-24	25-45	46+
Offline that must be done online	N/A	0	1-17	18-44	45-74	75-113	114+
Offline	N/A	0	1-17	18-44	45-74	75-113	114+
Pharmacy							
Overall recorded visits	N/A	0	1-4	5-12	13-26	27-42	43+
SMS[c] for prescription renewal							
Proposition sent by HMO[d] to patient	N/A	0	1	2	3	4	5+
Approval sent by patient to HMO	N/A	0	1	2	3	4	5+

[a]N/A: not applicable.

[b]ED: emergency department.

[c]SMS: short message service.

[d]HMO: health management organization.

Data Mining and Information Visualization

As we don't have prior knowledge on communication channel use, we used unsupervised learning algorithms, mainly k-means and hierarchical clustering [47-56] combined with the Ray-Turi criterion [49].

To investigate communication patterns over time, we built heatmaps for each year between 2008 and 2016 based on the previously generated hierarchical clustering of 2016 data. Furthermore, we concatenated the communication profile of each discovered patients' clusters over the years [56]. This visualization helps identify changes in communication profiles for each cluster.

All computations described above were performed using R 3.3.1 (R Foundation for Statistical Computing) with the following packages: data.table [57] (computing efficiency given the large data size), cluster [58] (k-means and hierarchical clustering), and gplots [59] (drawing heatmaps and hierarchical clustering dendrograms).

Results

Overview

A total of 311,168 individuals were included in the study cohort. As means of communication with health care professionals have changed throughout the research period, we chose 2016 as the base year to which communication behavior is compared because during this year, health care customers were offered the most diverse communication channels. Applying the methodology described above to the 2016 data revealed 13 communication profiles. The resulting heatmap is presented in Figure 2. Two main types of communication behaviors are evident from the figure. The first main cluster consists of 6 communication profiles describing low-to-moderate health care services consumption. The second one consists of 7 communication profiles describing moderate-to-high health care services consumption.

All differences between the sociodemographic, biological, clinical (Tables 2 and 3), and communication characteristics between the overall population and each one of the clusters were statistically significant (P<.001).

Figure 2. Communication patterns in 2016 of 311,168 patients with diabetes (members of the Clalit Health Services). HMO: health management organization.

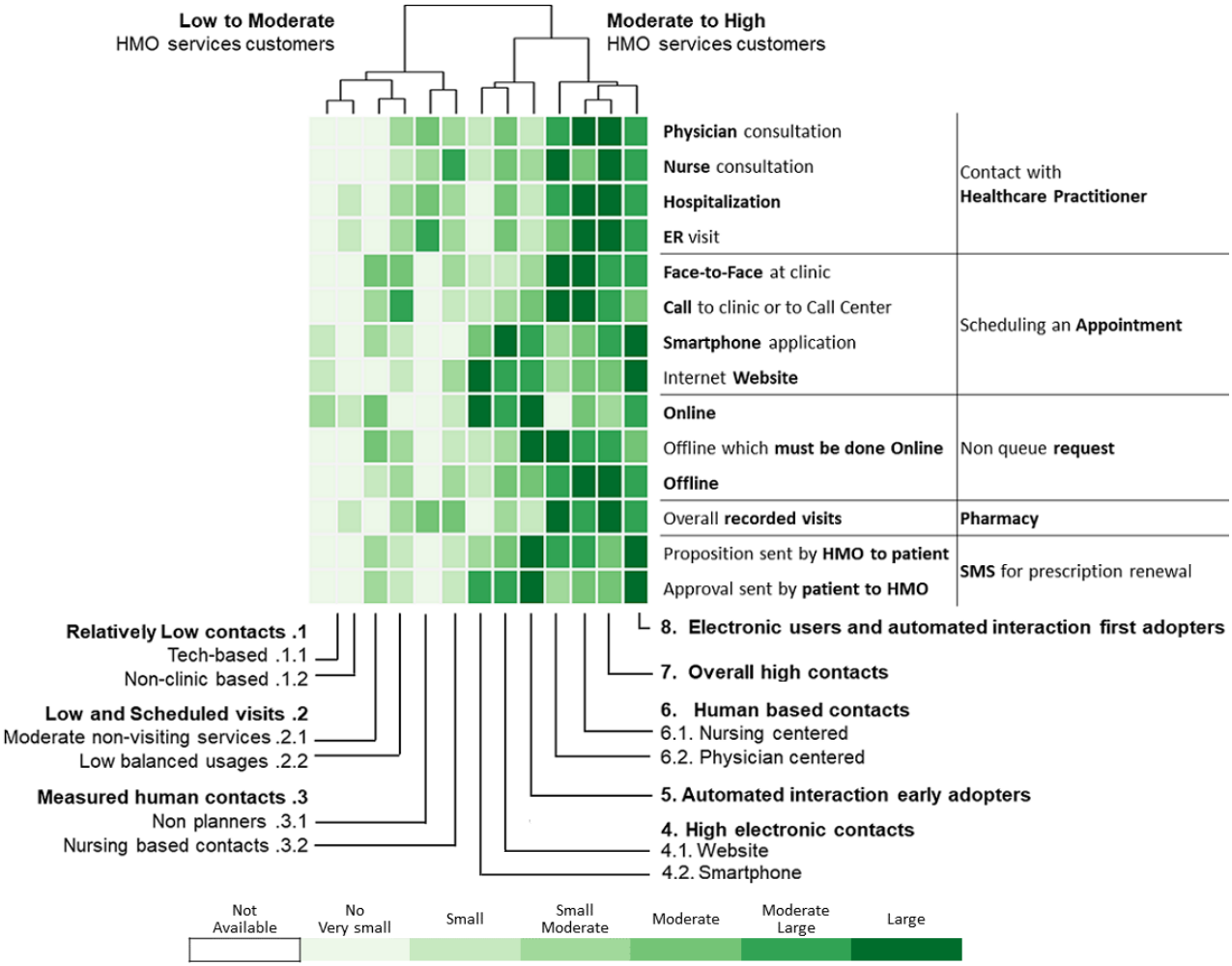


Table 2. Sociodemographic, clinical, and biological measurements data summary for patients with diabetes, in 2016, having a low-to-moderate health care services consumption.

Characteristic	Overall (n=311,168)	Relatively low contacts		Low and scheduled visits		Measured human contacts	
		Tech-based (n=35,719)	Non–clinic-based (n=34,060)	Moderate nonvisiting services (n=49,540)	Low balance use (n=41,735)	Nonplanners (n=22,275)	Nursing-based contacts (n=29,197)
Gender, n (%)							
Female	156,269 (50.2)	15,669 (43.9)	17,173 (50.4)	24,609 (49.7)	22,525 (54.0)	11,264 (50.6)	15,463 (53.0)
Male	154,899 (49.8)	20,050 (56.1)	16,887 (49.6)	24,931 (50.3)	19,210 (46.0)	11,011 (49.4)	13,734 (47.0)
Age in years, median (IQR[a])	68 (60-77)	60 (51-68)	69 (61-79)	67 (60-76)	70 (62-78)	68 (60-76)	70 (62-79)
Immigrant, n (%)							
No	152,533 (49.0)	23,602 (66.1)	18,712 (54.9)	21,710 (43.8)	15,921 (38.1)	14,504 (65.1)	16,085 (55.1)
Yes	158,635 (51.0)	12,117 (33.9)	15,348 (45.1)	27,830 (56.2)	25,814 (61.9)	7771 (34.9)	13,112 (44.9)
Ethnicity, n (%)							
General	242,022 (77.8)	24,232 (67.8)	23,486 (69.0)	41,613 (84.0)	34,218 (82.0)	12,456 (55.9)	19,899 (68.2)
Arab	60,619 (19.5)	10,534 (29.5)	9884 (29.0)	6140 (12.4)	6015 (14.4)	9441 (42.4)	8599 (29.5)
Ultra-Orthodox	8527 (2.7)	953 (2.7)	690 (2.0)	1787 (3.6)	1502 (3.6)	378 (1.7)	699 (2.4)
SES[b], n (%)							
High	97,556 (31.4)	10,444 (29.2)	10,740 (31.5)	15,873 (32.0)	11,897 (28.5)	6111 (27.4)	7304 (25.0)
Medium	126,057 (40.5)	12,458 (34.9)	11,629 (34.1)	22,136 (44.7)	18,920 (45.3)	6016 (27.0)	10,842 (37.1)
Low	83,677 (26.9)	12,479 (34.9)	11,238 (33.0)	11,138 (22.5)	10,568 (25.3)	9669 (43.4)	10,158 (34.8)
N/A[c]	3878 (1.2)	338 (0.9)	453 (1.3)	393 (0.8)	350 (0.8)	479 (2.2)	893 (3.1)
BMI, n (%)							
Obese	122,984 (39.5)	10,395 (29.1)	11,453 (33.6)	19,181 (38.7)	17,280 (41.4)	9830 (44.1)	13,210 (45.2)
Overweight	107,793 (34.6)	10,125 (28.3)	10,406 (30.6)	17,740 (35.8)	15,549 (37.3)	7845 (35.2)	10,355 (35.5)
Normal	47,193 (15.2)	4635 (13.0)	4416 (13.0)	7760 (15.7)	6748 (16.2)	2985 (13.4)	4481 (15.3)
Underweight	1255 (0.4)	124 (0.3)	104 (0.3)	192 (0.4)	162 (0.4)	76 (0.3)	146 (0.5)
Unavailable	31,943 (10.3)	10,440 (29.2)	7681 (22.6)	4667 (9.4)	1996 (4.8)	1539 (6.9)	1005 (3.4)
Smoking status, n (%)							
Nonsmoker	136,815 (44.0)	16,116 (45.1)	13,359 (39.2)	21,896 (44.2)	19,062 (45.7)	9622 (43.2)	12,928 (44.3)
Past smoker	67,300 (21.6)	6294 (17.6)	6249 (18.3)	10,050 (20.3)	8434 (20.2)	5061 (22.7)	6481 (22.2)
Current smoker	43,190 (13.9)	7064 (19.8)	4259 (12.5)	7649 (15.4)	5546 (13.3)	2959 (13.3)	3765 (12.9)
Unavailable	63,863 (20.5)	6245 (17.5)	10,193 (29.9)	9945 (20.1)	8693 (20.8)	4633 (20.8)	6023 (20.6)
ACG[d], median (IQR)	4 (3-5)	3 (2-4)	3 (3-4)	4 (3-4)	4 (4-5)	4 (4-5)	4 (4-5)
HbA_{1c}[e] (mmol/mol)							
n (%)	27,3491 (87.9)	22,748 (63.3)	27,951 (82.1)	43,651 (88.1)	39,015 (93.5)	20,687 (92.9)	23,249 (93.3)
Mean (SD)	7.19 (1.46)	7.31 (1.73)	7.16 (1.39)	7.22 (1.47)	7.15 (1.41)	7.22 (1.45)	7.39 (1.57)
Cholesterol (mg/dL)							
n (%)	282,583 (90.8)	24,541 (68.7)	29,139 (85.6)	45,069 (91.0)	40,035 (95.9)	21,306 (95.6)	27,832 (95.3)
Mean (SD)	167.78 (41.08)	186.01 (43.7)	167.41 (39.5)	169.09 (40.8)	167.25 (40.05)	166.54 (40.01)	164.40 (40.47)
Adherence, n (%)							
Not treated	65,873 (21.2)	14,213 (39.8)	7078 (20.8)	9126 (18.4)	7246 (17.4)	4004 (18.0)	5475 (18.8)
0%	10,501 (3.4)	2601 (7.3)	729 (2.1)	1731 (3.5)	1254 (3.0)	710 (3.2)	792 (2.7)

Characteristic	Overall (n=311,168)	Relatively low contacts		Low and scheduled visits		Measured human contacts	
		Tech-based (n=35,719)	Non–clinic-based (n=34,060)	Moderate nonvisiting services (n=49,540)	Low balance use (n=41,735)	Nonplanners (n=22,275)	Nursing-based contacts (n=29,197)
1%-19%	11,990 (3.9)	2141 (6.0)	804 (2.4)	1794 (3.6)	1691 (4.1)	1034 (4.6)	1229 (4.2)
20%-39%	17,573 (5.6)	3195 (8.9)	1312 (3.9)	2903 (5.9)	2497 (6.0)	1350 (6.1)	1583 (5.4)
40%-59%	27,107 (8.7)	3502 (9.8)	2538 (7.5)	4823 (9.7)	3987 (9.6)	1907 (8.6)	2411 (8.3)
60%-79%	31,447 (10.1)	3104 (8.7)	3168 (9.3)	5350 (10.8)	4690 (11.2)	2236 (10.0)	3031 (10.4)
≥80%	146,677 (47.1)	6963 (19.5)	18,431 (54.1)	23,813 (48.1)	20,370 (48.8)	11,034 (49.5)	14,676 (50.3)

[a]IQR: interquartile range.

[b]SES: socioeconomic status.

[c]N/A: not applicable.

[d]ACG: adjusted clinical group.

[e]HbA_{1c}: glycated hemoglobin.

Table 3. Sociodemographic, clinical, and biological measurements data summary for patients with diabetes who had a moderate-to-high health care services consumption in 2016.

Characteristic	Overall (n=311,168)	High electronic contacts		Automated interaction early adopters (n=26,290)	Human-based contacts		Overall high contact (n=9736)	Electronic driven interaction (n=14,647)
		Internet website (n=19,277)	Smartphone (n=13,279)		Nursing-centered (n=7276)	Physician-centered (n=8137)		
Gender, n (%)								
Female	156,269 (50.2)	8652 (44.9)	7250 (54.6)	15,034 (57.2)	3939 (54.1)	3815 (46.9)	4649 (47.8)	6227 (42.5)
Male	154,899 (49.8)	10,625 (55.1)	6029 (45.4)	11,256 (42.8)	3337 (45.9)	4322 (53.1)	5087 (52.2)	8420 (57.5)
Age in years, median (IQR[a])	68 (60-77)	68 (61-75)	64 (56-72)	72 (65-79)	70 (61-79)	73 (65-81)	71 (65-79)	65 (59-71)
Immigrant, n (%)								
No	152,533 (49.0)	8499 (44.1)	6645 (50.0)	9489 (36.1)	3230 (44.4)	3349 (41.2)	3556 (36.5)	7231 (50.6)
Yes	158,635 (51.0)	10,778 (55.9)	6634 (50.0)	16,801 (63.9)	4046 (55.6)	4788 (58.8)	6180 (63.5)	7416 (50.6)
Ethnicity, n (%)								
General	242,022 (77.8)	18,498 (96.0)	10,977 (82.7)	21,399 (81.4)	5745 (79.0)	6255 (76.9)	9241 (94.9)	14,003 (95.6)
Arab	60,619 (19.5)	478 (2.5)	2105 (15.9)	3895 (14.8)	1311 (18.0)	1656 (20.4)	313 (3.2)	248 (1.7)
Ultra-Orthodox	8527 (2.7)	301 (1.6)	197 (1.5)	996 (3.8)	220 (3.0)	226 (2.8)	182 (1.9)	396 (2.7)
SES[b], n (%)								
High	97,556 (31.4)	9961 (51.7)	3625 (27.3)	7079 (26.9)	1895 (26.0)	2101 (25.8)	4535 (46.6)	5991 (40.9)
Medium	126,057 (40.5)	7729 (40.1)	5981 (45.0)	12,231 (46.5)	3292 (45.2)	3525 (43.3)	4275 (43.9)	7023 (47.9)
Low	83,677 (26.9)	1430 (7.4)	3511 (26.4)	6683 (25.4)	2041 (28.1)	2417 (29.7)	842 (8.6)	1503 (10.3)
N/A[c]	3878 (1.2)	157 (0.8)	162 (1.2)	297 (1.1)	48 (0.7)	94 (1.2)	84 (0.9)	130 (0.9)
BMI, n (%)								
Obese	122,984 (39.5)	7421 (38.5)	6174 (46.5)	12,154 (46.2)	3128 (43.0)	3336 (41.0)	4071 (41.8)	5351 (36.5)
Overweight	107,793 (34.6)	7299 (37.9)	4512 (34.0)	9374 (35.7)	2514 (34.6)	2800 (34.4)	3749 (38.5)	5525 (37.7)
Normal	47,193 (15.2)	3123 (16.2)	1819 (13.7)	4316 (16.4)	1307 (18.0)	1621 (19.9)	1666 (17.1)	2316 (15.8)
Underweight	1255 (0.4)	48 (0.2)	53 (0.4)	112 (0.4)	55 (0.8)	97 (1.2)	42 (0.4)	44 (0.3)
Unavailable	31,943 (10.3)	1386 (7.2)	721 (5.4)	334 (1.3)	272 (3.7)	283 (3.5)	208 (2.1)	1411 (9.6)
Smoking status, n (%)								
Nonsmoker	136,815 (44.0)	8769 (45.5)	6669 (50.2)	11,955 (45.5)	3034 (41.7)	2951 (36.3)	4131 (42.4)	6323 (43.2)
Past smoker	67,300 (21.6)	5127 (26.6)	3086 (23.2)	5904 (22.5)	1582 (21.7)	2119 (26.0)	2835 (29.1)	4078 (27.8)
Current smoker	43,190 (13.9)	1862 (9.7)	1813 (13.7)	2799 (10.6)	1191 (16.4)	1087 (13.4)	756 (7.8)	2440 (16.7)
Unavailable	63,863 (20.5)	3519 (18.3)	1711 (12.9)	5632 (21.4)	1469 (20.2)	1980 (24.3)	2014 (20.7)	1806 (12.3)
ACG[d], median (IQR)	4 (3-5)	4 (3-5)	4 (4-5)	5 (4-5)	5 (4-5)	5 (5-6)	5 (4-5)	4 (3-5)
HbA_{1c}[e], (mmol/mol)								
n (%)	273,491 (87.9)	17,810 (92.4)	12,387 (93.3)	25,320 (96.3)	6769 (93.0)	7594 (93.3)	9373 (96.3)	12,936 (88.3)
Mean (SD)	7.19 (1.46)	6.86 (1.13)	7.19 (1.46)	7.22 (1.42)	7.24 (1.57)	7.26 (1.60)	6.91 (1.17)	7.15 (1.37)
Cholesterol (mg/dL)								
n (%)	282,583 (90.8)	18,399 (95.4)	12,713 (96.0)	25,849 (98.3)	6973 (95.8)	7875 (96.8)	9586 (98.5)	13,232 (80.3)

Characteristic	Overall (n=311,168)	High electronic contacts		Automated interaction early adopters (n=26,290)	Human-based contacts		Overall high contact (n=9736)	Electronic driven interaction (n=14,647)
		Internet website (n=19,277)	Smartphone (n=13,279)		Nursing-centered (n=7276)	Physician-centered (n=8137)		
Mean (SD)	167.78 (41.08)	165.51 (39.4)	166.77 (40.54)	161.96 (39.28)	169.36 (44.28)	159.30 (44.77)	160.79 (39.47)	165.85 (40.10)
Adherence, n (%)								
Not treated	65,873 (21.2)	4020 (20.9)	2487 (18.7)	4278 (16.3)	1552 (21.3)	2094 (25.7)	1889 (19.4)	2411 (16.5)
0%	10,501 (3.4)	503 (2.6)	448 (3.4)	617 (2.3)	245 (3.4)	253 (3.1)	239 (2.5)	379 (2.6)
1%-19%	11,990 (3.9)	458 (2.4)	548 (4.1)	941 (3.6)	357 (4.9)	341 (4.2)	269 (2.8)	383 (2.6)
20%-39%	17,573 (5.6)	723 (3.8)	823 (6.2)	1346 (5.1)	489 (6.7)	462 (5.7)	346 (3.6)	544 (3.7)
40%-59%	27,107 (8.7)	1325 (6.9)	1234 (9.3)	2156 (8.2)	697 (9.6)	741 (9.1)	560 (5.8)	1226 (8.4)
60%-79%	31,447 (10.1)	1711 (8.9)	1447 (10.9)	2712 (10.3)	849 (11.7)	819 (10.1)	804 (8.3)	1526 (10.4)
≥80%	146,677 (47.1)	10,537 (54.7)	6292 (47.4)	14,240 (54.2)	3087 (42.4)	3427 (42.1)	5629 (57.8)	8178 (55.8)

[a]IQR: interquartile range.

[b]SES: socioeconomic status.

[c]N/A: not applicable.

[d]ACG: adjusted clinical group.

[e]HbA_{1c}: glycated hemoglobin.

Figures 3-6 show charts comparing the differences between the low-to-moderate and moderate-to-high health care services consumer clusters.

Figure 3 shows the differences in health care services use:

- Administrative contacts relate to scheduling an appointment or submitting a nonqueue request (NQR) by a face-to-face meeting at clinic or call to clinic or call center
- Health care practitioner contacts in community relate to physician and nurse consultations and overall visits to pharmacies
- Service consumption at hospital relates to hospitalization and visits to emergency departments
- New communication channels relate to scheduling an appointment or submitting an NQR by using the HMO internet website or smartphone app or answering an SMS suggesting a prescription renewal

Low-to-moderate health care services customers use more administrative contacts than moderate-to-high patients, who use more of the other communication methods.

Figure 4 depicts sociodemographic data. The differences in age and gender are relatively small. There is a higher representation of females in the low-to-moderate cluster, and its patients are slightly younger. More profound differences are at the socioeconomic status (SES) and religious sector. In the low-to-moderate group, there is a higher representation of patients with medium SES, and it has a higher representation of patients from the Arab sector. The moderate-to-high cluster comprises patients with low and high SES mainly from the general (Jewish sector) and immigrant sectors.

Figure 5 compares bioclinical follow-up quality. Generally, the follow-up quality is better in the moderate-to-high cluster than in the low-to-moderate one.

Figure 6 relates to adherence to treatment. A higher proportion of patients from the low-to-moderate cluster are not treated for diabetes, and the adherence of the ones who are treated is much lower than patients in the moderate-to-high cluster.

Figure 3. Low-to-moderate versus moderate-to-high health care services consumption of patients with diabetes in 2016.

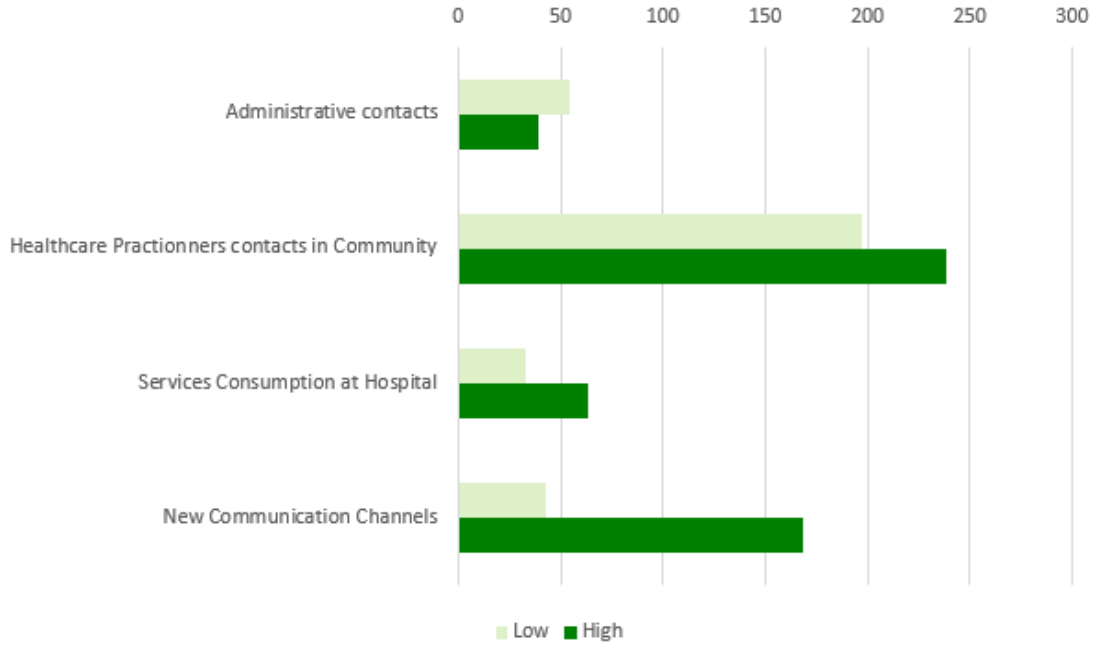


Figure 4. Differences in sociodemographic attributes between low-to-moderate and moderate-to-high health care services consumption of patients with diabetes in 2016. SES: socioeconomic status.

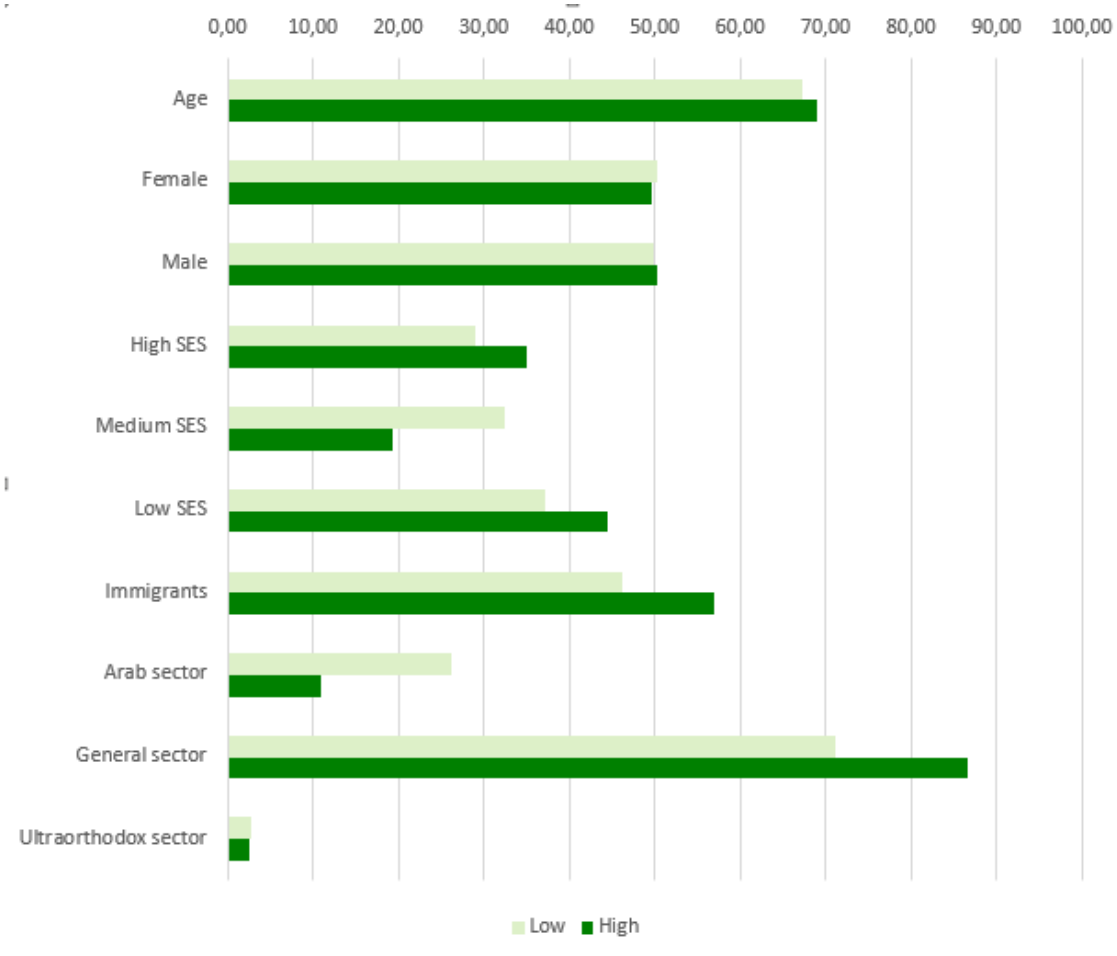


Figure 5. Differences of bioclinical follow-up quality between low-to-moderate and moderate-to-high health care services consumption of patients with diabetes in 2016. HbA1c: glycated hemoglobin.

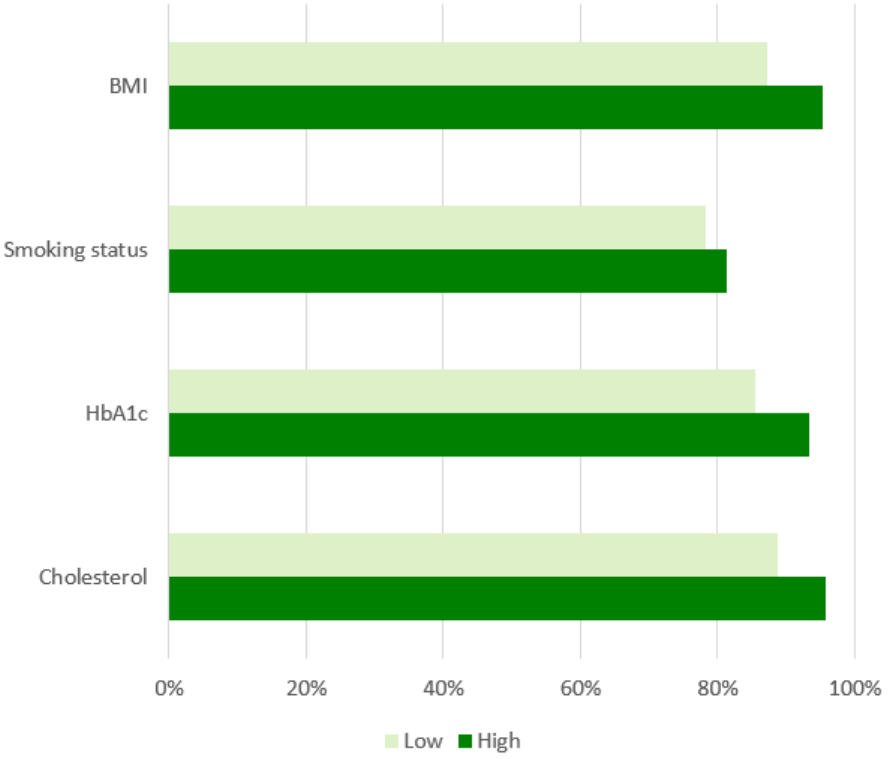
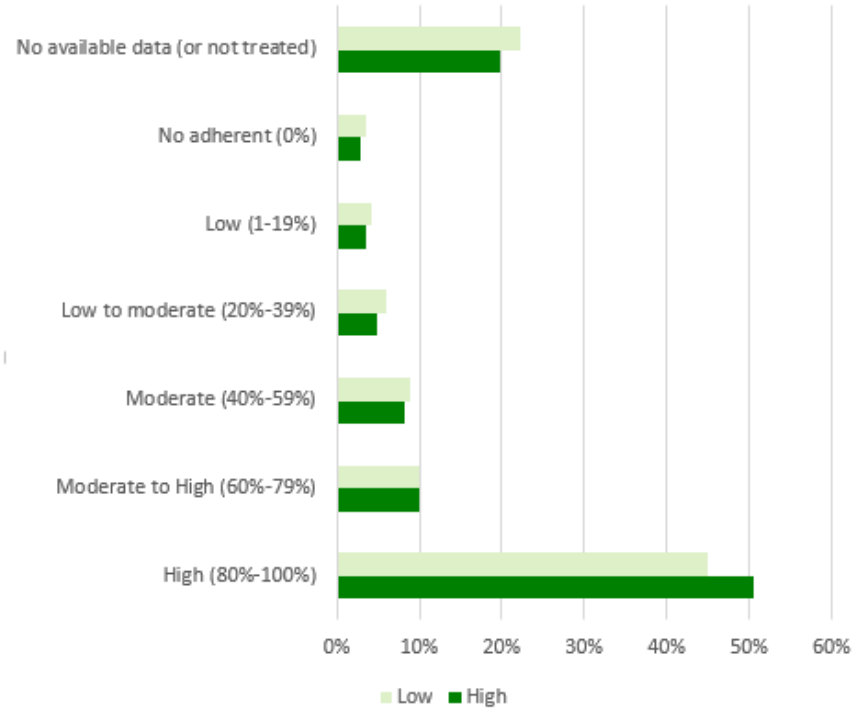


Figure 6. Differences of adherence to treatment between low-to-moderate and moderate-to-high health care services consumption of patients with diabetes in 2016.



Generally, low-to-moderate health care services customers tend to prefer direct contacts with health care professionals, and this is probably the cause for lower follow-up quality and adherence to treatment. Conversely, patients in the moderate-to-high clusters use myriad communication channels.

Below we describe the 13 communication profiles found in 2016 and characterize them based on sociodemographic and bioclinical data available. Keeping the population of each one constant, we describe how the communication behavior has

changed from 2009 to 2016 as health care professionals introduced new technological means of communications.

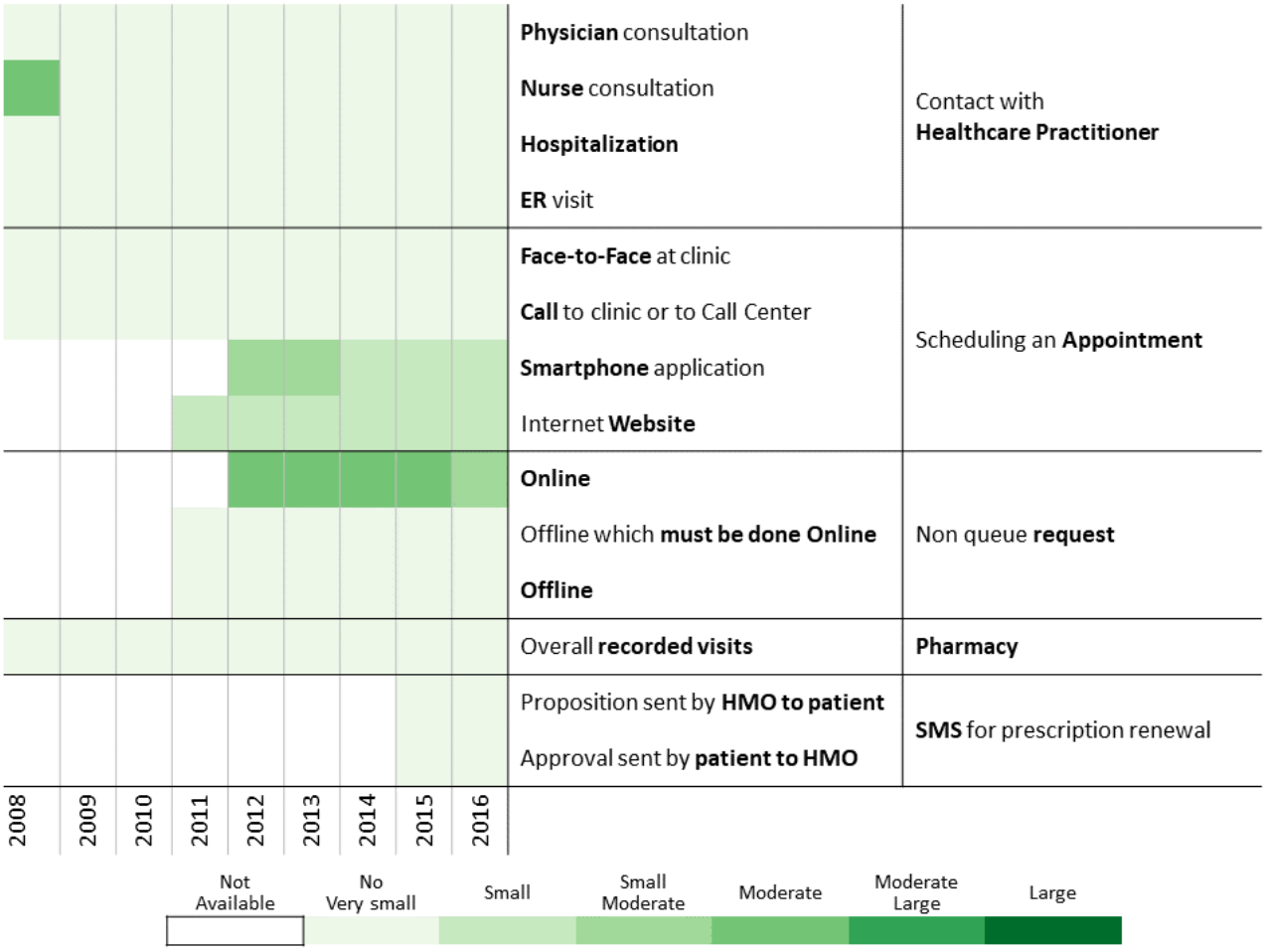
Low-to-Moderate Health Care Services Customers

Relatively Low Contact

Tech-Based

This cluster includes 11.48% (35,719/311,168) of the cohort. Patients in this group use fewer physical interactions and tend to be early adopters of new channels (Table 2 and Figure 7). They exhibit a relatively high use of electronic channels for scheduling appointments and online NQR tools when these channels became available. They are relatively young, men are highly represented (20,039/35,719, 56.10%), as are the Arab sector (10,537/35,719, 29.49%) and low SES population (12,479/35,719, 34.93%). Their follow-up quality is relatively poor but progressively improving. The missing measurements of BMI (2008: 37.30% [13,323/35,719]; 2016: 29.20% [10,430/35,719]) and HbA_{1c} (2008: 56.30% [20,110/35,719]; 2016: 36.30% [12,966/35,719]) decreased over time. Despite the aging, the HbA_{1c} average increased just a little (7.06 [SD 1.58] mmol/mol vs 7.31 [SD 1.7] mmol/mol). Of patients who were followed up, the percentage of patients treated for diabetes increased (2009: 32.8% [11,716/35,719]; 2016: 60.20% [21,503/35,719]), as did the proportion of highly adherent patients (2009: 13.40% [4786/35,719], 2016: 19.50% [6965/35,719]), with significant changes in 2011 and 2012 when NQRs and online (website and smartphone app) appointment scheduling were introduced. This group also started using the SMS channel for renewing prescriptions in 2015 when this channel was launched.

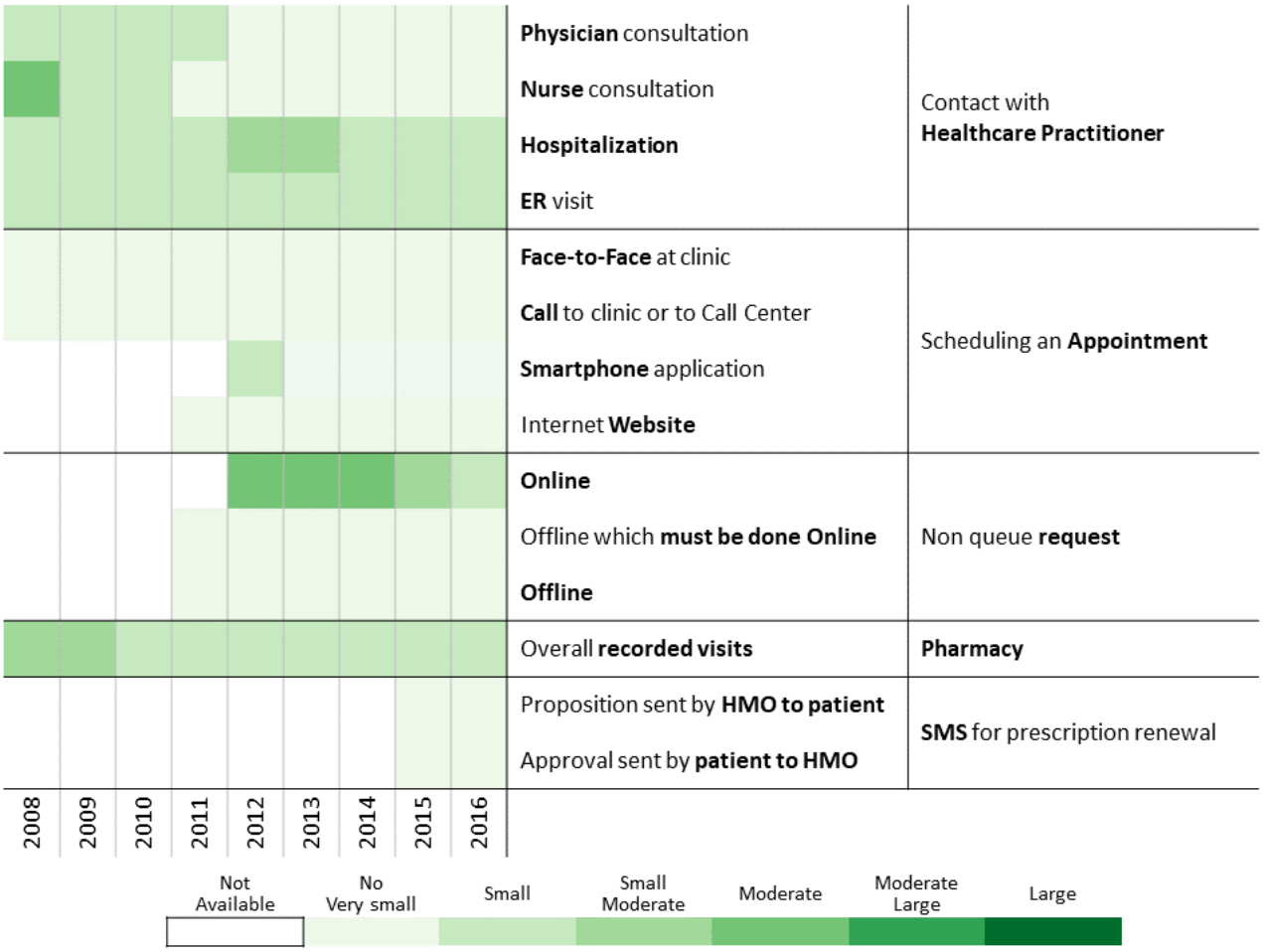
Figure 7. Communication pattern changes between 2008 and 2016 for the relatively low contacts — tech-based group.



Non–Clinic-Based

This cluster includes 10.95% of the cohort (34,060/311,168; Figure 8). Its median age is similar to the cohort, and the Arab sector is highly represented (9884/34,060, 29.02%). Until 2011, communication between patients and health care professionals was mainly achieved directly with the health care professionals. When electronic channels became available that year, the proportion of visits to health care professionals decreased in favor of online NQRs. The proportion of missing follow-up measurements is stable at around 20% each year. HbA_{1c} values are also relatively stable over time (7.16 [SD 1.39] mmol/mol in 2016). The percentage of patients treated for diabetes increased (2009: 56.50% [19,244/34,060]; 2016: 78.20% [26,635/34,060]) and is associated with an increase in high proportion of days covered (2009: 36.01% [12,266/34,060]; 2016: 54.11% [18,431/34,060]). This improvement in adherence over time is correlated with the use of the different electronic nonclinic contacts.

Figure 8. Communication pattern changes between 2008 and 2016 for the relatively low contacts — non–clinic-based group.



Tech-Based Versus Non–Clinic-Based

Analyzing the relatively low contact groups over time reveals that while the tech-based group prefers technology-based means of communication, the non–clinic-based group has a relatively high number of hospitalizations and visits to the emergency department compared with the rest of the population.

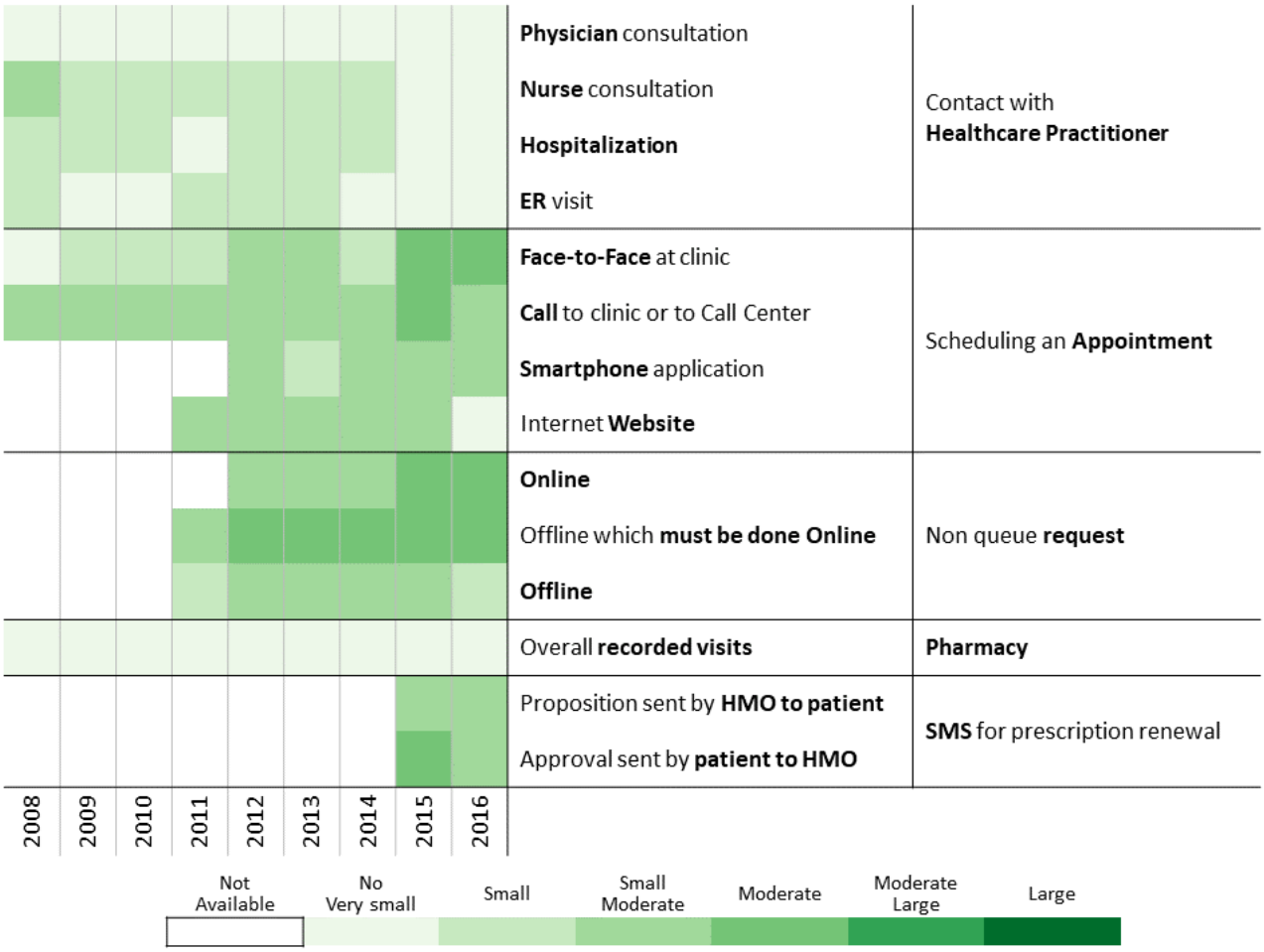
While the introduction of new digital communication channels is correlated with an increase in adherence to treatment for the tech-based group, patients in the non–clinic-based group do not exhibit this kind of behavior. On the contrary, there is a correlation between the introduction of new communication channels and a reduction in the average number of physician visits. For the tech-based cluster, the introduction of new communication channels might be viewed as an opportunity to decrease the number of contacts with the HMO.

Low and Scheduled Visits

Moderate Nonvisits

This group, which includes 15.92% (49,540/311,168) of the cohort, comprises patients who mainly use communication channels that do not involve face-to-face consultations (Table 2 and Figure 9). Its median age is close to the cohort median age of 67 years, and the general sector is highly represented (41,613/49,540, 84.00%). Until 2011 when the first electronic channels were introduced, patients in this group resorted to the available communication channels apart from physicians such as consulting nurses, hospitalizations, and emergency department visits. Once electronic options were introduced, the volume of contacts with health care professionals decreased in favor of the online tools for scheduling appointments and NQRs. Furthermore, this group was proactive and had a high answering rate to the SMS for automated prescription renewal. The proportion of patients taking a medication for diabetes jumped from 51.40% (25,464/49,540) in 2009 to 81.60% (40,425/49,540) in 2016. These values are associated with a high adherence to treatment, which increased by 17.70% between 2009 and 2016. Even though these patients consulted health care professionals less often, the proportion of missing annual measurements of follow-up metrics dropped from more than 20% in 2008 to approximatively 10% in 2016. For this subpopulation, additional channels are an opportunity to adjust communication patterns with the HMO to their preferences.

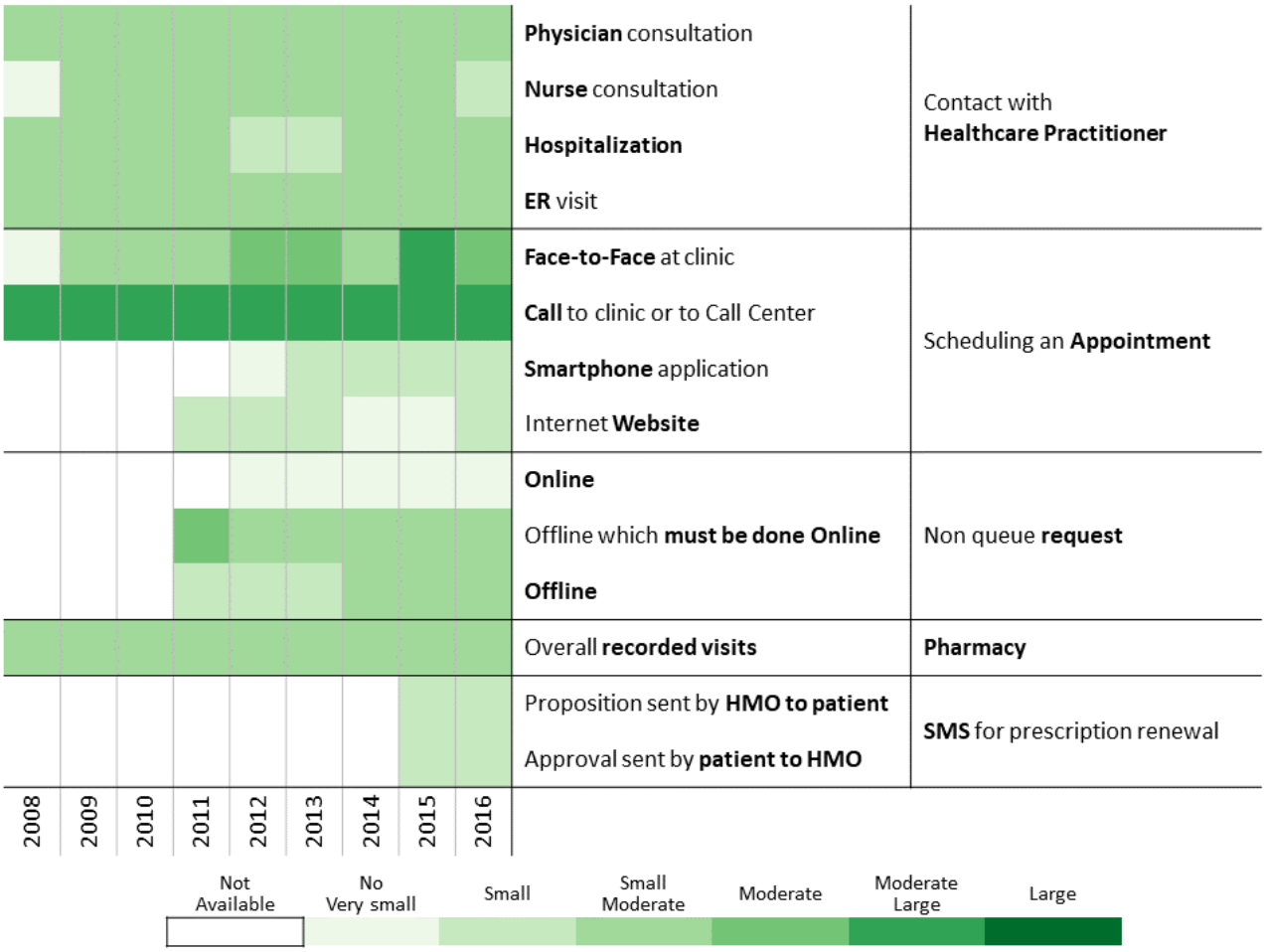
Figure 9. Communication patterns for patients having a low number but scheduled visits and preferring communication channels that do not involve a face-to-face meeting.



Low Balance Use

This cluster includes 13.41% (41,735/311,168) of the cohort (Table 2 and Figure 10). These patients are using all available services at a relatively low rate. The cluster median age is higher than the cohort, and there is a higher proportion of immigrants. Despite the diversity of available and newly introduced communication channels, health care services consumption is stable. These patients prefer traditional channels for appointment scheduling (ie, face-to-face at clinic, call to clinic or call center). Adding technology-based channels has a marginal effect on communication with health care professionals. Clinically, adjusted clinical group (ACG) level remains relatively stable over time (2010: 4 [3;4]; 2016: 4 [4;5]), and HbA_{1c} level is globally controlled (2016: 7.15 [SD 1.41] mmol/mol). Despite low use of available channels, patients have high rates of adherence to follow-up and treatment that increase over time. The proportion of missing annual measurements of follow-up metrics dropped (2008: around 15%; 2016: around 7%), and the percentage of patients with high adherence increased. This change may be attributed to aging and changes in therapeutic status (patients taking a noninsulin medication for diabetes jumped from 51.40% [25,464/49,540] in 2009 to 81.60% [40,425/49,540] in 2016).

Figure 10. Communication patterns for patients with diabetes having low balance use of health care services and scheduled visits.



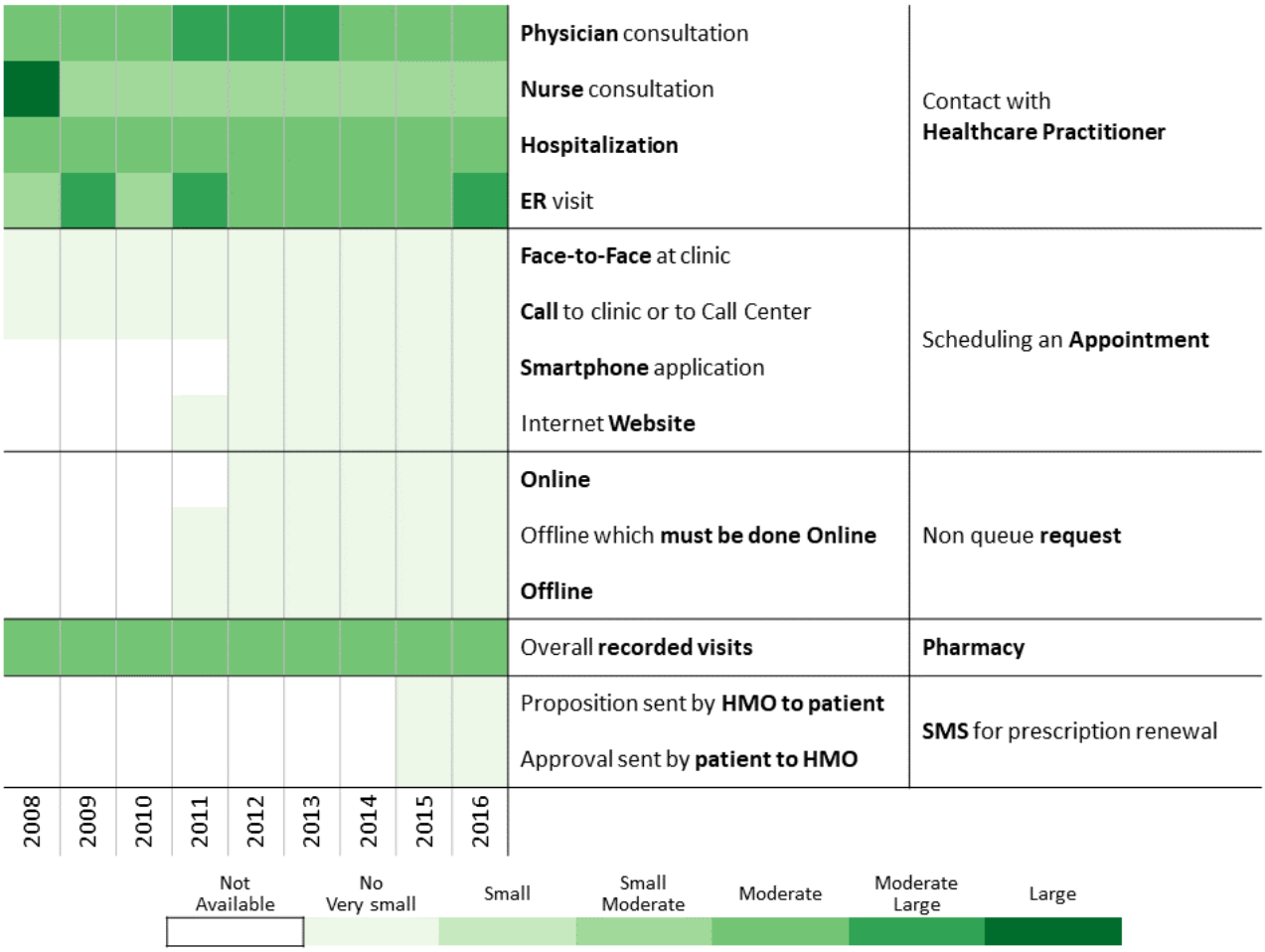
Measured Human Contacts

Measured human contacts means a relatively low use of appointment scheduling, NQRs, and SMS channels. Despite relatively low human contacts, these patients exhibit a moderate level of consulting health care professionals. Their median comorbidity level is relatively stable over time (2010: 4 [3;5]; 2016: 4 [4;5]) in view of the population aging.

Nonplanners

Nonplanners constitute 7.16% (22,275/311,168) of the cohort. They have a relatively moderate-to-high human health care services consumption but a low tendency to use nonhuman means of communication (Table 2 and Figure 11). This group consists of a larger proportion of patients from the Arab sector (9441/22,275, 42.38%), which is associated with a higher proportion of nonimmigrants (14,504/22,275, 65.11%) and people with low SES level (9669/22,275, 43.40%). The preference of nonplanned human contact does not have a negative clinical effect. HbA_{1c} values of the nonplanners are controlled over time (2010: 7.26 [SD 1.52] mmol/mol; 2016: 7.22 [SD 1.45] mmol/mol) with very good follow-up and adherence levels. The proportion of missing follow-up measurements dropped over time, and this improvement is associated with an increased number of patients treated with diabetes medication (2010: 61.51% [13,701/22,275]; 2016: 82.02% [18,271/22,275]) and better adherence over time (2010: 36.35% [8097/22,275]; 2016: 49.54% [11,034/22,275]).

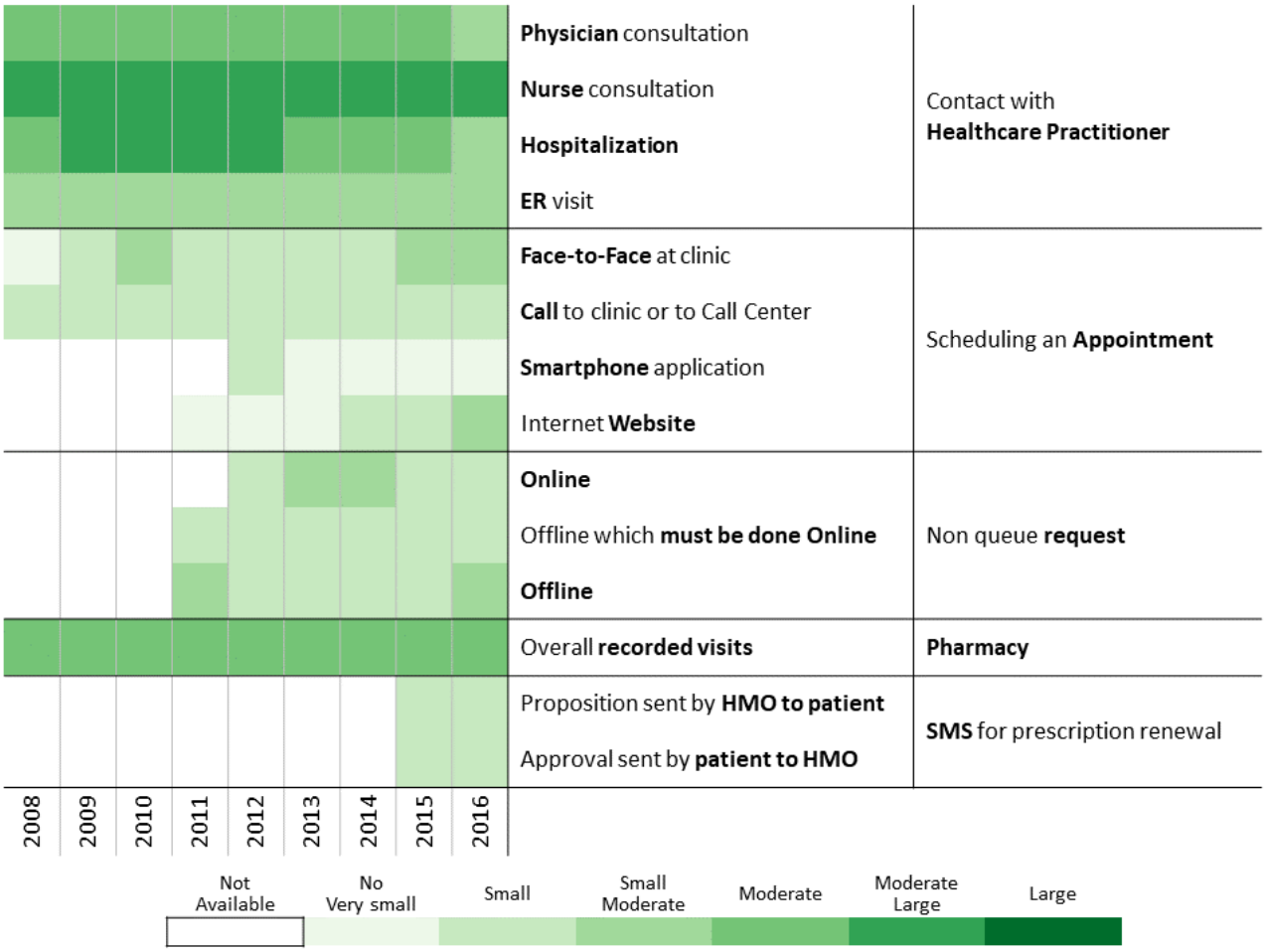
Figure 11. Communication patterns for patients having measured human contacts without generally scheduling their visits.



Nursing Contacts

The nursing-based contacts cluster constitutes 9.38% (29,197/311,168) of the cohort (Table 2 and Figure 12), and its members prefer to consult with nurses. They also tend to schedule their appointments by using all available channels. The proportion of females is slightly higher than in the cohort (15,463/29,197, 52.99%), as are the proportions of nonimmigrant patients (16,085/29,197, 55.09%), Arab sector representation (8599/29,197, 29.45%), and low SES population (10,158/29,197, 34.79%). These demographics may explain the tendency to rely on contacts with health care professionals (human contacts). The average HbA_{1c} level is stable over time (2010: 7.42 [SD 1.60] mmol/mol; 2016: 7.39 [SD 1.57] mmol/mol) and higher than the overall cohort. Although these results show merely a correlation between use of the nurse-patient channel and high levels of follow-up and adherence, they raise the hypothesis that the nurse-patient channel is very effective in inducing follow-up and adherence levels of patients. The percentage of patients who missed their annual measurements of follow-up metrics dropped over time (2008: around 15%, 2016: around 4%) and the adherence level increased (2010: 42.07% [12,284/29,197]; 2016: 50.27% [14,676/29,197]), as did the proportion of patients receiving treatment for diabetes (2010: 65.90% [19,240/29,197]; 2016: 81.25% [23,722/29,197]).

Figure 12. Communication pattern changes for patients with diabetes having measured human contacts mainly by consulting nurses.



High Electronic Contacts

Patients in these clusters have a high level of use of online and electronic channels for communicating with the HMO (Table 3, Figure 13, and Figure 14). The website group (those using a personal computer) comprises 6.20% of the cohort (19,277/311,168) and the smartphone (those using a mobile app) group comprises 4.27% (13,279/311,168).

Figure 13. Communication pattern changes for patients having a high volume of electronic contacts over the health management organization website.

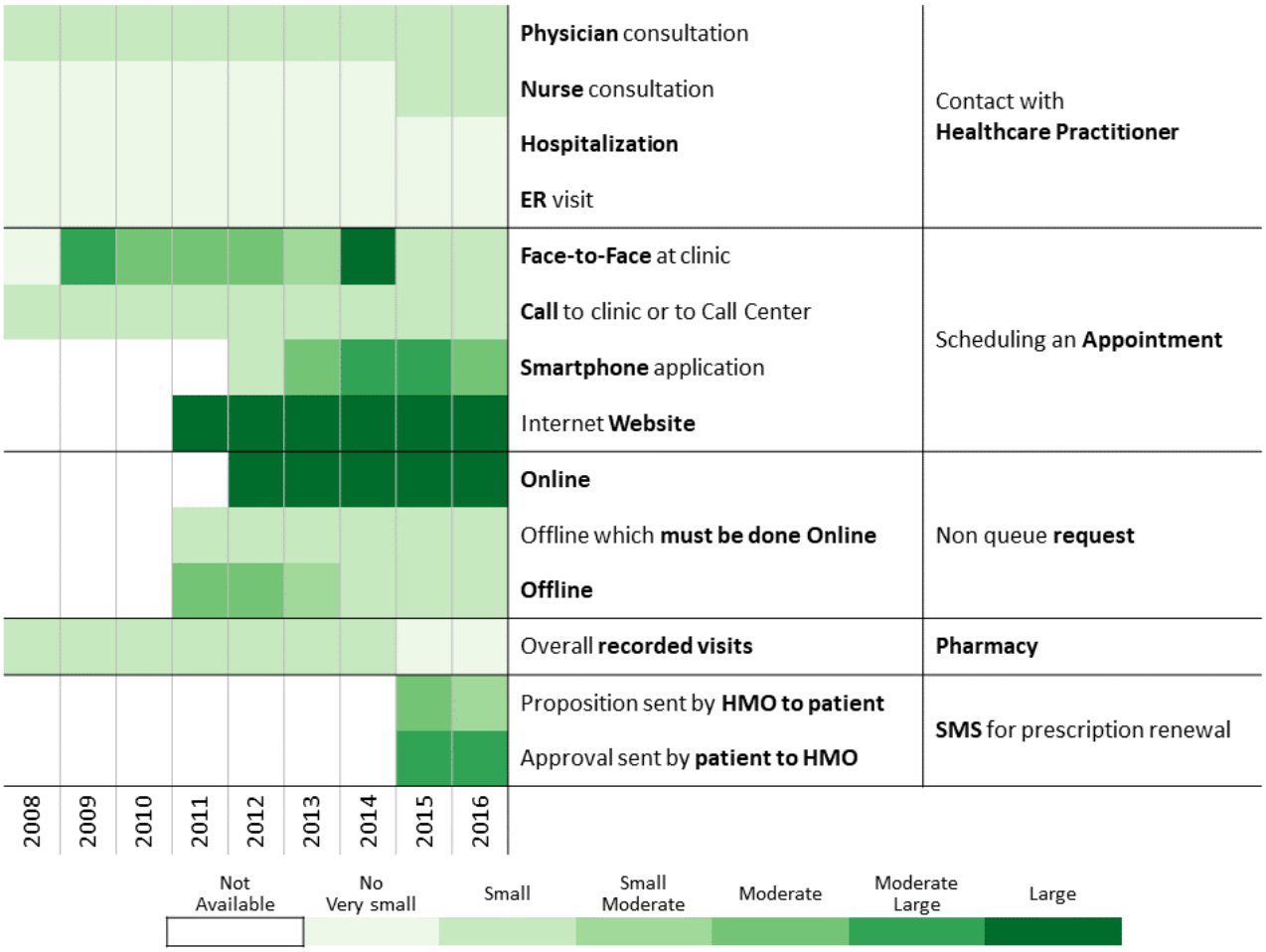
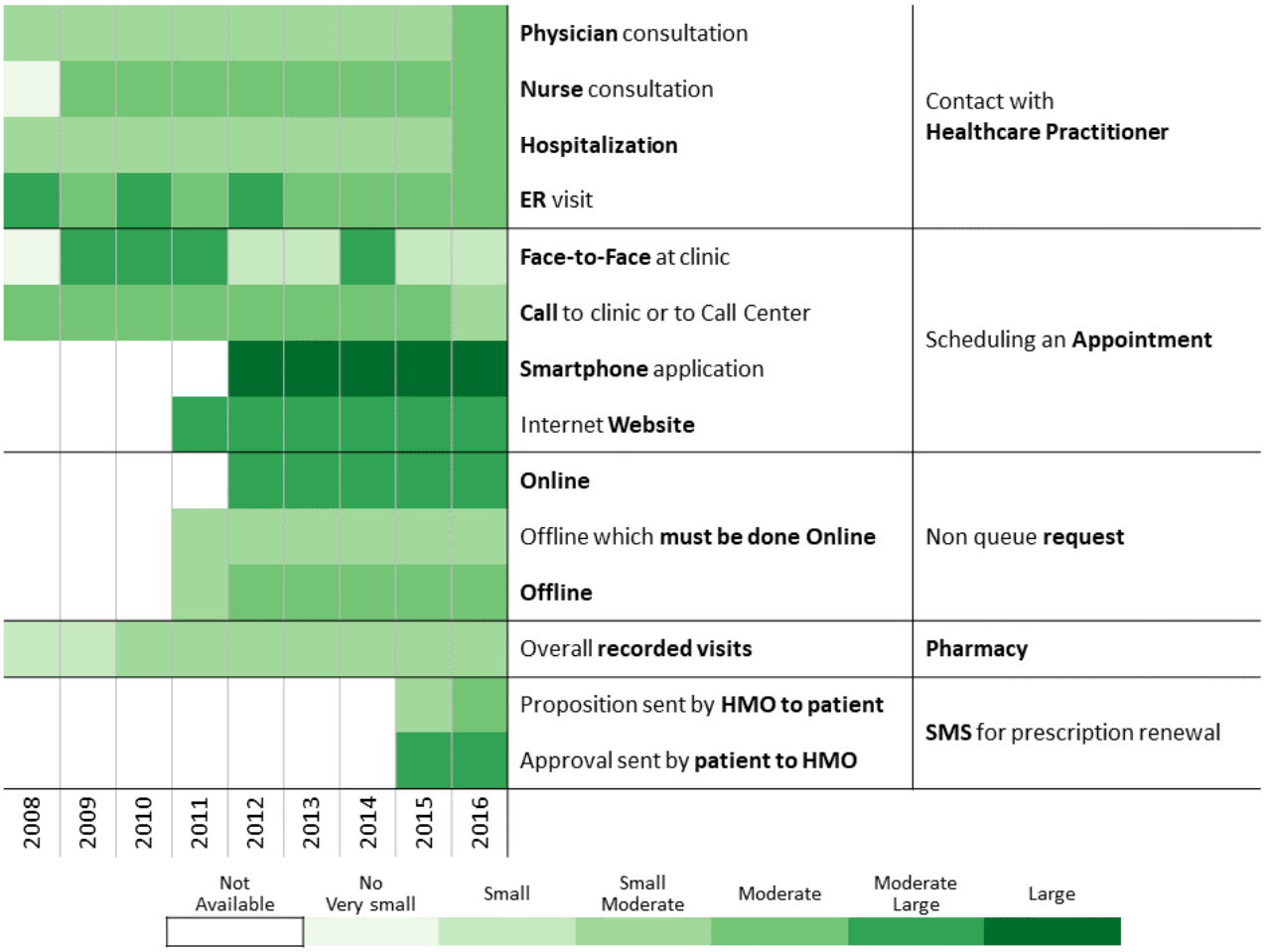


Figure 14. Communication pattern changes between 2008 and 2016 for patients with diabetes having a high volume of electronic contacts over the health management organization smartphone app.



Differences Between Website and Smartphone Clusters

These patients exhibit substantial use of the most prominent technological interfaces developed in the past 20 years, with smartphone users being younger. This phenomenon might relate to the lower penetration rate of new technologies in older populations. The gender profiles of the two groups differ (website: males, 55.12% [10,625/19,277]; smartphone: females: 54.60%, [7250/13,279]). Moreover, in both groups (website/internet and smartphone users) there is a higher representation of the general population (respectively 95.96% [18,498/19,277] and 82.66% [10,977/13,279] vs the overall population 77.8% (242,022/311,168]) and a higher representation of medium and high SES (respectively, the website/internet users having a high SES are 51.67% [9961/19,277] and 45.04% [5981/13,279] and the people of the smartphone cluster have a medium SES). This observation conforms to prior research, which found that lower SES populations gravitate toward smartphones [60,61].

Common Findings Between Website and Smartphone Clusters

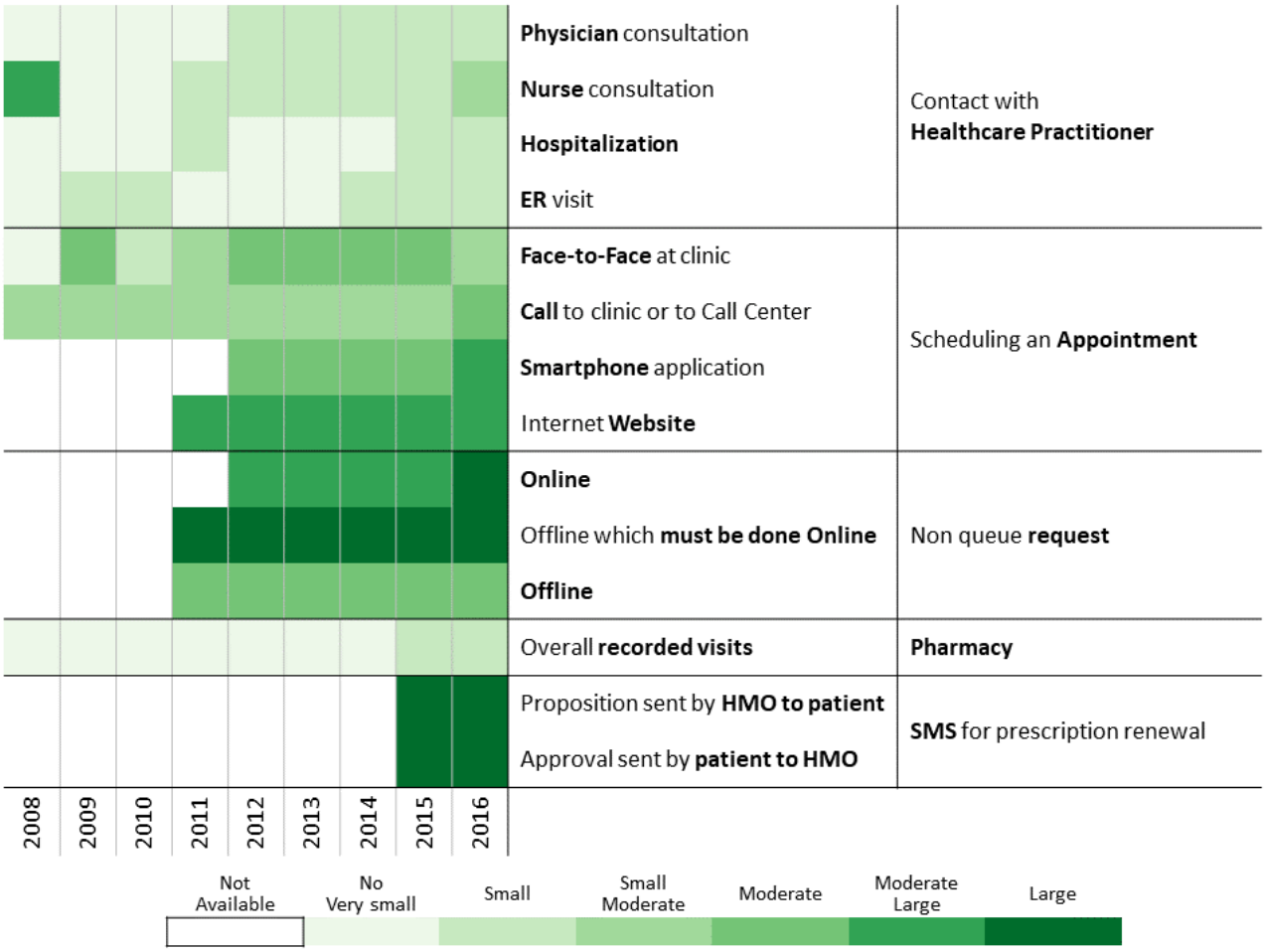
The heatmaps (Figure 13 and Figure 14) show high use of the website and smartphone app. Nevertheless, it is possible to see that the website cluster has a relatively low volume of direct contacts with health care professionals and prefers tools that enable nondirect and distant contacts. The clinical follow-up of these two clusters is better than that of the overall population, the proportion of missing follow-up indicators being lower in 2016 (Table 3). Furthermore, treatment adherence was better in 2016 than in the cohort population and also increased over the years in parallel with the number of newly treated patients. Patients in these two clusters tend to reduce their use of other means of communication and contacts with the HMO in favor of electronic media while maintaining a high follow-up quality and treatment adherence. By considering what seems like a positive impact of smartphone presence, the HMO should incorporate more functions into the smartphone app.

Automated Interaction Early Adopters

These patients are relatively young (Table 3 and Figure 15), and females and the general sector are largely represented (15,034/26,290, 57.19%, and 19,310/26,290, 73.45%, respectively). Despite being early adopters of new interaction services, this cluster uses all human contact–based services over time. The new automated interaction tools improve the quality of contacts with the HMO and do not serve as a replacement to previously existing channels. The comorbidity of this relatively young group is high (ACG: 5 [4;5]), and the proportion of missing bioclinical and follow-up measurements in 2016 is relatively low. Considering the aging process and the diabetes treatment policy change (lowering the HbA_{1c} threshold from 7.5% to 6.5%), the number of patients taking a noninsulin

treatment for diabetes increased over time (2010: 66.81% [17,565/26,290]; 2016: 83.73% [22,012/26,290]). Moreover, the proportion of missing follow-up measurements decreased (eg, BMI: 2010, 6.83% [1796/26,290]; 2016, 1.27% [334/26,290]; HbA_{1c}, 2010, 17.08% [4492/26,290]; 2016, 3.69% [970/26,290]) and the proportion of adherence to treatment increased (2010: 44.85% [11,792/26,290]; 2016: 54.17% [14,240/26,290]). These results may indicate that the new tools allow patients to improve their engagement with the HMO. As their disease progresses, an increase in their services consumption is expected, but instead we observe a slight decrease in some of the services consumed. One contributing factor to the high prevalence of patients from the general sector in the cluster may be related to the sole use of Hebrew in the internet and smartphone app.

Figure 15. Communication pattern changes between 2008 and 2016 for patients with diabetes being early adopters of automated interaction tools.



Human-Based Contacts

Differences Between Nursing-Centered and Physician-Centered Clusters

Patients having mainly human-based contacts (Table 3) with the HMO are divided in two clusters: nursing-centered (7276/311,168, 2.34%; Figure 16) and physician-centered (8137/311,168, 2.61%; Figure 17). It should be noted that one of the main differences between these two populations is the proportion of patients missing their follow-up measurements in the physician-centered group. This highlights the importance of nurse involvement in patient follow-up.

Figure 16. Communication pattern changes between 2008 and 2016 for patients with diabetes having mainly human-based contacts based on interactions with nursing.

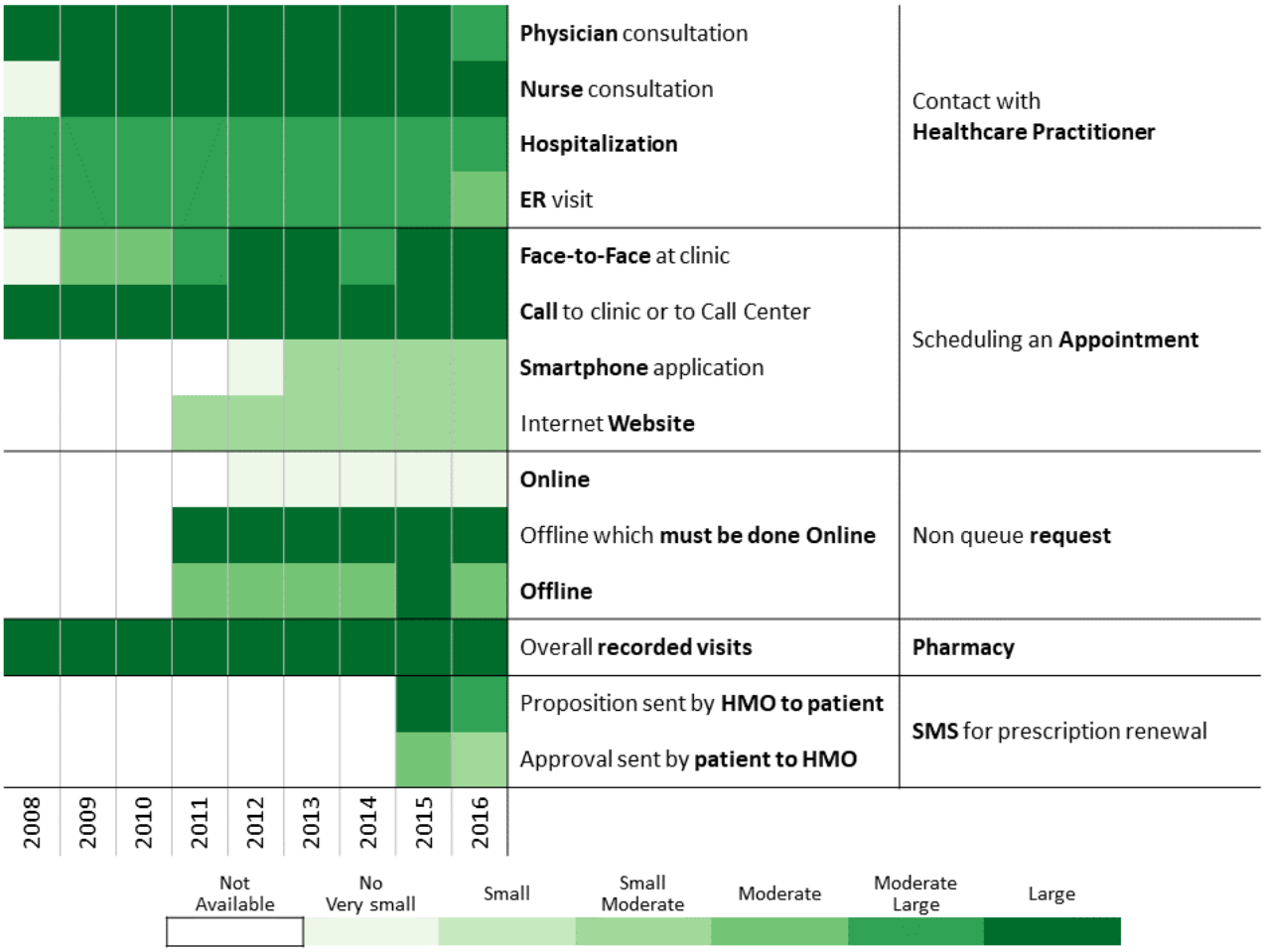
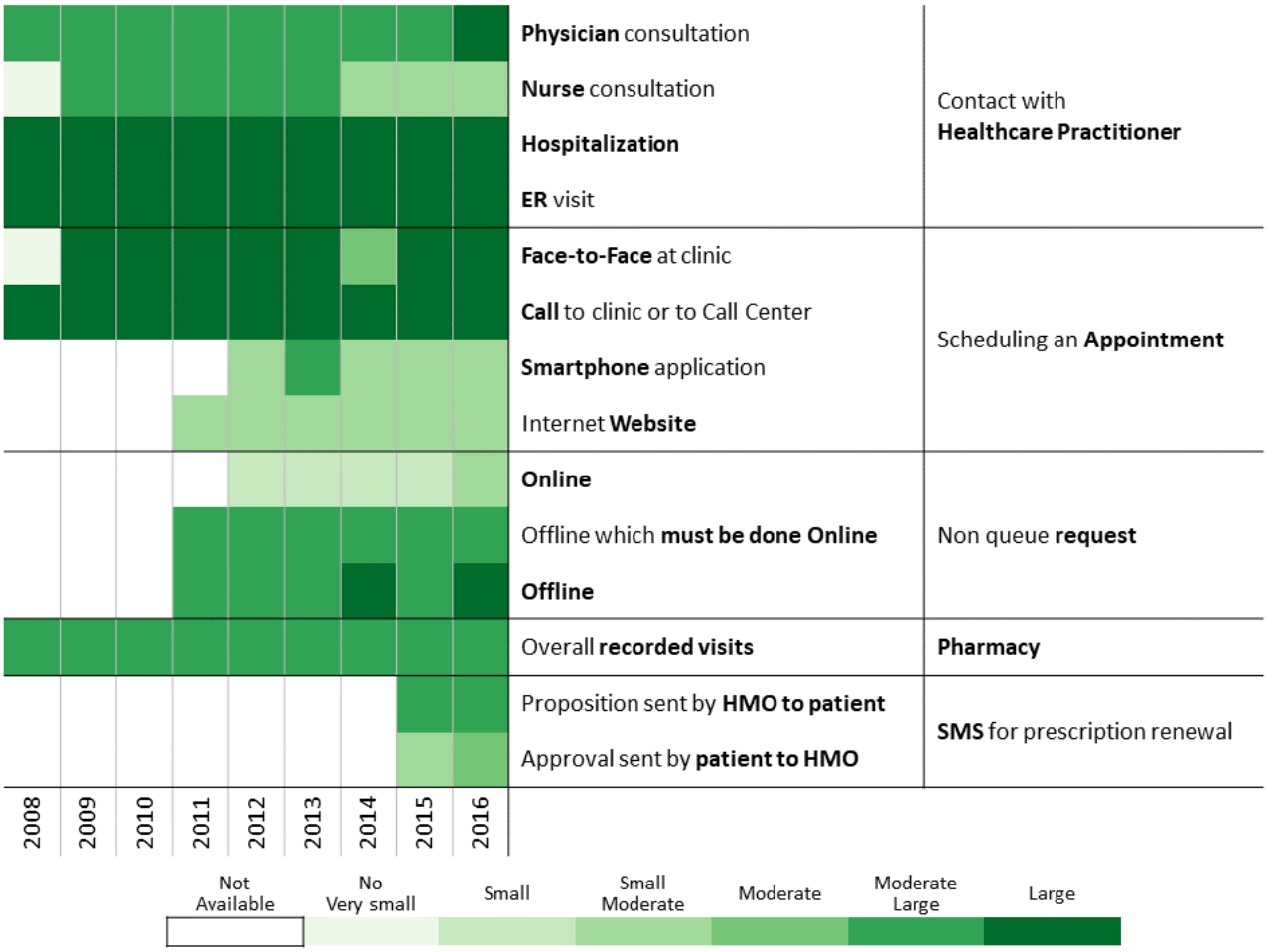


Figure 17. Communication pattern changes between 2008 and 2016 for patients with diabetes having mainly human-based contacts based on interactions with physicians.



Common Findings in Nursing-Centered and Physician-Centered Clusters

These two clusters are relatively similar. In 2016, the proportion of female patients is relatively high and the population is older, with a majority of immigrants, and a higher representation of patients from medium-low SES groups. Patients in these two clusters had a relatively high ACG score over time. This increasing level of comorbidity can justify the high volume of nurse and physician consultations and high follow-up quality scores.

Higher Resource Consumers Having Overall High Contacts and Electronic Driven Interactions

Overall high contact represents 3.13% (9736/311,168) of the cohort (Table 3 and Figure 18) and the electronic driven interaction 4.71% (14,647/311,168; Table 3 and Figure 19). They comprise older people (mean age 71 years) and a higher proportion of men and immigrants. From ethnicity and SES perspectives, the distributions for the overall high contact group resemble the cohort, while the electronic driven interaction group has a higher representation of the general sector (9241/9736, 94.92%) and patients with high SES (4535/9736, 46.58%).

Figure 18. Communication pattern changes between 2008 and 2016 for patients with diabetes having an overall high number of contacts with health care services.

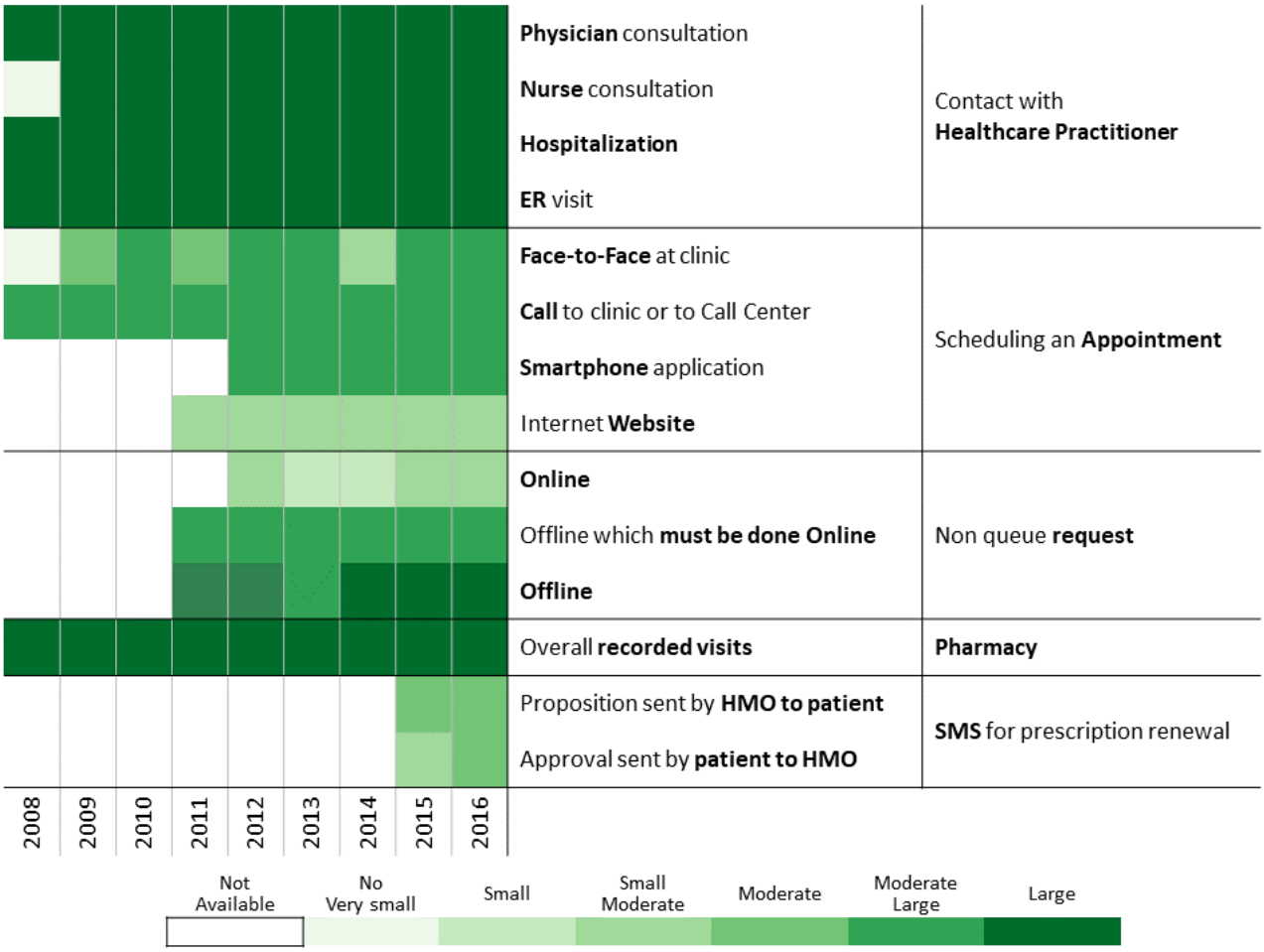
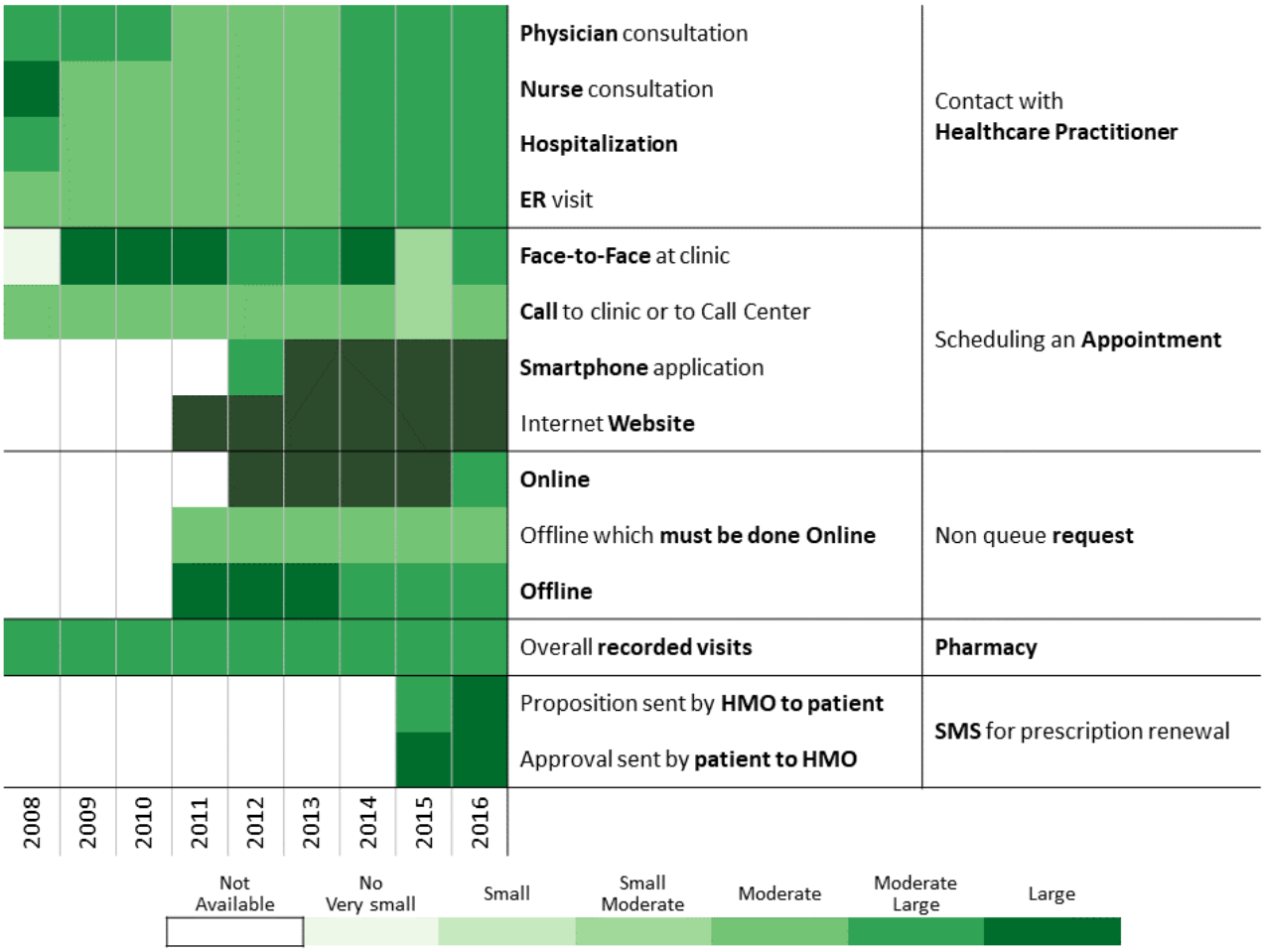


Figure 19. Communication pattern changes between 2008 and 2016 for patients with diabetes leading electronic-driven interaction with the health management organization.



The main differences in the communicational profiles are in the interaction strength. The volume of contacts is high in both groups but higher for the overall high contact group compared with the electronic driven interaction group. This can be explained by the lower SES in the first cluster and Hebrew-speaking abilities, which may be lower than in the second. Lower SES may be considered as a proxy for confidence and ability to use technologies, and immigrant status, age, and SES as proxies for defining language abilities. The introduction of new communication channels over time only increased the global number of contacts. The large number of visits at the clinic may be justified by the higher comorbidity level. Follow-up measurements in both clusters are consistently better than the cohort. Treatment compliance is better for the electronic driven interaction cluster and increasing over time and in parallel to new communication channels introduction. Adherence to treatment of the overall high contact group is around 42% over time and not influenced by the addition of technological channels.

The age and comorbidity levels in these two clusters strongly influence the number of contacts with health care professionals over time and, as a by-product, the quality of follow-up improves. However, results suggest that SES and immigrant status influence the use of new technologies and increase the number of contacts with the HMO and patient adherence to treatment.

Discussion

Principal Findings

In this study, we identified and characterized 13 media profiles of patients. We have shown how communication behavior is influenced by the means of communication that the health organization provides to the patient. Additionally we have pointed out how different patients respond to technology-based communication and change the way they communicate with the health organization. Finally, we highlighted that some patients prefer to communicate with the organization by technological means and respond adequately to text messages, others prefer to communicate with the physician, and others with the nurse.

Identifying the channels of communication with the health organization and health care professionals preferred by each patient creates an opportunity to convey messages adapted to the patient in the most suitable communication channel. The greater the likelihood that the therapeutic message is received by the patient, the greater the patient's response to treatment, and the better the health of the patient.

Strengths and Limitations

Clalit insures and provides medical services to more than 54% of the Israeli population. It is the largest health care organization and insurer in Israel. However, although Clalit covers most of

the population, the overall ethnic distribution of its health care customers does not accurately reflect the Israeli demographic composition: it has a higher proportion of Arabs, a lower proportion of ultra-Orthodox, and a higher proportion of members with a low SES [19].

Patients With Diabetes and Generalization to the Overall Chronic Patient Population

This retrospective analysis looks at Israeli patients treated by an Israeli HMO. The Israeli health care system, culture, and norms are factors affecting patient behavior in a specific way that do not allow a direct generalization of the results to other parts of the world.

This study overcomes a limitation of prior research dealing with the identification and description of health care customer communication patterns among individuals with diabetes in Clalit in 2015 [62]. Analyzing data that spans 9 years provides a better understanding of the changes of communication channel use over time and impact of socioeconomic factors, which cannot be easily and clearly understood with a 1-year snapshot.

Effect of Digital Communication Tools

The digital tools introduced between 2009 and 2016 for patients diagnosed with diabetes influence their follow-up and communication pathways with health care professionals. However, for more than half of the population investigated in this research, we found only a negligible influence of the digital tools on the communicational behavior (relatively low contacts, low and scheduled visits–low balance uses, measured human contacts, human-based contacts, and overall high contact clusters, 54.30%).

Digital tools, such as NQRs and SMS for prescription renewal, allow patients to reduce or avoid visits to the clinic or hospital. For patients initially having a relatively low number of health care practitioner contacts over time, these digital tools may induce a reduction in the number of visits to the clinic or nurse station. Eliminating potential visits due to the introduction of digital tools might influence the follow-up quality because these visits could have served as another opportunity for a human contact with the patient (eg, for discussing treatment issues) or at least to measure the patient's condition [61,62]. Almost a third of the research population use technology to reduce their engagement with health care professionals. Not surprisingly, these are patients who tend to have a relatively small number of interactions with health care professionals (relatively low contacts–tech-based, low and scheduled visits–moderate nonvisits, electronic users, and automated interaction early adopters, 32.10%). We would like to emphasize that the results do not show that the introduction of digital tools deteriorates the health condition of this one-third of the population. Nor do we claim that the reduction in visits to health care professionals is inherently an unwanted outcome. On the contrary, this is exactly what EMR systems are designed for. Rather, we claim that for targeted populations, which do not communicate efficiently with health care professionals, new digital tools might have negative consequences on the quality of the follow-up. This danger can be mitigated by using additional, human-based communication channels, akin to the guided-care approach [63], which have already proven to be effective. We can now build tools to identify these patients based on their behavior and target the efforts on the population that needs it.

For about 13.60% of the population, the introduction of new and digital channels in the communication arsenal of the HMO is effective. These are the patients who belong to the following clusters: high electronic contacts–website, high electronic contacts–smartphone, and electronic driven interaction.

Patients in the low-to-moderate clusters were found to have different health outcomes due to a lower health care services consumption impacting their follow-up quality and adherence to treatment. On the other hand, patients who consume more services, the ones in the moderate-to-high clusters, have better health outcomes (despite being generally older and with a higher ACG). To sum up, it looks like that the effect of digital communication tools is to improve the follow-up and adherence to treatment instead of replacing human interactions with health care professionals.

Current and Potential Future Directions

As time progresses, the population becomes more accustomed to using digital channels and new communication channels are introduced (eg, an online counseling services with video calls to physicians when clinics are closed, available in Clalit since 2017). Communication patterns should be monitored in the face of the rapid changes in population behavior and services offered.

Furthermore, by tuning its communication tools to patient preferences and special needs (eg, by translating the user interfaces of electronic communications tools to languages such as Arabic, English, Russian, Amharic, French, Spanish), the health organization would realize the following:

- Improve and increase accessibility to health care services, achieve better patient engagement and responsiveness to treatment, and improve quality of treatment and treatment experience within existing budgetary constraints
- Increase patient engagement with the treatment process by transforming the communication scheme with each patient to a more proactive scheme to better fit patient profile
- Allow patient-reported outcome measures [64] for some follow-up measurements such as BMI (or more specifically, weight) and smoking status in an effort to reduce nurse work (over)load while continuing and improving patient follow-up

Finally, we investigated only diabetic patients. This research and its related methodology can be generalized and extended to other chronic and acute patients.

Conclusion

In this paper we presented and demonstrated a methodology to identify communication profiles over time within health care systems. We applied this methodology to the data of more than 300,000 diabetic patients from Clalit Health Services in Israel and found 13 such profiles. These profiles enabled health care professionals and the insurer to adapt the communication and message conveyed to patients based on their communication profile. This methodology can be applied in other organizations in other geographical locations.

We found that 22.40% of patients have very low health services consumption, and an additional 45.90% have low-to-moderate health services consumption, which indicated a low level of patient engagement. We showed that the introduction of technological communication channels didn't substantially improve the engagement of these patients and for some of the patients it even reduced communication with health care professionals. Based on these findings, we think that improving patient engagement cannot rely solely on technological solutions; rather, these solutions must be accompanied by complementary means [65].

Acknowledgments

The research was supported by a grant from the Israel National Institute for Health Policy (#188-15).

Conflicts of Interest

None declared.

References

1. Axén I, Bodin L, Bergström G, Halasz L, Lange F, Lövgren P, et al. Clustering patients on the basis of their individual course of low back pain over a six month period. BMC Musculoskelet Disord 2011 May 17;12:99 [FREE Full text] [doi: 10.1186/1471-2474-12-99] [Medline: 21586117]
2. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 02;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
3. Hoffman AS, Volk RJ, Saarimaki A, Stirling C, Li LC, Härter M, et al. Delivering patient decision aids on the Internet: definitions, theories, current evidence, and emerging research areas. BMC Med Inform Decis Mak 2013;13 Suppl 2:S13 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S13] [Medline: 24625064]
4. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
5. Moick M, Terlutter R. Physicians' motives for professional internet use and differences in attitudes toward the internet-informed patient, physician-patient communication, and prescribing behavior. Med 2 0 2012;1(2):e2 [FREE Full text] [doi: 10.2196/med20.1996] [Medline: 25075230]
6. Kritz M, Gschwandtner M, Stefanov V, Hanbury A, Samwald M. Utilization and perceived problems of online medical resources and search tools among different groups of European physicians. J Med Internet Res 2013;15(6):e122 [FREE Full text] [doi: 10.2196/jmir.2436] [Medline: 23803299]
7. Dugdale DC, Epstein R, Pantilat SZ. Time and the patient-physician relationship. J Gen Intern Med 1999 Jan;14 Suppl 1:S34-S40 [FREE Full text] [doi: 10.1046/j.1525-1497.1999.00263.x] [Medline: 9933493]
8. Weiner JP. Doctor-patient communication in the e-health era. Isr J Health Policy Res 2012;1(1):33 [FREE Full text] [doi: 10.1186/2045-4015-1-33] [Medline: 22929000]
9. Peleg R, Avdalimov A, Freud T. Providing cell phone numbers and email addresses to patients: the physician's perspective. BMC Res Notes 2011 Mar 23;4:76 [FREE Full text] [doi: 10.1186/1756-0500-4-76] [Medline: 21426591]
10. Peleg R, Nazarenko E. Providing cell phone numbers and e-mail addresses to patients: the patient's perspective, a cross sectional study. Isr J Health Policy Res 2012 Aug 28;1(1):32 [FREE Full text] [doi: 10.1186/2045-4015-1-32] [Medline: 22929801]
11. Ha JF, Longnecker N. Doctor-patient communication: a review. Ochsner J 2010;10(1):38-43 [FREE Full text] [Medline: 21603354]
12. Rothlind E, Fors U, Salminen H, Wändell P, Ekblad S. Circling the undefined: a grounded theory study of intercultural consultations in Swedish primary care. PLoS One 2018;13(8):e0203383 [FREE Full text] [doi: 10.1371/journal.pone.0203383] [Medline: 30161227]
13. Shang L, Zuo M, Ma D, Yu Q. The antecedents and consequences of health care professional-patient online interactions: systematic review. J Med Internet Res 2019 Sep 25;21(9):e13940 [FREE Full text] [doi: 10.2196/13940] [Medline: 31573908]
14. Lupton D. Digital health and health care. In: Scambler G, editor. Sociology as Applied to Health and Medicine, 7th Edition. London: Macmillan Publishers; 2018:277-290.
15. Husted GR, Weis J, Teilmann G, Castensøe-Seidenfaden P. Exploring the influence of a smartphone app (Young with Diabetes) on young people's self-management: qualitative study. JMIR Mhealth Uhealth 2018 Feb 28;6(2):e43 [FREE Full text] [doi: 10.2196/mhealth.8876] [Medline: 29490897]
16. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]

17. Henao R, Murray J, Ginsburg G, Carin L, Lucas JE. Patient clustering with uncoded text in electronic medical records. AMIA Annu Symp Proc 2013;2013:592-599 [FREE Full text] [Medline: 24551361]
18. Sewitch MJ, Leffondré K, Dobkin PL. Clustering patients according to health perceptions: relationships to psychosocial characteristics and medication nonadherence. J Psychosom Res 2004 Mar;56(3):323-332. [doi: 10.1016/S0022-3999(03)00508-7] [Medline: 15046970]
19. Benis A, Harel N, Barak Barkan R, Srulovici E, Key C. Patterns of patients' interactions with a health care organization and their impacts on health quality measurements: protocol for a retrospective cohort study. JMIR Res Protoc 2018 Nov 07;7(11):e10734 [FREE Full text] [doi: 10.2196/10734] [Medline: 30404769]
20. Webster C. EHR business process management: from process mining to process improvement to process usability. Healthcare Systems Process Improvement Conf., Las Vegas 2012:1-6.
21. Markides M. The importance of good communication between patient and health professionals. J Ped Hematol Oncol 2011;33:S123-S125. [doi: 10.1097/mph.0b013e318230e1e5]
22. Paternotte E, van Dulmen S, van der Lee N, Scherpbier AJ, Scheele F. Factors influencing intercultural doctor-patient communication: a realist review. Patient Educ Couns 2015 Apr;98(4):420-445. [doi: 10.1016/j.pec.2014.11.018] [Medline: 25535014]
23. Cohen-Stavi CJ, Balicer RD, Roberts ML. Innovation in health care for proactive care delivery and strategic clinical decision-making: integrating research, technology and practice. Public Health Panorama 2018 Sep;4(3):470-474 [FREE Full text] [doi: 10.1093/oxfordjournals.pubmed.a043518]
24. Jaffe DH, Flaks-Manov N, Benis A, Gabay H, DiBonaventura M, Rosenbaum H, et al. Population-based cohort of 500 patients with Gaucher disease in Israel. BMJ Open 2019 Jan 21;9(1):e024251 [FREE Full text] [doi: 10.1136/bmjopen-2018-024251] [Medline: 30670517]
25. Rennert G, Peterburg Y. Prevalence of selected chronic diseases in Israel. Isr Med Assoc J 2001 Jun;3(6):404-408 [FREE Full text] [Medline: 11433630]
26. Karpati T, Cohen-Stavi CJ, Leibowitz M, Hoshen M, Feldman BS, Balicer RD. Towards a subsiding diabetes epidemic: trends from a large population-based study in Israel. Popul Health Metr 2014;12(1):32 [FREE Full text] [doi: 10.1186/s12963-014-0032-y] [Medline: 25400512]
27. Ifrah A, Libroder C, Karolinsky D. In: Shohat T, Keinan-Boker L, Ifrah A, Libruder C, editors. Highlights of Health in Israel 2016. Israel: Israel Center for Disease Control, Israel Ministry of Health; May 2017:1-111.
28. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. AI Magazine 1996;17(3):37. [doi: 10.1609/aimag.v17i3.1230]
29. Fayyad U, Piatetsky-Shapiro G, Smyth P. The KDD process for extracting useful knowledge from volumes of data. Commun ACM 1996;39(11):27-34. [doi: 10.1145/240455.240464]
30. Frawley P, Piatetsky-Shapiro G, Matheus CJ. Knowledge discovery in databases: an overview. AI Magazine 1992;13(3):57-70 [FREE Full text] [doi: 10.1609/aimag.v13i3.1011]
31. Zaki M, Meira W. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge: Cambridge University Press; 2014.
32. Mitchell T. Machine Learning. New York: McGraw Hill; 1997.
33. O'Connor PJ, Bodkin NL, Fradkin J, Glasgow RE, Greenfield S, Gregg E, et al. Diabetes performance measures: current status and future directions. Diabetes Care 2011 Jul;34(12):1651-1659 [FREE Full text] [doi: 10.2337/dc11-0735] [Medline: 21709298]
34. Turner M, Burns SM, Knight L, Ward K, Garo A, Morris T, et al. Weight management practices among heart and vascular health care providers in an ambulatory setting. Medsurg Nurs 2012;21(4):222-232. [Medline: 22966525]
35. Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus: abbreviated report of a WHO consultation. Geneva: World Health Organization; 2011.
36. Reid RJ, Roos NP, MacWilliam L, Frohlich N, Black C. Assessing population health care need using a claims-based ACG morbidity measure: a validation analysis in the Province of Manitoba. Health Serv Res 2002 Oct;37(5):1345-1364 [FREE Full text] [Medline: 12479500]
37. Norwegian Institute of Public Health. Guidelines for ATC classification and DDD assignment. Oslo: WHO Collaborating Centre for Drug Statistics Methodology; 2018. URL: https://www.drugsandalcohol.ie/29364/1/WHO%20Collaborating%20Centre%20for%20Drug%20Statistics%20Methodology.pdf [accessed 2020-07-24]
38. Ben-Gal I. Outlier detection. In: Maimon O, editor. Data Mining and Knowledge Discovery Handbook, 2nd Edition. New York: Springer; 2011:117-130.
39. Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J Exper Soc Psychol 2013 Jul;49(4):764-766. [doi: 10.1016/j.jesp.2013.03.013]
40. Bansal R, Gaur N, Singh S. Outlier detection: applications and techniques in data mining. 2016 Presented at: 2016 6th International Conference—Cloud System and Big Data Engineering (Confluence); 2016; Noida. [doi: 10.1109/confluence.2016.7508146]
41. Rahm E, Hong H. Data cleaning: problems and current approaches. IEEE Bull Tech Comm Data Engineer 2000;23:1-11 [FREE Full text]

42. Han J, Kamber M. Data Mining: Concepts and Techniques, 3rd Edition. Burlington: Morgan Kaufmann; 2011.
43. Escalante HJ. A comparison of outlier detection algorithms for machine learning. In: Memoria del XIV Congreso Internacional de Computación (CIC 2005). 2005 Presented at: Congreso Internacional en Computacion - IPN, 2005; September 2005; México URL: https://pdfs.semanticscholar.org/cf06/9b7460ce1b5a0434a6a19f420544a780f35d.pdf [doi: 10.1002/0471725382.ch5]
44. Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Comput Surv 2009;41(3):1-58. [doi: 10.1145/1541880.1541882]
45. Fodor A. A survey of dimension reduction techniques. In: Technical Report. CA (US): Lawrence Livermore National Laboratory; 2002.
46. Cunningham P. Dimension reduction. In: Cord M, Cunningham P, editors. Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval. Berlin, Heidelberg: Springer; 2008:91-112.
47. Rokach L. A survey of clustering algorithms. In: Maimon O, Rokach L, editors. Data Mining and Knowledge Discovery Handbook, 2nd Edition. New York: Springer; 2011:269-298.
48. MacQueen J. Some methods for classification and analysis of multivariate observations. 1967 Presented at: Proc Fifth Berkeley Symp Math Stat Probability Vol 1; 1967; Berkeley p. 281-297 URL: https://projecteuclid.org/euclid.bsmsp/1200512992
49. Ray S, Turi R. Determination of number of clusters in K-means clustering and application in colour segmentation. 1999 Presented at: The 4th International Conference on Advances in Pattern Recognition and Digital Techniques; 1999; Victoria.
50. Ester M, Kriegal H, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. 1996 Presented at: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96); 1996; Munchen.
51. Ankerst M, Breunig M, Kriegel H, Sander J. Optics: ordering points to identify the clustering structure. 1999 Presented at: ACM SIGMOD Record; 1999; New York. [doi: 10.1145/304182.304187]
52. Madeira SC, Oliveira AL. Biclustering algorithms for biological data analysis: a survey. IEEE/ACM Trans Comput Biol Bioinform 2004;1(1):24-45. [doi: 10.1109/TCBB.2004.2] [Medline: 17048406]
53. Zhao H, Liew A, Wang D, Yan H. Biclustering analysis for pattern discovery: current techniques, comparative studies and applications. Curr Bioinform 2012 Mar 01;7(1):43-55. [doi: 10.2174/157489312799304413]
54. Aggarwal C, Han J. Frequent Pattern Mining. New York: Springer International Publishing; 2014.
55. Card SK, Mackinlay J, Shneiderman B, editors. Readings in Information Visualization: Using Vision to Think. Burlington: Moegan Kaufmann; 1999:712.
56. van der Aalst W. Process Mining, 2nd Edition. Berlin: Springer-Verlag; 2016.
57. Dowle M, Srinivasan A, Gorecki J. CRAN: package data.table. URL: https://cran.r-project.org/web/packages/data.table/ [accessed 2020-07-24]
58. Mächler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. Cluster: cluster analysis basics and extensions. 2012. URL: https://cran.r-project.org/web/packages/cluster/cluster.pdf [accessed 2020-07-24]
59. Warnes GR, Bolker B, Bonebakker L, Liaw A, Huber W, Lumley T, et al. gplots: Various R programming tools for plotting data, version 2.17.0. 2015. URL: https://CRAN.R-project.org/package=gplots [accessed 2020-07-26]
60. Duran JL. A Smart (Phone) Solution: Improving Early Intervention for Families of Low Socioeconomic Status [Thesis]. St. Louis: Washington University in St. Louis School of Medicine; 2015.
61. Ernsting C, Dombrowski SU, Oedekoven M, Kanzler M, Kuhlmey A, Gellert P. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
62. Benis A, Harel N, Barkan R, Sela T, Feldman B. Identification and description of healthcare customer communication patterns among individuals with diabetes in Clalit Health Services: a retrospective database study. Stud Health Technol Inform 2017;244:18-22. [doi: 10.3233/978-1-61499-824-2-18] [Medline: 29039369]
63. Aliotta SL, Grieve K, Giddens JF, Dunbar L, Groves C, Frey K, et al. Guided care: a new frontier for adults with chronic conditions. Prof Case Manag 2008;13(3):151-158. [doi: 10.1097/01.PCAMA.0000319968.76605.b3] [Medline: 18562909]
64. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. BMJ 2015 Feb 10;350:g7818. [doi: 10.1136/bmj.g7818] [Medline: 25670183]
65. Roosan D, Weir C, Samore M, Jones M, Rahman M, Stoddard GJ, et al. Identifying complexity in infectious diseases inpatient settings: an observation study. J Biomed Inform 2017 Jul;71S:S13-S21 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.018] [Medline: 27818310]

Abbreviations

ACG: adjusted clinical groups
Clalit: Clalit Health Services
CRI: Clalit Research Institute
DWH: data warehouse

EMR: electronic medical record
ER: emergency department
HbA_{1c}: glycated hemoglobin
HMO: health management organization
KDD: knowledge discovery in databases
NQR: nonqueue request
SES: socioeconomic status
PDC: proportion of days covered
SMS: short message service

Edited by G Eysenbach; submitted 01.12.19; peer-reviewed by A Dormanesh, D Roosan; comments to author 01.01.20; revised version received 19.02.20; accepted 15.06.20; published 11.08.20.

Please cite as:
Benis A, Barak Barkan R, Sela T, Harel N
Communication Behavior Changes Between Patients With Diabetes and Healthcare Providers Over 9 Years: Retrospective Cohort Study
J Med Internet Res 2020;22(8):e17186
URL: http://www.jmir.org/2020/8/e17186/
doi:10.2196/17186
PMID:32648555

Viewpoint

How Can Artificial Intelligence Make Medicine More Preemptive?

Usman Iqbal[1,2], PharmD, MBA, PhD; Leo Anthony Celi[3,4,5], MD, MSc; Yu-Chuan Jack Li[2,6,7,8], MD, PhD

[1]Master Program in Global Health & Development, PhD Program in Global Health & Health Security, College of Public Health, Taipei Medical University, Taipei, Taiwan

[2]International Center for Health Information Technology, Taipei Medical University, Taipei, Taiwan

[3]Laboratory for Computational Physiology, Massachusetts Institute of Technology, Cambridge, MA, United States

[4]Division of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States

[5]Department of Biostatistics, Harvard TH Chan School of Public Health, Harvard University, Boston, MA, United States

[6]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[7]Department of Dermatology, Taipei Municipal Wan-Fang Hospital, Taipei, Taiwan

[8]International Medical Informatics Association, Geneva, Switzerland

Corresponding Author:
Yu-Chuan Jack Li, MD, PhD
Graduate Institute of Biomedical Informatics
College of Medical Science and Technology
Taipei Medical University
No 172-1, Sec 2, Keelung Rd
Daan District
Taipei, 10675
Taiwan
Phone: 886 6638 2736 ext 7601
Email: jack@tmu.edu.tw

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2020/8/e23645/

Abstract

In this paper we propose the idea that Artificial intelligence (AI) is ushering in a new era of "Earlier Medicine," which is a predictive approach for disease prevention based on AI modeling and big data. The flourishing health care technological landscape is showing great potential—from diagnosis and prescription automation to the early detection of disease through efficient and cost-effective patient data screening tools that benefit from the predictive capabilities of AI. Monitoring the trajectories of both in- and outpatients has proven to be a task AI can perform to a reliable degree. Predictions can be a significant advantage to health care if they are accurate, prompt, and can be personalized and acted upon efficiently. This is where AI plays a crucial role in "Earlier Medicine" implementation.

(J Med Internet Res 2020;22(8):e17211) doi:10.2196/17211

KEYWORDS

artificial intelligence; digital health; eHealth; health care technology; medical innovations; health information technology; advanced care systems

In this paper we propose the idea that artificial intelligence (AI) is ushering in a new era of "Earlier Medicine." Advanced digital health solutions play a significant role in improving health care by enabling accurate diagnosis, prescription automation, and early prediction of diseases with potential AI capabilities [1,2]. "Earlier Medicine" refers to a temporally predictive and proactive approach to individualized health enabled by innovative AI modeling plus longitudinal/personal health big data. It calls on medical practice to not just react or manage present situations but also medical events of the foreseeable future. Such an approach will save money, lives, and our health care ecosystem from inefficiencies and disintegration.

Monitoring the trajectories of both in- and outpatients has proven to be a task AI can perform reliably [3]. Predictive tools present a considerable advantage to medicine if they are accurate,

prompt, customizable, and actionable. This is where AI plays a crucial role in “Earlier Medicine” implementation. In the late 1940s, Leavell and Clark [4] proposed three levels of prevention: primary, secondary, and tertiary. The term “primary prevention” was used to describe “measures applicable to a particular disease or group of diseases to intercept the causes of disease before they involve man” [4]. It is important to note, however, that the concept of prevention has evolved with time. Leavell and Clark's definition is disease-oriented, but the application of prevention overall extends beyond medical problems and addresses other societal concerns. Through prevention, we can create a society that fosters good health to improve the quality of care [5].

In the 1970s, Schwartz [6] speculated that by the year 2000, a sizeable amount of the thought process involved in medicine could be augmented by a division of AI known as expert systems. Recently, Topol [7] proposed that new technologies will improve the precision and accuracy of diagnosis and in doing so will enhance treatment selection. In the last 2 years, the World Health Organization (WHO) has been reviewing evidence on digital technologies from consulting experts around the world in order to assemble recommendations on ways such tools can be used for maximum impact on health care. In 2018, governments unanimously adopted a World Health Assembly resolution, calling on the WHO to develop a global digital health strategy to support national efforts to achieve universal health coverage [8]. The United States Food and Drug Administration has been proposing ways to establish greater oversight over this rapidly evolving segment of AI products to regulate these systems whose performance constantly change based on exposure to new patient data in different clinical settings [9]. Keeping these initiatives in mind, the saying, “prevention is better than cure” can be realized in a way like never before due to the advent of AI for “Earlier Medicine.” This brings us to propose the following levels of prevention—Actionable, Accurate, Timely, and Individualized (Figure 1).

Figure 1. Artificial intelligence (AI) for "Earlier Medicine" with primary, secondary, and tertiary prevention levels.

First, we propose that primary prevention using AI should be targeted at those who are well. In the case of screening all healthy individuals for risk of disease, AI can be used for earlier rather than early detection for risk reduction. However, current screening protocols are oversimplified and suffer from low compliance. More robust models are needed, for example, for the risk prediction for nonmelanoma skin cancer [10]. Screening procedures are often the first step leading to early interventions that are more cost-effective than intervening once symptoms appear. In United States, positive results from screening mammograms were of little benefit; they resulted in increases to costs, and did little to nothing to ensure quality of life and decrease mortality rate [11]. Early detection serves little purpose for patients with an illness. We believe that society would benefit more from increased precision in the selection of groups for screening with AI-based earlier risk reduction using AI prediction technology. Other examples of the possible use of AI prediction technology is in the diagnosis of some forms of melanoma from an atypical mole such as cutaneous pigmented lesion screening, using smartphone - generated images and clinical information simultaneously [12]. The incorporation of genomic information in electronic health records as part of one’s personalized treatment will drive the greater use of AI in primary prevention [13].

Secondly, we advocate that “Earlier Medicine” for secondary prevention is just as crucial. For a patient who is at risk of or suffering from a disease, AI can compute a management plan tailored to the patient’s individual needs. Secondary prevention AI includes recurrence prediction (eg, predicting non–ST-elevation myocardial infarction for patients with chest pain [14]).

Third, “Earlier Medicine” for tertiary prevention focuses on the deterrence of consequences of disease such as complications and disability. It also focuses on the overall improvement of quality of life through AI-based earlier interventions. Most countries are experiencing increases in the proportion of older people that comprise their population. It is estimated that by 2030, around 22% of the world’s population will be over 60 years of age [15]. There is an increasing prevalence of chronic disease in this age bracket globally and assisted daily living levels vary among regions [16,17]. Current health care systems that deal with this leave much to be desired in terms of long-term care, palliative care, and the expenses needed to maintain the system. Once a developed disease has been treated during its acute clinical phase, tertiary prevention seeks to soften the impact caused by the disease and potential complications on the patient's function, longevity, and quality of life. Tertiary prevention can include modifying risk factors. In cases where the condition is not reversible, tertiary prevention focuses on rehabilitation, assisting the patient to accommodate to disability [18]. For example, “Earlier Medicine” for tertiary prevention based on AI-computed individualized risk would help to prevent falls at home by 90% while reducing at least 10% of the dependent cost [19]. The key goal of tertiary prevention is to enhance quality of life by focusing on home health care services for either a short or long period as a result of illness, impaired health, old age, or other factors [20,21].

We earnestly believe that AI for “Earlier Medicine” will not only transform the practice of medicine but also radically reshape health care around the world. Harnessing the power of digital technologies is essential for achieving universal health coverage and to reviving humane medical practices for improving the quality of care [22]. The idea of prevention has evolved over the years from primary to tertiary and from a doctor-driven to patient-centered care model. AI for “Earlier Medicine” can serve as virtual medical assistants for clinicians, allowing for the resurgence of empathy-based care [23].

Acknowledgments

LAC is funded by the National Institutes of Health through NIBIB R01 EB017205.

Conflicts of Interest

None declared.

References

1. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ 2019 Oct;7:e7702 [FREE Full text] [doi: 10.7717/peerj.7702] [Medline: 31592346]
2. Kukafka R. Digital Health Consumers on the Road to the Future. J Med Internet Res 2019 Nov 21;21(11):e16359 [FREE Full text] [doi: 10.2196/16359] [Medline: 31750835]
3. Parreco J, Hidalgo A, Parks JJ, Kozol R, Rattan R. Using artificial intelligence to predict prolonged mechanical ventilation and tracheostomy placement. J Surg Res 2018 Aug;228:179-187. [doi: 10.1016/j.jss.2018.03.028] [Medline: 29907209]
4. Top FH. Preventive Medicine for the Doctor in His Community: An Epidemiologic Approach. Arch Intern Med 1959 Jan 01;103(1):164. [doi: 10.1001/archinte.1959.00270010170027]
5. Glossary of terms used in the "Health for All" series, no. 1-8. World Health Organization. 1984. URL: https://apps.who.int/iris/handle/10665/39565 [accessed 2020-07-02]
6. Schwartz WB. Medicine and the Computer. N Engl J Med 1970 Dec 03;283(23):1257-1264. [doi: 10.1056/nejm197012032832305]
7. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. UK: Hachette; Mar 12, 2019.
8. Seventy-first World Health Assembly update, 25 May. World Health Organization. 2018 May 25. URL: https://www.who.int/news-room/detail/25-05-2018-seventy-first-world-health-assembly-update-25-may [accessed 2020-07-02]
9. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SAMD): discussion paper and request for feedback. United States Food and Drug Administration. 2019. URL: https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf [accessed 2020-07-02]
10. Wang H, Wang Y, Liang C, Li Y. Assessment of Deep Learning Using Nonimaging Information and Sequential Medical Records to Develop a Prediction Model for Nonmelanoma Skin Cancer. JAMA Dermatol 2019 Sep 04;155(11):1277. [doi: 10.1001/jamadermatol.2019.2335] [Medline: 31483437]
11. Morris E, Feig SA, Drexler M, Lehman C. Implications of Overdiagnosis: Impact on Screening Mammography Practices. Popul Health Manag 2015 Sep;18 Suppl 1(S1):S3-11. [doi: 10.1089/pop.2015.29023.mor] [Medline: 26414384]
12. Chin Y, Hou ZY, Lee MY, Chu HM, Wang HH, Lin YT, et al. A patient-oriented, general-practitioner-level, deep-learning-based cutaneous pigmented lesion risk classifier on a smartphone. Br J Dermatol 2020 Jun 01;182(6):1498-1500. [doi: 10.1111/bjd.18859] [Medline: 31907926]
13. Hazin R, Brothers KB, Malin BA, Koenig BA, Sanderson SC, Rothstein MA, et al. Ethical, legal, and social implications of incorporating genomic information into electronic health records. Genet Med 2013 Oct 12;15(10):810-816 [FREE Full text] [doi: 10.1038/gim.2013.117] [Medline: 24030434]

14. Wu C, Hsu W, Islam MM, Poly TN, Yang H, Nguyen P, et al. An artificial intelligence approach to early predict non-ST-elevation myocardial infarction patients with chest pain. Comput Methods Programs Biomed 2019 May;173:109-117. [doi: 10.1016/j.cmpb.2019.01.013] [Medline: 31046985]
15. Department of Economic and Social Affairs, Population Division. World Population Ageing 2019. New York, NY: United Nations; 2019. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2019-Highlights.pdf [accessed 2020-07-02]
16. Scheel-Hincke LL, Möller S, Lindahl-Jacobsen R, Jeune B, Ahrenfeldt LJ. Cross-national comparison of sex differences in ADL and IADL in Europe: findings from SHARE. Eur J Ageing 2020 Mar 14;17(1):69-79. [doi: 10.1007/s10433-019-00524-y] [Medline: 32158373]
17. Kuo H, Lin K, Lan C, Li I. Activities of daily living trajectories among institutionalised older adults: A prospective study. J Clin Nurs 2017 Dec 23;26(23-24):4756-4767. [doi: 10.1111/jocn.13828] [Medline: 28334483]
18. Cattelani L, Palumbo P, Palmerini L, Bandinelli S, Becker C, Chesani F, et al. FRAT-up, a Web-based fall-risk assessment tool for elderly people living in the community. J Med Internet Res 2015 Feb 18;17(2):e41 [FREE Full text] [doi: 10.2196/jmir.4064] [Medline: 25693419]
19. Gomez F, Curcio CL, Brennan-Olsen SL, Boersma D, Phu S, Vogrin S, et al. Effects of the falls and fractures clinic as an integrated multidisciplinary model of care in Australia: a pre-post study. BMJ Open 2019 Jul 29;9(7):e027013 [FREE Full text] [doi: 10.1136/bmjopen-2018-027013] [Medline: 31362962]
20. Kang H, Han J, Kwon GH. An Ecological Approach to Smart Homes for Health Care Services: Conceptual Framework of a Smart Servicescape Wheel. JMIR Mhealth Uhealth 2019 Feb 08;7(2):e12425 [FREE Full text] [doi: 10.2196/12425] [Medline: 30735145]
21. Willis M, Duckworth P, Coulter A, Meyer ET, Osborne M. The Future of Health Care: Protocol for Measuring the Potential of Task Automation Grounded in the National Health Service Primary Care System. JMIR Res Protoc 2019 Apr 09;8(4):e11232 [FREE Full text] [doi: 10.2196/11232] [Medline: 30964437]
22. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med 2019 Apr 04;380(14):1347-1358. [doi: 10.1056/nejmra1814259]
23. Shen J, Zhang CJP, Jiang B, Chen J, Song J, Liu Z, et al. Artificial Intelligence Versus Clinicians in Disease Diagnosis: Systematic Review. JMIR Med Inform 2019 Aug 16;7(3):e10010 [FREE Full text] [doi: 10.2196/10010] [Medline: 31420959]

Abbreviations

AI: artificial intelligence
WHO: World Health Organization

Edited by G Eysenbach; submitted 26.11.19; peer-reviewed by A Hussain, D Gunasekeran; comments to author 20.05.20; revised version received 05.06.20; accepted 25.06.20; published 11.08.20.

Please cite as:
Iqbal U, Celi LA, Li YCJ
How Can Artificial Intelligence Make Medicine More Preemptive?
J Med Internet Res 2020;22(8):e17211
URL: https://www.jmir.org/2020/8/e17211
doi:10.2196/17211
PMID:32780024

Original Paper

Artificial Intelligence for Rapid Meta-Analysis: Case Study on Ocular Toxicity of Hydroxychloroquine

Matthew Michelson[1,2], PhD; Tiffany Chow[1], PharmD; Neil A Martin[3], MD; Mike Ross[1], BS; Amelia Tee Qiao Ying[1]; Steven Minton[2], PhD

[1]Evid Science, El Segundo, CA, United States

[2]InferLink, El Segundo, CA, United States

[3]Pacific Neuroscience Institute, Providence St John's Health Center, Santa Monica, CA, United States

Corresponding Author:
Matthew Michelson, PhD
Evid Science
2361 Rosencrans Ave Ste 348
El Segundo, CA, 90245-4929
United States
Phone: 1 626 765 1903
Email: mmichelson@evidscience.com

Abstract

Background: Rapid access to evidence is crucial in times of an evolving clinical crisis. To that end, we propose a novel approach to answer clinical queries, termed rapid meta-analysis (RMA). Unlike traditional meta-analysis, RMA balances a quick time to production with reasonable data quality assurances, leveraging artificial intelligence (AI) to strike this balance.

Objective: We aimed to evaluate whether RMA can generate meaningful clinical insights, but crucially, in a much faster processing time than traditional meta-analysis, using a relevant, real-world example.

Methods: The development of our RMA approach was motivated by a currently relevant clinical question: is ocular toxicity and vision compromise a side effect of hydroxychloroquine therapy? At the time of designing this study, hydroxychloroquine was a leading candidate in the treatment of coronavirus disease (COVID-19). We then leveraged AI to pull and screen articles, automatically extract their results, review the studies, and analyze the data with standard statistical methods.

Results: By combining AI with human analysis in our RMA, we generated a meaningful, clinical result in less than 30 minutes. The RMA identified 11 studies considering ocular toxicity as a side effect of hydroxychloroquine and estimated the incidence to be 3.4% (95% CI 1.11%-9.96%). The heterogeneity across individual study findings was high, which should be taken into account in interpretation of the result.

Conclusions: We demonstrate that a novel approach to meta-analysis using AI can generate meaningful clinical insights in a much shorter time period than traditional meta-analysis.

(J Med Internet Res 2020;22(8):e20007) doi:10.2196/20007

KEYWORDS

meta-analysis; rapid meta-analysis; artificial intelligence; drug; analysis; hydroxychloroquine; toxic; COVID-19; treatment; side effect; ocular; eye

Introduction

The capacity of artificial intelligence (AI) to aggregate and process massive volumes of information is emerging as particularly crucial in the current moment, especially as the large amount of data available can be overwhelming for humans to evaluate [1]. AI technology can alleviate the burden of some of this overload by automatically processing the written text of medical papers, and converting the text into a more consumable, structured set of data that can be easily searched and analyzed. Essentially, AI turns all of the written articles into spreadsheets of results.

Further, although meta-analysis and systematic literature review are the gold standards for evidence [2], these analyses require significant time and effort to produce (often as long as 1 year [3]) and are therefore rarely updated [4,5].

Therefore, to produce this evidence in a more timely manner, we here propose the rapid meta-analysis (RMA). An RMA follows the same general framework methodology of a traditional meta-analysis, but leverages technology at each step, yielding a much faster time to production. Some data quality may be compromised due to the emphasis on fast time to production, but the ability to generate answers so quickly may warrant this tradeoff.

We were motivated to develop the RMA method based on a practical example of the current need for obtaining a rapid consensus on evidence from the literature. Hydroxychloroquine has been available since the 1950s [6] and has been used to treat malaria, lupus erythematosus, and rheumatoid arthritis. Most recently, hydroxychloroquine has been highlighted as a potential intervention to support patients with coronavirus disease (COVID-19). Although the efficacy outcomes of hydroxychloroquine are different in each clinical condition for which it is used, adverse events tend to be consistent. In this study, we used RMA to answer a specific clinical question regarding hydroxychloroquine and the degree to which ocular toxicity is a side effect. This is an important clinical question; however, we were not able to find a suitable aggregation of results.

The core innovation of an RMA is replacing as many of the steps of manual meta-analysis as possible with machine intelligence, as has been proposed previously [4,7]. Machines are not yet at the point where they can simply provide an answer to a posed question; therefore, RMA instead replaces as many manual steps as possible with machine assistance (or entirely AI). The goal is that each step could eventually be replaced with AI.

Figure 1 provides a schematic to make this idea more concrete. The left of the figure shows the standard steps (at a high level) for meta-analysis and the right side shows the equivalent steps with technology replacement.

For this RMA, we leveraged the Evid Science clinical outcomes database [8] for searching and screening (although any suitable AI system could provide a similar benefit). This database was built using the Evid Science AI, which is capable of turning written text of results into a "structured" representation (eg, a row in a database or spreadsheet).

Figure 2 shows a sentence from an article about toxicity detected for a set of patients, which has been parsed by the Evid Science AI. The AI was able to break this sentence down into fields (such as result, intervention, and outcome) automatically. In particular, it knows that 18 is the number of patients, and since that represents 30.5% of the patients, it must be 18 of 59. It also knows that 18 was associated with "Retinal toxicity being detected" in contrast to 5, which is associated with "color vision impairments."

Previous AI-related approaches have attempted to identify sentences associated with Patient/Problem, Intervention, Comparison, Outcome (PICO) parameters from studies [9,10], surface more relevant articles for screening [11], and even study characteristics, including bias [12]. However, we were not able to find another AI that was purposely built to parse the full, numeric results from the text directly (eg, numbers and their associated fields), which are the inputs required for an advanced investigation such as a meta-analysis.

To train the AI to perform this task, researchers at Evid Science employed supervised machine learning. In this methodology, the researchers initially gave the system very explicit examples of the type of output they wanted; similar to the format shown in Figure 2, these comprised sets of sentences and the associated structured results.

The machine was then trained with a dataset of thousands of such examples from a wide variety of articles in the literature. The learning process enables the machine to produce these types of output for brand new sentences. To be clear, the articles chosen for training were selected from multiple disease topics and with various interventions, and were not only focused on hydroxychloroquine. As the system improves, it can even be taught to correct mistakes, rather than having to start with fresh examples each time, thereby limiting the effort involved in refining its learning.

Figure 1. Traditional meta-analysis (simplified, left) vs rapid meta-analysis using artificial intelligence (AI) (right).

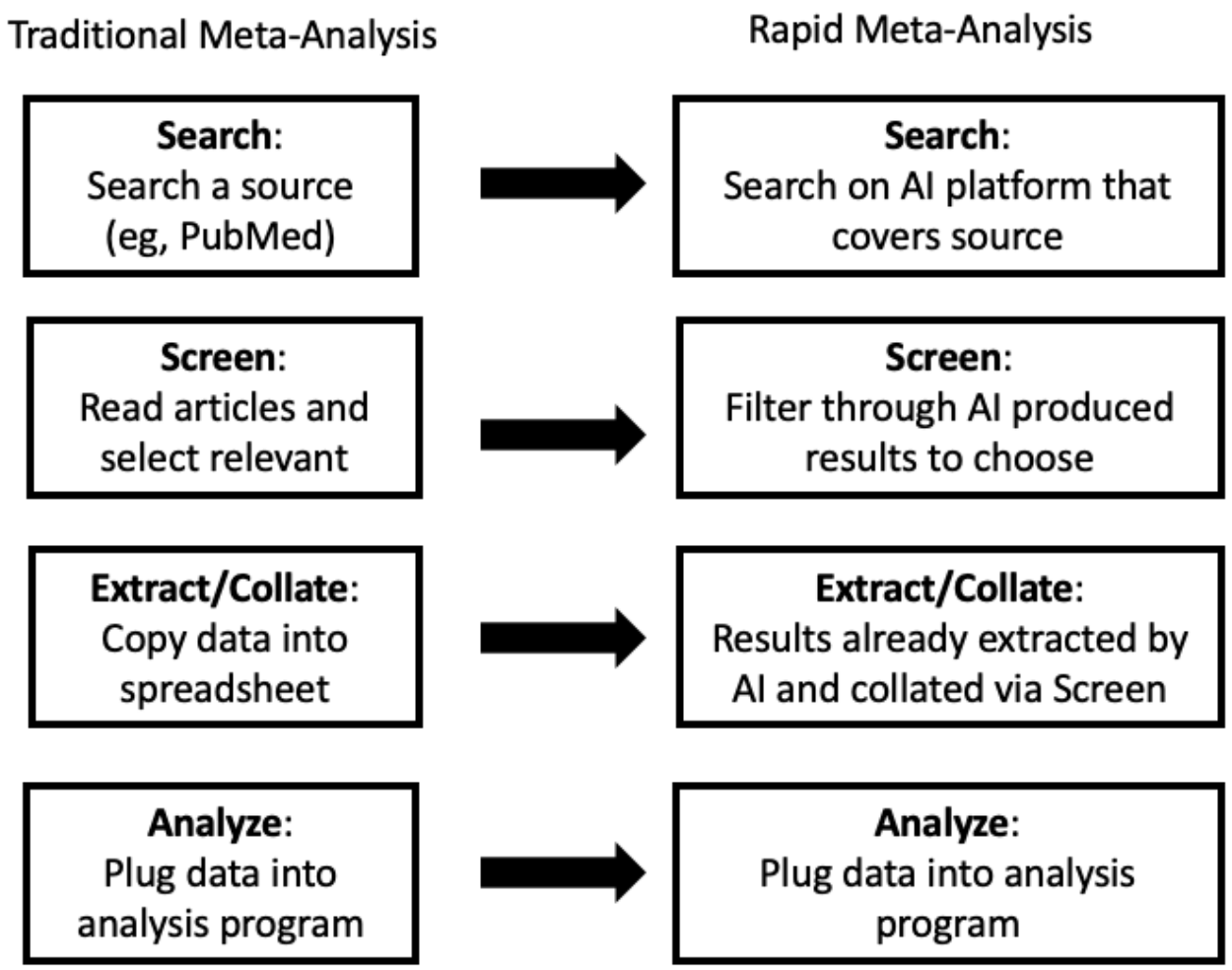


Figure 2. Example of artificial intelligence–generated results from article text.

Retinal toxicity was detected in 18 (30.5%) of the patients, and 5 (8.5 %) developed color vision impairments.

Result	Intervention	Outcome
18 of 59	hydroxychloroquine	Retinal toxicity was detected
5 of 59	hydroxychloroquine	developed color vision impairments

Methods

Evid Science AI

The Evid Science AI is a deep-learning model, written in python, constructed from layers of transformers and bidirectional long-short term memory (bi-LSTM) units. Our model first encodes the inputs using the transformer language model (SciBERT [13]), which turns the words into a mathematical space where similar words are grouped together. These embedded inputs are then passed along through the bi-LSTM layers of the network, which traverse words in the text and labels them appropriately. We trained our algorithm on 24,614 labeled records.

Model Performance

Recently, we performed a dual-annotator analysis of extraction accuracy. One hundred results were randomly selected from the database, each of which contains a result record (eg, numerator, denominator, percent, measurement value, unit, intervention/study group, outcome) and a sentence. Of note, a single sentence can be associated with multiple results; however, in randomizing, we chose one result to label for accuracy. Further, not all sentences have values for all fields. We then labeled the extracted fields (numerator, denominator, percent, measurement value, unit, intervention/study group, outcome) for accuracy. Our labels are provided as a spreadsheet in Multimedia Appendix 1.

The labels could be "perfect" (eg, a field was perfectly extracted); "near perfect" (eg, the field contained extra words or missed a few words, but was otherwise understandable, such as an outcome of "attained remission" contains the extra word "attained"); or "incorrect." We also included "missing" as a means for estimating recall (true recall is hard to measure, given that we would require full labeling of all documents). The 100 sentences and labels are shown in Multimedia Appendix 1.

Specifically, for each field, we report the estimated recall, precision (which is accuracy), perfect precision (accuracy only considering perfect extractions), and F-measure (harmonic mean between recall and precision).

Using the Model for RMA

From a practical perspective, in previous work, we demonstrated a similar process to RMA using the Evid Science AI to replicate the results from a systematic literature review [14]. Crucially, by leveraging AI, we produced the results in 6 days rather than the months it took to produce the original. In addition, given the time between the original publication and AI-assisted version, 22 new relevant results had been published. Of note, the current version of the AI used for this RMA is significantly more powerful than the version previously used for systematic literature review replication.

The Evid Science clinical outcomes database used in this RMA is the result of running the AI over the entirety of the publicly (freely) available medical literature (PubMed). The current database has nearly 70,000,000 "facts" associated with results from articles, which users can search and screen through.

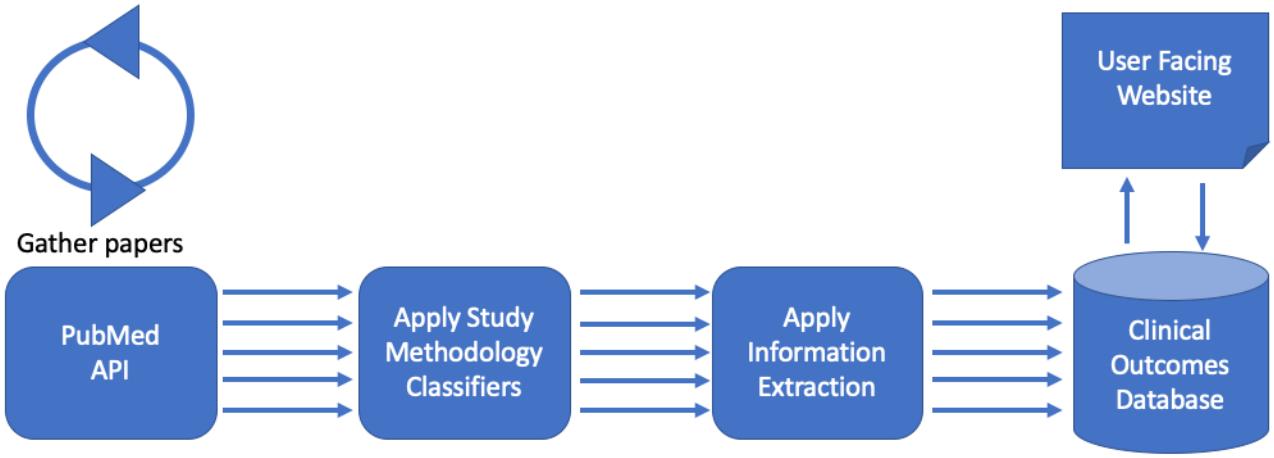
The Evid Science platform has already indexed the entirety of PubMed, and each night, it pulls in the latest papers. The architecture for the system is shown in Figure 3. Starting at the left, articles come into the system via the PubMed application programming interface (API). Machine-learning classifiers are then applied to articles, determining the study type (eg, trial, observational study) and other methodology information. The extraction algorithm described above is then applied to the article. Each result from the extraction algorithm is then stored in our database and users can retrieve these results via searching and filtering via a web-based user interface. Final results can be exported (in CSV format) so that they can be analyzed in sophisticated statistical programs such as R.

Therefore, our RMA proceeds by searching and screening through this database, as described below. The search itself (Step 1) leverages PubMed APIs, and therefore returns equivalent articles to PubMed. That is, any search on our platform is passed to the PubMed API, and the returned articles are then matched against what our AI has extracted. Therefore, the initial search results are equivalent.

Screening is then simplified, since the AI has processed the text into structured records that can be filtered and screened efficiently. For instance, we can simply filter results associated with "toxicity" in the outcome (or other outcomes of interest). This is more efficient than manually reading each returned abstract, since one only screens articles in the filtered set.

After searching and screening, a user obtains the final dataset for analysis (which the AI helped to produce via extraction). One can perform many analyses directly within the Evid Science web-based tool or export data to Excel and then analyze it with other programs (as we have done).

Figure 3. Evid Science artificial intelligence architecture. API: application programming interface.



Results

RMA Process for the Association of Eye Issues With Hydroxychloroquine Use

We initially performed a search for hydroxychloroquine on the Evid Science platform, and filtered down to results in which the outcome was major vision impairments (eg, "maculopathy," "blind," "toxicity"). In this study, we focused solely on PubMed abstracts, since they are freely available. This yielded 22 candidate articles from a possible set of 5010 articles related to hydroxychloroquine, 1352 of which were identified as primary studies (eg, clinical trial or observational study) by our AI and were therefore included as possible articles to process results from.

After screening, we were left with 11 papers for our RMA; the other 11 excluded articles were published before 2000 or focused on the diagnosis of ocular issues. The search took less than 1

minute and the screening took 22 minutes. Most of the work involved selecting the papers and lightly cleaning the results to make the table of results easier to read. Two results required "more significant" human intervention: one result was reported as having "all cases" of documented blindness attributed to causes other than hydroxychloroquine, and we therefore needed to invert this result to be 0 cases attributed to hydroxychloroquine; the other result had a misattributed denominator, which was manually fixed. All other changes involved removing single words, which was required very infrequently in concordance with the accuracy results shown in Table 1.

The results from the search and screening processes are shown in Table 2, which served as the input for our meta-analysis computation. Although we did not conduct the equivalent screening manually, in previous work, we were able to use our AI to match a published systematic literature review on inflammatory bowel disease [14], and therefore have already demonstrated that we can generate the equivalent screening using our tool to that obtained with a manual process.

Table 1. Extraction results based on 100 randomly selected results, dual-screened.

Result extracted	Recall	Precision	Perfect Precision	F-measure	F-measure (Perfect)
Numerator (N=42)	92.86%	95.12%	95.12%	93.98%	93.98%
Denominator (N=23)	91.30%	91.30%	91.30%	91.30%	91.30%
Percent (N=40)	90.00%	94.74%	92.11%	92.31%	91.04%
Continuous Value (N=29)	89.66%	92.86%	92.86%	91.23%	91.23%
Continuous Unit (N=24)	83.33%	95.24%	90.48%	88.89%	86.76%
Intervention (N=85)	84.71%	94.74%	86.84%	89.44%	85.76%
Outcome (N=100)	95.00%	97.94%	79.38%	96.45%	86.49%

Table 2. Results from included papers.

Events (n)	Patients (n)	Intervention	Outcome	Reference
2	400	patients who were treated with recommended dosages of the drug for a mean of 8.7 years	incidence of hydroxychloroquine-related retinopathy	Mavrikakis et al [15]
0	526	hydroxychloroquine	retinal toxicity was noted during the first 6 years of treatment	Mavrikakis et al [15]
46	845	chloroquine, hydroxychloroquine, or both	ophthalmological alterations, confirmed by the ophthalmological examination	Spinelli et al [16]
3	12	800 mg/day hydroxychloroquine	developed retinal toxicity with scotomas in the Amsler grid and Humphrey 10-2 automated perimetry, as well as abnormal multifocal electroretinography	Navajas et al [17]
0	11	long-term hydroxychloroquine	documented blindness, in all cases attributed to a cause other than hydroxychloroquine-related ocular toxicity	Singh et al [18]
35	678	hydroxychloroquine	had hydroxychloroquine toxicity	Chiu et al [19]
1	121	hydroxychloroquine	prevalence of toxic retinopathy	Cabral et al [20]
18	59	hydroxychloroquine	retinal toxicity was detected	Espandar et al [21]
9	778	antimalarial drugs	suffered definite presence of antimalarial retinopathy	Jover et al [22]
11	36	hydroxychloroquine	had abnormal response densities in one or both eyes	Maturi et al [23]
3	26	hydroxychloroquine	results from electrophysiological and clinical evaluation, toxicity (bull's eye maculopathy)	Tzekov et al [24]
4	93	chloroquine and hydroxychloroquine therapy	developed typical bull's eye maculopathy	Neubauer et al [25]

RMA Outcome

We then performed a meta-analysis of the results from the screened out articles using a generalized linear mixed model (in R), as these are binary occurrences of having an eye issue. We chose a random-effects model for the analysis, demonstrating a result of 3.4 events of eye issues per 100 observations (95% CI 1.11-9.96). The code for this analysis

was already written; therefore, plugging in the data (Table 2) and running it took roughly 2 minutes, including exporting the data to Excel, renaming and selecting columns to conform to R code input, and running the code.

The forest plot of the meta-analysis is shown in Figure 4. Clearly, there was heterogeneity (I^2=97%) among studies; therefore, these results warrant deeper inspection and cautious interpretation. The funnel plots of the results are shown in Figure 5. Each step of our RMA, its output, and its timing are shown in Figure 6, showing that altogether the RMA process took roughly 30 minutes to complete.

Figure 4. Forest plot for our rapid meta-analysis.

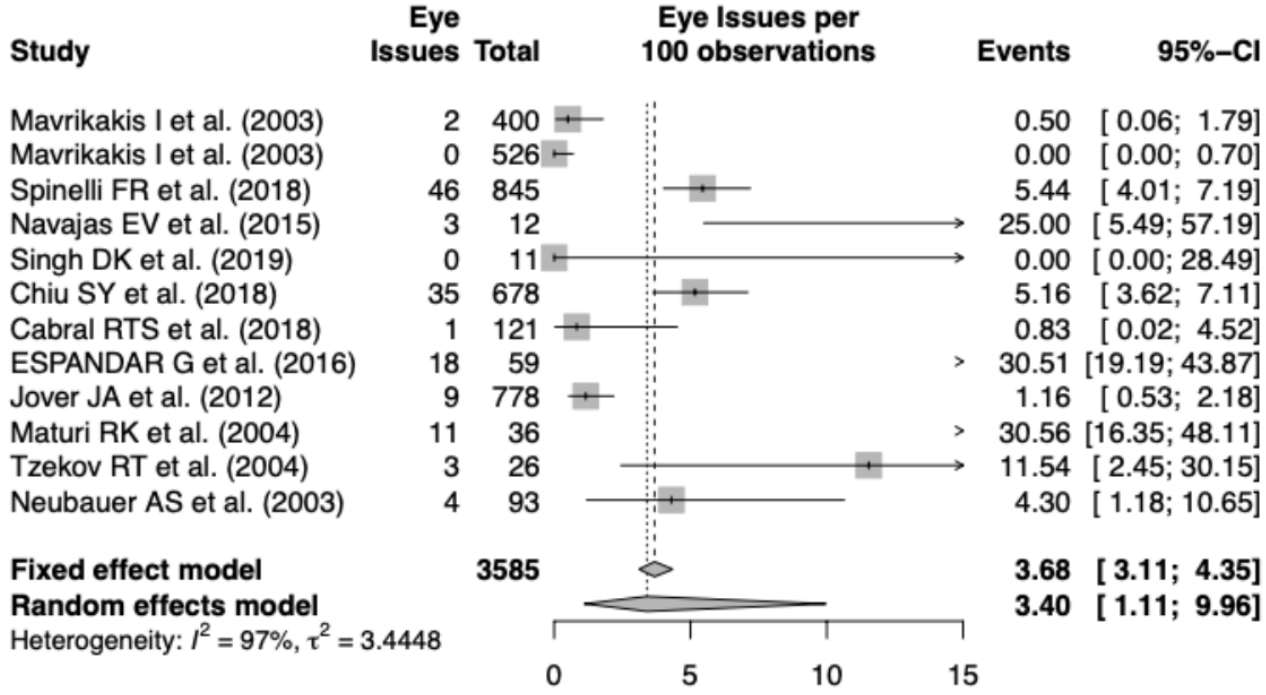


Figure 5. Funnel plot for our rapid meta-analysisis.

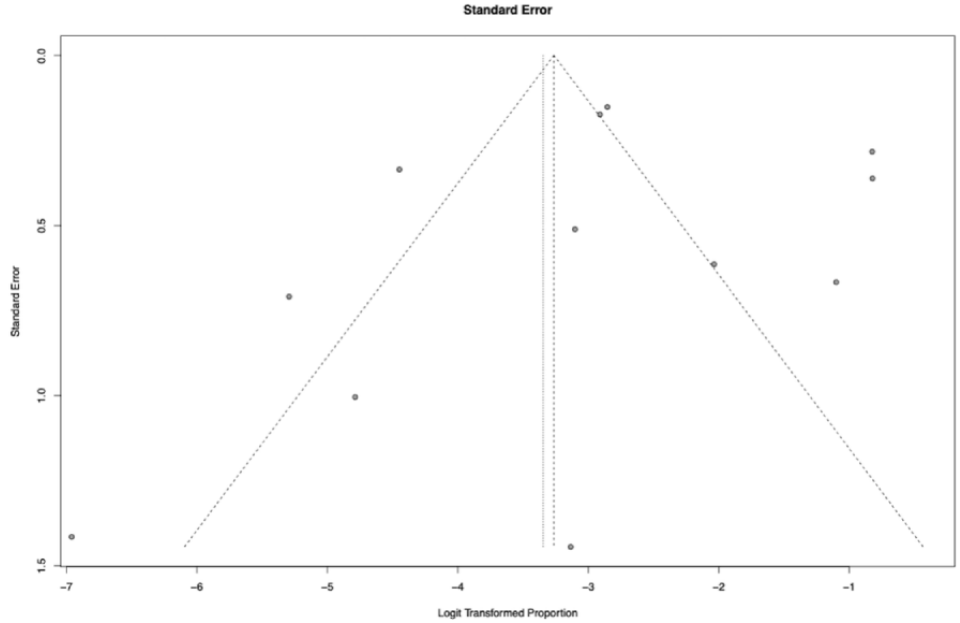
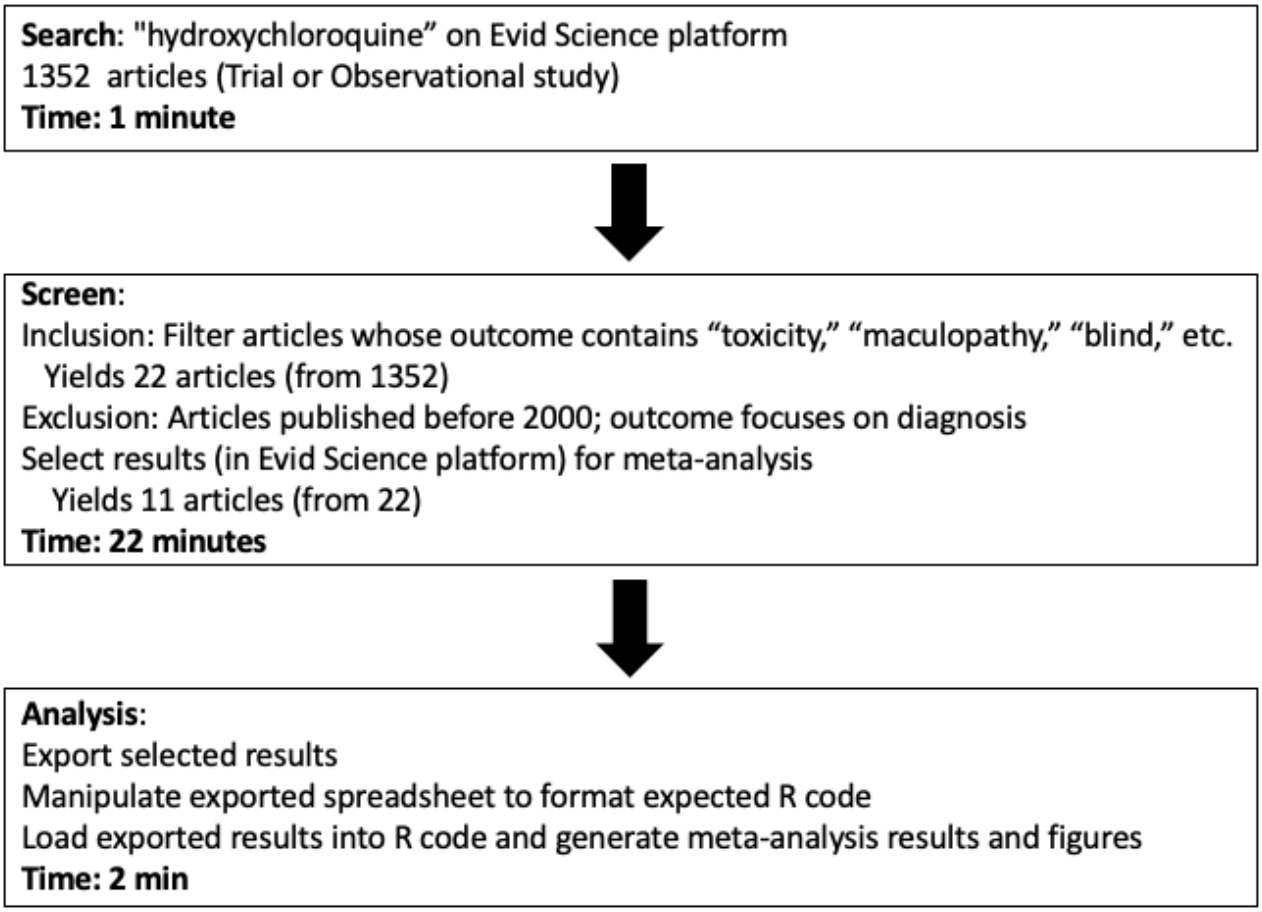


Figure 6. Overview of the rapid meta-analysis for ocular toxicity associated with hydroxychloroquine.



Discussion

Principal Findings

For the RMA, we leveraged the Evid Science clinical outcomes database to find relevant studies, screened the database for studies focused on hydroxychloroquine and vision issues, and then performed the meta-analysis computation. Most crucially, the entire process from the search to analysis took less than 30 minutes. Based on results from 11 studies (N=3585), we could expect to see major eye issues 3.4% (95% CI: 1.11%-9.96%) of the time when using hydroxychloroquine. We note the high heterogeneity across the studies (I^2=97%), requiring caution when interpreting these results. Notably, an RMA such as the present analysis is meant to raise awareness, and should not be treated as a full systematic literature review or meta-analysis that should strictly guide treatment. The data and results presented herein are current as of April 11, 2020.

The screening of articles step is one of the key areas of time savings in RMA. Although we described the various accuracy metrics in Table 1, the AI is also able to surface study characteristics that can be helpful in screening for meta-analysis as well. For example, Table 3 shows various study characteristics for a few of the chosen articles, including time period (retrospective or prospective), cohort focus (groups designated based on different drugs or conditions), study type (trial or observational), and finally methodology sentences. These characteristics are all generated by the machine, except for the methodology sentences, which are surfaced from the text automatically (rather than applied as "tags" to an article). In our RMA, we chose to include as many data points as possible, which perhaps led to our high level of heterogeneity (although population size differences clearly influence heterogeneity as well), but we could have been more specific, focusing on certain study characteristics using the values supplied by the AI. For instance, the methodology sentences include geographic locations, or we could have focused solely on papers that group patients by drug, rather than condition. For a complete list of characteristics for all articles, refer to Multimedia Appendix 2.

Table 3. Study characteristics, pulled via artificial intelligence (see Multimedia Appendix 2 for the characteristics of all papers).

Methodology Sentences	Time period	Cohort focus	Study type	Reference
The incidence of irreversible retinal toxicity in patients treated with hydroxychloroquine: a reappraisal. To define the risk of hydroxychloroquine (HCQ)-related retinal toxicity in patients with rheumatoid arthritis (rheumatoid arthritis) and systemic lupus erythematosus (systemic lupus erythematosus) who are receiving recommended dosages of the drug (< or =6.5 mg/kg/day). Prospective cohort study, from 1985 to 2000. Greek patients with rheumatoid arthritis (n=335) and systemic lupus erythematosus (n=191) treated with hydroxychloroquine, 400 of whom had completed at least 6 years of treatment.	Prospective	Drug Therapy	Observational	Mavrikakis et al [15]
Treating lupus patients with antimalarials: analysis of safety profile in a single-center cohort. This longitudinal retrospective study aims at describing the safety profile and the reasons for discontinuation of antimalarials in patients with systemic lupus erythematosus (systemic lupus erythematosus) and discoid lupus erythematosus (discoid lupus erythematosus), focusing on ocular toxicity. We analyzed the clinical data of 845 systemic lupus erythematosus and discoid lupus erythematosus patients; 59% of them were taking antimalarials: 1.4% chloroquine (chloroquine), 88.5% hydroxychloroquine (hydroxychloroquine) and 10.1% both.	Retrospective	Drug Therapy	Observational	Spinelli et al [16]
Retinal toxicity of high-dose hydroxychloroquine in patients with chronic graft-versus-host disease. To evaluate retinal toxicity in patients treated with high-dose hydroxychloroquine (hydroxychloroquine) (Plaquenil, Sanofi Pharmaceuticals) for chronic graft-versus-host disease (graft versus host disease). Twelve patients with chronic graft versus host disease treated with 800 mg/day hydroxychloroquine between June 2005 and December 2010.	Retrospective	Drug Therapy	Observational	Navajas et al [17]

Limitations

Of course, there are limitations to our approach as well. One important aspect to note is that although RMA can very rapidly produce answers to clinical questions, nothing (yet) replaces human ingenuity and creativity (and most importantly, common sense). A major limitation of RMA currently is that the AI is not sophisticated enough to present more than data and mathematical results; that is, it cannot make meaningful interpretations.

In this case, for instance, the result is indicative (3.4 events per 100 observations), but the confidence interval is wide and the I^2 is high. Therefore, the result of our RMA warrants cautious interpretation. A machine cannot produce this nuance summary but can only provide the data and results for a person to then interpret.

Another limitation is that models may make mistakes. Of course, human beings make mistakes as well, but model mistakes can be counterintuitive. For instance, in this study, one of the results extracted was that all cases had documented blindness associated with something other than hydroxychloroquine. This implies zero cases for hydroxychloroquine, but the machine did not pick up on this. It is obvious to us as humans that the inverse result is what we want, but that is a common sense observation. Therefore, there is a tradeoff between assuming there will be some mistakes and the rapid nature of RMA. We note that there are often mistakes in human analysis as well.

A final limitation is acceptability. We introduce RMA as a means to more rapidly produce evidence that can be helpful in clinical decision making. However, without trust in and the adoption of AI-assisted evidence, such results might exist in a vacuum. If that is the case, clinical practice will not benefit from the advancement. Therefore, this limitation necessitates that, to be useful, current AI-generated evidence must be accepted in some manner. We hope that our transparency in this article (presenting results and data) helps to bring some change in this regard.

Conclusion

In this article, we have presented a new framework for answering clinical questions when time is at a premium and can be traded off for data quality. We call this approach RMA, and we demonstrated its utility in answering a clinical question about ocular toxicity associated with hydroxychloroquine as a proposed treatment for COVID-19. By leveraging RMA, in roughly 30 minutes we were able to discern a potential association with an incidence of 3.4 events per 100 observations (95% CI 1.11-9.96). Although the results raise further questions that need to be considered (eg, regarding the high heterogeneity), they nevertheless raise attention to a relevant clinical issue with the drug hydroxychloroquine. Importantly, the whole assessment was completed in less than 30 minutes, representing huge time savings compared to the months it takes for traditional meta-analysis conducted by hand.

Conflicts of Interest

MM, TC, MR, ATQY, and SM are all employees, former employees, or board members of Evid Science.

Multimedia Appendix 1
Labels from dual-annotator labeling of extractions.
[XLSX File (Microsoft Excel File), 25 KB - jmir_v22i8e20007_app1.xlsx]

Multimedia Appendix 2
The full study characteristics for Table 3.
[XLSX File (Microsoft Excel File), 12 KB - jmir_v22i8e20007_app2.xlsx]

References

1. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med 2010 Sep 21;7(9):e1000326 [FREE Full text] [doi: 10.1371/journal.pmed.1000326] [Medline: 20877712]
2. Sackett D, Straus S, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM, 2nd ed. Edinburgh: Churchill Livingstone; 2000.
3. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. 2011. URL: https://handbook-5-1.cochrane.org/ [accessed 2020-04-10]
4. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. Syst Rev 2014 Jul 09;3(1):74-74 [FREE Full text] [doi: 10.1186/2046-4053-3-74] [Medline: 25005128]
5. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med 2007 Aug 21;147(4):224-233. [doi: 10.7326/0003-4819-147-4-200708210-00179] [Medline: 17638714]
6. Loughlin EH, Rice JB, Wells HS, Rappaport I, Joseph AA. The treatment of Plasmodium falciparum malaria with a single dose antimalarial; a preliminary report of the use of hydroxychloroquine, 7-chloro-4(4-(N-ethyl-N-B-hydroxyethylamino)-1-methylbutylamino)-quinoline diphosphate. Antibiot Chemother (Northfield) 1952 Apr;2(4):171-174. [Medline: 24541823]
7. Michelson M. Automating Meta-Analyses of Randomized Clinical Trials: A First Look. 2014 Jan 01 Presented at: AAAI Fall Symposium on Expanding the Boundaries of Health Informatics Using AI (HIAI); 2014; Arlington, VA.
8. EvidScience. URL: https://www.evidscience.com/ [accessed 2020-04-11]
9. Wallace BC, Kuiper J, Sharma A, Zhu MB, Marshall IJ. Extracting PICO Sentences from Clinical Trial Reports using Supervised Distant Supervision. J Mach Learn Res 2016;17:132 [FREE Full text] [Medline: 27746703]
10. Scells H, Zuccon G, Koopman B, Deacon A, Azzopardi L, Geva S. Integrating the Framing of Clinical Questions via PICO into the Retrieval of Medical Literature for Systematic Reviews. 2017 Nov 06 Presented at: ACM on Conference on Information and Knowledge Management; 2017; Singapore.
11. Przybyła P, Brockmeier AJ, Kontonatsios G, Le Pogam M, McNaught J, von Elm E, et al. Prioritising references for systematic reviews with RobotAnalyst: A user study. Res Synth Methods 2018 Sep 30;9(3):470-488 [FREE Full text] [doi: 10.1002/jrsm.1311] [Medline: 29956486]
12. Marshall I, Kuiper J, Banner E, Wallace B. Automating Biomedical Evidence Synthesis: RobotReviewer. 2017 Presented at: Annual Meeting of the Association of Computational Linguistics, System Demonstrations; July 30, 2017; Vancouver, Canada. [doi: 10.18653/v1/p17-4002]
13. scibert. GitHub. URL: https://github.com/allenai/scibert [accessed 2020-07-30]
14. Michelson M, Ross M, Minton S. AI2 Leveraging machine-assistance to replicate a systematic review. Value in Health 2019 May;22:S34. [doi: 10.1016/j.jval.2019.04.006]
15. Mavrikakis I, Sfikakis PP, Mavrikakis E, Rougas K, Nikolaou A, Kostopoulos C, et al. The incidence of irreversible retinal toxicity in patients treated with hydroxychloroquine: a reappraisal. Ophthalmology 2003 Jul;110(7):1321-1326. [doi: 10.1016/S0161-6420(03)00409-3] [Medline: 12867385]
16. Spinelli FR, Moscarelli E, Ceccarelli F, Miranda F, Perricone C, Truglia S, et al. Treating lupus patients with antimalarials: analysis of safety profile in a single-center cohort. Lupus 2018 Sep 28;27(10):1616-1623. [doi: 10.1177/0961203318781008] [Medline: 29954281]
17. Navajas EV, Krema H, Hammoudi DS, Lipton JH, Simpson ER, Boyd S, et al. Retinal toxicity of high-dose hydroxychloroquine in patients with chronic graft-versus-host disease. Can J Ophthalmol 2015 Dec;50(6):442-450. [doi: 10.1016/j.jcjo.2015.08.003] [Medline: 26651304]
18. Singh D, Muhieddine L, Einstadter D, Ballou S. Incidence of blindness in a population of rheumatic patients treated with hydroxychloroquine. Rheumatol Adv Pract 2019;3(1):rkz009-rkz009 [FREE Full text] [doi: 10.1093/rap/rkz009] [Medline: 31431997]
19. Chiu S, Shaw J, Luong T, Fong D, Modjtahedi B. Coding patterns used by ophthalmologists for hydroxychloroquine retinal toxicity. Clin Ophthalmol 2018;12:2261-2265. [doi: 10.2147/OPTH.S170789] [Medline: 30464387]

20. Cabral R, Klumb E, Couto M, Carneiro S. Evaluation of toxic retinopathy caused by antimalarial medications with spectral domain optical coherence tomography. Arq Bras Oftalmol 2019;82(1):12-17 [FREE Full text] [doi: 10.5935/0004-2749.20190002] [Medline: 30403262]
21. Espandar G, Moghimi J, Ghorbani R, Pourazizi M, Seiri MA, Khosravi S. Retinal Toxicity in Patients Treated With Hydroxychloroquine: A Cross-Sectional Study. Med Hypothesis Discov Innov Ophthalmol 2016;5(2):41-46 [FREE Full text] [Medline: 28293646]
22. Jover JA, Leon L, Pato E, Loza E, Rosales Z, Matias MA, et al. Long-term use of antimalarial drugs in rheumatic diseases. Clin Exp Rheumatol 2012;30(3):380-387. [Medline: 22339928]
23. Maturi RK, Yu M, Weleber RG. Multifocal electroretinographic evaluation of long-term hydroxychloroquine users. Arch Ophthalmol 2004 Jul;122(7):973-981. [doi: 10.1001/archopht.122.7.973] [Medline: 15249360]
24. Tzekov RT, Serrato A, Marmor MF. ERG findings in patients using hydroxychloroquine. Doc Ophthalmol 2004 Jan;108(1):87-97. [doi: 10.1023/b:doop.0000018410.08802.d3]
25. Neubauer AS, Samari-Kermani K, Schaller U, Welge-Lübetaen U, Rudolph G, Berninger T. Detecting chloroquine retinopathy: electro-oculogram versus colour vision. Br J Ophthalmol 2003 Jul;87(7):902-908 [FREE Full text] [doi: 10.1136/bjo.87.7.902] [Medline: 12812896]

Abbreviations

AI: artificial intelligence
API: application programming interface
bi-LSTM: bidirectional long-short term memory
COVID-19: coronavirus disease
RMA: rapid meta-analysis

Edited by G Eysenbach; submitted 08.05.20; peer-reviewed by D Smith, A Doganer, KM Kuo; comments to author 02.06.20; revised version received 15.06.20; accepted 22.07.20; published 17.08.20.

Please cite as:
Michelson M, Chow T, Martin NA, Ross M, Tee Qiao Ying A, Minton S
Artificial Intelligence for Rapid Meta-Analysis: Case Study on Ocular Toxicity of Hydroxychloroquine
J Med Internet Res 2020;22(8):e20007
URL: http://www.jmir.org/2020/8/e20007/
doi:10.2196/20007
PMID:32804086

Original Paper

Evaluating Smart Assistant Responses for Accuracy and Misinformation Regarding Human Papillomavirus Vaccination: Content Analysis Study

John Ferrand[1*], MPH; Ryli Hockensmith[1*]; Rebecca Fagen Houghton[1*], MPA; Eric R Walsh-Buhi[1*], MPH, PhD

School of Public Health-Bloomington, Indiana University, Bloomington, IN, United States

[*]all authors contributed equally

Corresponding Author:
Eric R Walsh-Buhi, MPH, PhD
School of Public Health-Bloomington
Indiana University
1025 E. 7th Street, Room 116
Department of Applied Health Science
Bloomington, IN, 47405
United States
Phone: 1 8128554867
Email: erwals@iu.edu

Abstract

Background: Almost half (46%) of Americans have used a smart assistant of some kind (eg, Apple Siri), and 25% have used a stand-alone smart assistant (eg, Amazon Echo). This positions smart assistants as potentially useful modalities for retrieving health-related information; however, the accuracy of smart assistant responses lacks rigorous evaluation.

Objective: This study aimed to evaluate the levels of accuracy, misinformation, and sentiment in smart assistant responses to human papillomavirus (HPV) vaccination–related questions.

Methods: We systematically examined responses to questions about the HPV vaccine from the following four most popular smart assistants: Apple Siri, Google Assistant, Amazon Alexa, and Microsoft Cortana. One team member posed 10 questions to each smart assistant and recorded all queries and responses. Two raters independently coded all responses (κ=0.85). We then assessed differences among the smart assistants in terms of response accuracy, presence of misinformation, and sentiment regarding the HPV vaccine.

Results: A total of 103 responses were obtained from the 10 questions posed across the smart assistants. Google Assistant data were excluded owing to nonresponse. Over half (n=63, 61%) of the responses of the remaining three smart assistants were accurate. We found statistically significant differences across the smart assistants (N=103, χ^2_2=7.807, *P*=.02), with Cortana yielding the greatest proportion of misinformation. Siri yielded the greatest proportion of accurate responses (n=26, 72%), whereas Cortana yielded the lowest proportion of accurate responses (n=33, 54%). Most response sentiments across smart assistants were positive (n=65, 64%) or neutral (n=18, 18%), but Cortana's responses yielded the largest proportion of negative sentiment (n=7, 12%).

Conclusions: Smart assistants appear to be average-quality sources for HPV vaccination information, with Alexa responding most reliably. Cortana returned the largest proportion of inaccurate responses, the most misinformation, and the greatest proportion of results with negative sentiments. More collaboration between technology companies and public health entities is necessary to improve the retrieval of accurate health information via smart assistants.

(J Med Internet Res 2020;22(8):e19018) doi:10.2196/19018

KEYWORDS

digital health; human papillomavirus; smart assistants; chatbots; conversational agents; misinformation; infodemiology; vaccination

Introduction

Background

Voice assistants, a form of chatbot or conversational agent often referred to colloquially as "smart assistants," are devices that respond to human voices and can be commanded to do a variety of tasks [1]. Smart assistants have existed in their most contemporary form since 2010, with the introduction of Siri, and they are used for numerous tasks such as automation (ie, integration with climate control devices and entertainment devices), retrieval of information about certain topics, and shopping [1]. Smart assistants have been integrated into smartphones, laptops, speakers, and other devices, creating a large network of consumers who frequently utilize smart assistants to acquire information on a range of topics [1].

Prevalence of Smart Assistant Use

According to the Pew Research Center, out of a total sample of 2045 Americans, nearly half (46%) reported using smart assistants in 2017 [2]. Of 4272 Americans surveyed in 2019, one-quarter (25%) had a stand-alone smart assistant device (eg, Amazon Echo, Google Home) in their homes [3]. Households that reported having a stand-alone smart assistant device (n=1067) also reported higher income (34% earned US $75,000 or more per year; 15% earned below US $30,000) [3]. The same report indicated that younger Americans frequently use stand-alone smart assistants, with 32% of users aged between 18 and 29 years and 28% of users aged between 30 and 49 years [3]. This difference may be due to varied ways of assessing use (eg, ever use and daily use). In contrast, smartphone smart assistants are more frequently adopted by those aged between 18 and 29 years (81%), whereas those aged between 45 and 60 years report the most daily active use of smartphone smart assistants [2]. These usage trends depict nuanced usage patterns wherein younger users are the most common adopters of smartphone smart assistants and older users, once they begin using smartphone smart assistants, use them more frequently than do other age groups. These varying adoption and usage behaviors provide multiple avenues for targeting different age groups through smart assistants.

Uses of Smart Assistants

More than half of Americans report that a major reason for using smart assistants is the ability to interact with their devices without using their hands [2]. One-quarter of Americans say that they use smart assistants to remotely control other devices such as heating systems, door locks, and lights [2]. Given that approximately 35% of Americans report searching online (ie, using Google or another search engine) to self-diagnose a medical condition, it seems logical to assume that individuals may turn to smart assistants for this information as well [4]. Unfortunately, smart assistants are relatively new technologies, and their responses have not been rigorously evaluated in many contexts.

Previous Assessments of Smart Assistants

In general, smart assistants have been evaluated for their response accuracy [5] and their general usability [6]. While considerable research has been conducted assessing digital health approaches, such as text message–based approaches and mobile apps for a range of health behaviors [7-9], few have included smart assistant responses to health-related queries. For instance, a pilot study comparing two smart assistants (Google Assistant and Apple Siri) to a standard Google Search on the topic of smoking cessation resources found that Google Assistant provided a greater number of evidence-based responses [10]. Other studies have specifically examined consumer experiences with the natural language processing of smart assistants [5].

Human Papillomavirus Prevention

To effectively assess smart assistant responses for accuracy and misinformation, human papillomavirus (HPV) has been identified as a controversial content area with a substantial evidence base. The evidence base of HPV is scientifically valid but hotly contested. HPV is the most common sexually transmitted infection and is a known cause of cervical cancer, as well as several other types of cancers [11]. While most sexually active people will contract HPV at some point in their lives, how the infection resolves varies from person to person, and a federally approved vaccine has been shown to be effective at reducing both the incidence of HPV transmission and the incidences of genital and anogenital HPV infection and cervical lesions [12]. Despite studies affirming the positive effects of the HPV vaccine while debunking inaccurate claims, there continues to be large-scale misinformation efforts (mostly conducted online through social media platforms) surrounding this issue, which are driven in part by the antivaccination movement [13-20]. These issues of misinformation have contributed to mistrust of medical professionals [21-23] and misunderstanding of diseases and their risks [24,25]. In January 2020, the World Health Organization (WHO) released a list of urgent global health challenges for the new decade, including stopping vaccine-preventable diseases and earning the public's trust [26]. These two challenges are closely intertwined, as trust helps to shape whether individuals rely on provider recommendations (eg, whether parents will vaccinate their children) [22,27] and how misinformation disseminated online and across social media platforms influences vaccine refusal or vaccine delay [28]. In fact, the WHO has stated that vaccine hesitancy is one of the top 10 threats to global health [29].

In this study, we attempted to answer the following research question: do responses to smart assistant queries vary between different smart assistants, with regard to accuracy, misinformation, and sentiment toward HPV vaccination?

Methods

Query Development

In order to effectively assess smart assistants' responses for accuracy and misinformation, we utilized questions from the chat-text service of Planned Parenthood (personal communication by Nicole Levitz; March 18, 2019). We chose to focus on questions around the HPV vaccine owing to the previously identified issues of accuracy and misinformation surrounding this topic [30,31]. We chose variations of 10 evidence-based questions from the Planned Parenthood system, allowing us to better evaluate responses to those questions for

accuracy and misinformation. The questions are listed in Textbox 1.

Textbox 1. Queries posed to smart assistants.

1. Does the HPV vaccine work?
2. Does the HPV vaccine cause autism?
3. Does Gardasil work?
4. Does Gardasil cause autism?
5. Is the HPV vaccine dangerous?
6. Is Gardasil dangerous?
7. Who can get the HPV vaccine?
8. Where can I get the HPV vaccine?
9. Does Gardasil kill?
10. How much does the HPV vaccine cost?

Search Process

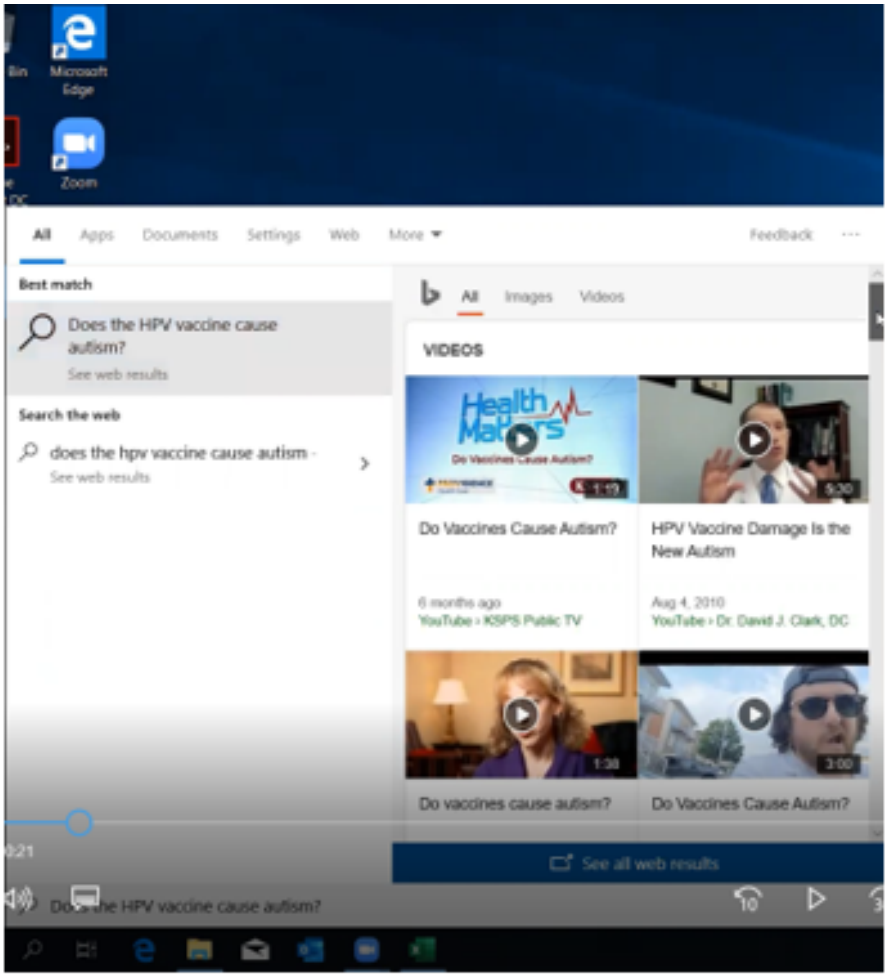
One member of the research team queried each of the four most popular smart assistants (Apple Siri, Google Assistant, Amazon Alexa, and Microsoft Cortana), which were identified in the 2019 Voice Report by Microsoft [32]. In the event that the smart assistant provided a nonresponse to a query (eg, "I don't know how to answer that"), the team member queried the smart assistant a maximum of three times to ensure that it was not an errored response due to misunderstanding the query, background noise, or some other unrelated reason.

Smart assistants provided varying numbers of results on their respective first pages for each query. Alexa and Google Assistant provided only one oral result for each query, whereas Siri provided five text results for each query. Cortana provided between three and nine text or video results for each query. Multimedia Appendix 1 displays the number of results each smart assistant provided per query. The team member, responsible for conducting the searches, recorded the first page of the results for data extraction based on previous studies indicating that users more frequently click on the top 10 results, which tend to be concentrated on the first page of most search results [33]. There is also evidence suggesting that people do not necessarily only click the first result on a page. In a 2020 study of search engine optimization, 16% of respondents in the study reported clicking on only the first result compared with 17% and 14% of respondents who reported clicking on three and five results, respectively [34]. This finding suggests that the remaining results on the first page should be extracted to best replicate actual human search behavior.

Recording Process

The team member, who posed questions to the smart assistants, used video and audio recording software to record both the queries and the smart assistant responses. We used these recordings in the data extraction and coding process. Figure 1 depicts a recording of a query posed to Cortana.

Figure 1. Example of a video result from a query to Cortana.



Data Extraction and Coding Process

We developed a web-based (Qualtrics) survey to aid in data extraction and coding of the smart assistant responses. Survey items were designed to extract relevant data from the smart assistant responses, including the types of responses (eg, video, web page, blog, and journal article), sources of the responses (eg, business, doctors, and health information provider), sentiments of the responses toward the HPV vaccine (eg, positive, neutral, and negative), accuracy of the responses, any incidence of misinformation in the responses, and topics discussed in the responses (eg, cancer, sexual behavior, and conspiracy theory).

We defined *accuracy* as a response that reflected the existing evidence. We coded accuracy using a dichotomy approach (0 for not accurate; 1 for accurate), which we applied specifically to the query (ie, accuracy of unrelated tangential content was not considered). If the response did not answer the query with the correct reply or positioned the correct reply as dubious or incorrect, we considered it "not accurate." We defined *misinformation* as either deliberate or accidental promotion of previously disproved or unproven beliefs, attitudes, and behaviors. We coded misinformation using a dichotomy approach (0 if it did not provide misinformation; 1 if it did provide misinformation), which we applied to the entirety of the response (not just the answer to the question posed). If any misinformation was found in the smart assistant's response, we coded it as having provided misinformation. Table 1 depicts examples of accuracy and misinformation in smart assistant responses to several queries. Since accuracy and misinformation were coded based on different aspects of the smart assistant responses, there were some cases with both an accurate response to the query and some misinformation in the result. We coded sentiment toward vaccines as one of the following four potential categories: negative (mostly negative statements), neutral (neither positive nor negative statements), positive (mostly positive statements), and ambiguous (both negative and positive statements). We applied sentiment coding to the entirety of the response.

Two independent team members extracted and coded the data with an almost perfect agreement (κ=0.85) [35]. We resolved discrepancies through discussion.

Table 1. Examples of accuracy and misinformation in smart assistant responses.

Query	Accuracy	Misinformation
Does the HPV[a] vaccine work?	"In the trials that led to the approval of Gardasil and Cervarix, these vaccines were found to provide nearly 100% protection against persistent cervical infections with HPV types 16 and 18 and the cervical cell changes that these persistent infections can cause."	N/A[b]
Does the HPV vaccine cause autism?	"There is no link between vaccines and autism."	"The scientific literature is now starting to fill up with case reports and studies and articles that irrefutably show that there is a connection between this vaccine (and it's an ugly vaccine) and neurological damage."
Does Gardasil work?	"Gardasil works by preventing the infection of the types of HPV that can lead to cervical cancer..."	"The Gardasil HPV vaccine has been proved to have caused the deaths of 32 women."
Does Gardasil cause autism?	"There has never been a study that has shown that vaccines cause autism."	N/A
Is the HPV vaccine dangerous?	"Findings from many vaccine safety monitoring systems and more than 160 studies have shown that HPV vaccines have a favorable safety profile—the body of scientific evidence overwhelmingly supports their safety."	"Aluminum in the vaccines is toxic enough to be harmful."
Is Gardasil dangerous?	"Although this information is accurate in a strictly literal sense, it is a misleading presentation of raw data that does not in itself establish a causal connection between Gardasil and the posited medical dangers."	"The Gardasil HPV vaccine has been proved to have caused the deaths of 32 women."
Who can get the HPV vaccine?	"All people ages 9 to 45 can get the HPV vaccine to protect against genital warts and/or different types of HPV that can cause cancer."	N/A
Where can I get the HPV vaccine?	"The HPV vaccine is available at: Healthcare Clinic for patients aged 11-26. Walgreens Pharmacy. Ages vary by state."	N/A
Does Gardasil kill?	"I cannot stress this enough, based on this report alone you can't make a determination that the vaccine caused the deaths."	N/A
How much does the HPV vaccine cost?	"Each dose of the vaccine can cost about $250."	N/A

[a]HPV: human papillomavirus.

[b]N/A: not applicable.

Analyses

The sample consisted of 128 total data points across all four smart assistants. We excluded any nonresponse data points (eg, "I don't know how to answer that") and responses that were categorized as *other* (eg, responded with completely irrelevant content to the query) from the final analyses. Google Assistant provided nonresponses to every query and was removed, thus reducing our final analysis sample to 103 responses across three smart assistants. These nonresponses and other responses were removed because they did not in any way address the query posed. Multimedia Appendix 1 displays the frequencies of responses and nonresponses for each smart assistant. Descriptive frequencies and chi-square difference tests were used to examine the levels of accuracy and misinformation among the smart assistants.

Results

Table 2 displays the source and content characteristics of the smart assistant responses. Smart assistant responses contained content published by an organization in 82 out of 103 (79.6%) responses, whereas 20 out of 103 (19.4%) responses were published by an individual. Content was provided by organizations including some type of health information provider (eg, Planned Parenthood and Mayo Clinic) in 36 out of 103 (35.0%) responses and by a government entity (eg, Centers for Disease Control and Prevention) in 25 out of 103 (24.3%) responses. Cortana provided content published by individuals (eg, physicians and journalists) in 18 out of 61 (29.5%) responses, whereas Siri and Alexa provided content published by individuals in only 2 out of 36 (5.6%) and 0 out of 6 (0.0%) responses, respectively. The most common type of individual who provided content (10 out of 103 responses, 9.7%) was some classification of physician. The type of content provided in the smart assistant responses varied, with videos provided in 30 out of 103 (29.1%) responses, driven entirely by Cortana, which was the only smart assistant to provide video responses. Siri's responses were classified as frequently asked question (FAQ) pages in 13 out of 36 (36.1%) responses, whereas Cortana's responses were classified as videos in 30 out of 61 (49.2%) responses and as general web pages in 12 out of 61 (19.7%) responses.

Table 2. Source and content characteristics of smart assistant responses.

Variable	Value (N=103), n (%)
Source[a]	
Health information provider	36 (35.0)
Federal/city/state	25 (24.3)
Nonprofit/advocacy group	15 (14.6)
Physician	10 (9.7)
Other	9 (8.7)
News/media organization	8 (7.8)
Business	6 (5.8)
Journalist/press	2 (1.9)
Politician	2 (1.9)
Health care organization	1 (1.0)
Content	
Format of the smart assistant responses	
Video	30 (29.1)
General web page	24 (23.3)
Article	14 (13.6)
FAQ[b] site	19 (18.4)
Other	6 (5.8)
Blog post	3 (2.9)
Location/map/directions	1 (1.0)
Topics mentioned in the smart assistant responses[a]	
Cancer	77 (74.8)
Vaccines and side effects	57 (55.3)
Sexual behaviors	26 (25.2)
Access to medical care	20 (19.4)
Direct impacts on loved ones	13 (12.6)
Conspiracy theories	6 (5.8)
Vaccines are ineffective	6 (5.8)
Mass media	6 (5.8)
Risk of disease is lower than risk of adverse vaccination events	2 (1.9)
Sentiment expressed toward vaccines in smart assistant responses	
Positive	65 (63.1)
Neutral	18 (17.5)
Ambiguous	11 (10.7)
Negative	7 (6.8)

[a]Totals may exceed 100% owing to the availability of multiple response options.

[b]FAQ: frequently asked question.

Tables 3 and 4 depict the differences among smart assistants in terms of primary outcomes. Smart assistant responses contained accurate answers in 63 out of 103 (61.2%) responses. Neither response accuracy (P=.10) nor response sentiment (P=.22) was significantly different among the devices. The number of responses provided for each query varied across the devices, but 4 out of 6 (66.7%) responses by Alexa and 26 out of 36 (72.2%) responses by Siri were accurate. In contrast, 33 out of 61 (54.1%) responses by Cortana were accurate. There were no cases in which responses contained both accurate answers and

misinformation. Chi-square tests indicated that the three smart assistants significantly varied in terms of whether they provided misinformation in their responses (N=103, χ^2_2=7.807, *P*=.02) and whether they provided any evidence to support the claims made in their responses (N=103, χ^2_2=11.054, *P*=.004). Smart assistant responses contained at least one instance of misinformation in 18 out of 103 (17.5%) responses. Responses by Alexa contained no misinformation, whereas misinformation was present in 2 out of 36 (5.6%) responses by Siri and 16 out of 61 (26.2%) responses by Cortana. Alexa provided evidence to support each of its responses, whereas Siri provided evidence in 21 out of 36 (58.3%) responses and Cortana provided evidence in 23 out of 61 (37.7%) responses.

In general, smart assistant responses contained positive sentiment toward vaccines, with 65 out of 103 (63.1%) responses containing primarily positive sentiment. The second most common sentiment was "neutral," which was found in 18 out of 103 (17.5%) responses. The least common sentiment was "negative," which was found in only 7 out of 103 (6.8%) responses, and all were provided by Cortana.

Table 3. Response differences among smart assistants in terms of accuracy, evidence provided, and misinformation.

Response quality of the smart assistants	Value, n/N (%)	Chi-square effect size (df)	*P* value
Accurate response? (Yes)	63/103 (61.2%)		
Alexa	4/6 (66.7%)	4.558 (2)	.10
Siri	26/36 (72.2%)	4.558 (2)	.10
Cortana	33/61 (54.1%)	4.558 (2)	.10
Evidence provided? (Yes)	50/103 (48.5%)		
Alexa	6/6 (100%)	11.054 (2)	.004
Siri	21/36 (58.3%)	11.054 (2)	.004
Cortana	23/61 (37.7%)	11.054 (2)	.004
Misinformation provided? (Yes)	18/103 (17.5%)		
Alexa	0/6 (0%)	7.807 (2)	.02
Siri	2/36 (5.6%)	7.807 (2)	.02
Cortana	16/61 (26.2%)	7.807 (2)	.02

Table 4. Sentiment differences among the smart assistants.

Smart assistant	Content sentiment					*P* value
	Neutral (N=18)	Negative (N=7)	Ambiguous (N=11)	Positive (N=65)	Chi-square effect size (df)	
Alexa (N=6), n (%)	1 (16.7%)	0 (0%)	0 (0%)	5 (83.3%)	8.20 (6)	.22
Siri (N=35), n (%)	9 (25.7%)	0 (0%)	5 (14.3%)	21 (60%)	8.20 (6)	.22
Cortana (N=60), n (%)	8 (13.3%)	7 (11.7%)	6 (10%)	39 (65%)	8.20 (6)	.22

Discussion

Principal Results

This study sought to determine whether responses to queries vary among smart assistants in terms of accuracy, misinformation, and sentiment related to HPV vaccination. Our results indicate that smart assistants responded differently when asked about this topic. Specifically, smart assistants showed variations in the level of misinformation provided in responses, as well as the provision of evidence to support claims made in their responses. Smart assistants did not differ statistically in terms of the accuracy of their responses to queries involving HPV vaccination.

To our knowledge, this study is the first to report on smart assistants and their responses to HPV vaccine–related queries. Previous studies of HPV vaccination behaviors did not focus on smart assistants, and they indicated that antivaccination messages may influence one's decisions to vaccinate themselves or their children [15]. However, given the growing number of persons and households adopting and utilizing smart assistants and the fact that smart assistants may be able to provide HPV vaccine information, it is necessary to examine whether responses delivered through smart assistants are disseminating these same antivaccination messages. This study is the first step in establishing a foundation for the types of vaccine sentiments that are present in content disseminated by smart assistants.

Importantly, our study revealed that, in general, smart assistants largely provide accurate and positive information regarding HPV vaccination (just under two-thirds of all smart assistant responses were accurate and positive), with no relevant differences across devices. We should reiterate here, however, that Google Assistant provided nonresponses to every query and, accordingly, was removed from the study. While accurate information on something as beneficial as HPV vaccination is necessary to mitigate HPV transmission and cancer incidence, just over one-third of responses contained *inaccurate* answers,

suggesting that work is needed in this area to improve the provision of health information via smart assistants. On the other hand, misinformation was more of a concern across devices, as Cortana yielded the greatest number of responses containing misinformation (one-quarter of Cortana's responses contained misinformation).

Comparison With Prior Work

Previous studies of smart assistant responses have focused on whether the devices responded correctly or whether they understood the query [5,6]. Only a handful of studies examined smart assistant responses in the context of a health behavior [10]. This study further assessed smart assistant responses to determine whether they provide accurate evidence to specific health behavior questions. Evidence is critical in combatting misinformation, at least on social media platforms. For example, in an experimental study that exposed Facebook users to simulated misinformation and different correction mechanisms about the Zika virus, the authors found that correction can work when it provides supporting evidence or appropriate sources to accompany refutation (eg, Centers for Disease Control and Prevention) [36]. The authors reported that this finding was maintained even in those who were rated higher in conspiracy beliefs.

Limitations

This study has limitations that should be considered when evaluating its results. The devices used their specific locations when responding to queries, which may have produced less generalizable responses. Some of the smart assistants employed in the study were previously used devices and may have been influenced by prior searches of the former user. To the best of our knowledge, no private mode exists in these smart assistants, which would prevent the queries and responses from influencing future searches. The sample size was limited for several reasons, including nonresponses and other unrelated responses to queries. Specifically, Google Assistant provided no response data, which may have influenced the results. Any responses that were entirely unrelated to the queries should be explored in future studies, since this may be an indication of a larger issue with smart assistants not comprehending the queries. Based on our knowledge about natural language processing in these smart assistants, it is suggested that syntax and dictation are some of the many influencers of smart assistant responses [37], and the queries posed to each device may not have been appropriately worded to elicit the most effective responses. For example, our use of a single investigator to ask the questions could be viewed as a limitation (ie, we do not know how the smart assistants would have responded to an investigator with a different gender, tone, or accent). The investigator only queried a single example of a smart assistant (ie, only one device was used to query Alexa), which could be a limitation, as smart assistants may respond differently depending on whether a stand-alone smart assistant or a smartphone-based smart assistant is used. The underlying search engine behind each smart assistant determines the responses to queries, which could have limited the information provided by smart assistants that are not connected to large search engines such as Google. The varying number of results provided in the smart assistant responses to queries could have impacted our ability to compare across the devices; however, since we found no relevant differences between the devices with regard to accuracy or sentiment, we do not believe that more results from one device impacted our findings or conclusions.

Despite these limitations, there are several strengths that provide further evidence of the study's validity and importance. First, we utilized questions submitted to a Planned Parenthood chat service, which represent real-world issues on which consumers have previously sought additional information. Second, the questions used as queries were chosen because responses to these queries could be scored as either correct or incorrect, limiting "gray areas" in assessing accuracy of the smart assistant responses. Third, the queries were conducted by a single member of the research team to maintain consistency in language processing factors (eg, tone and syntax), and the search and coding processes were systematic to maintain consistency as well. Overall, this research furthers our understanding of how an emerging technology disseminates health information and whether such information can be classified as accurate and evidence-based. This study provides a basic framework for future evaluation of these smart assistants on which more advanced devices may be based.

Our findings paint a picture of smart assistants as newly emerged potential health promotion tools that are yet to be rigorously evaluated in this area. Smart assistants have only recently been introduced as potential health promotion tools (most notably from an executive perspective [38]) despite the growing proportion of households and individuals that use them [2,3]. The untapped potential of these devices for evidence-based information dissemination to consumers should be further explored. The potential of the devices may also be determined by their manufacturers who have, in some cases, provided platforms on which specialized topical information can be consolidated and further explored. For example, Alexa has an option for developers to create an Alexa Skill, which is essentially an application for Alexa that provides predefined information when queried specifically for that information. For example, an Alexa Skill that specifically searches in predefined evidence-based sources when queried for information on HPV vaccination could be developed. The downside to this potential approach for improving public health is that, as our results show, there is risk of disseminating misinformation through smart assistant responses, potentially reducing the positive impacts of a health promotion intervention. More needs to be done to better understand the susceptibility of these devices and their respective skills to outside influences.

Conclusions

Our results suggest that not all smart assistants are created equal with regard to utility, at least when it comes to provided evidence and misinformation in their responses. These findings should spark further research into how the proprietors of smart assistants procure the information that is disseminated to consumers of their products. Specifically, we suggest that manufacturers of these and other smart assistants collaborate with researchers to further evaluate the accuracy of smart assistants as public health tools and determine together how to

disseminate information and what fact-checking assessments should be used for such information. With the high rate of HPV transmission globally and in the United States, smart assistants and their manufacturers are well positioned to deliver evidence-based health information to consumers. However, such a practice necessitates strong communication between technology companies, who may not be as focused on the accuracy or source of their most heavily promoted content, and public health entities. The focus on collaboration to address the issues of information accuracy and misinformation is paramount if we are to adequately respond to the WHO's list of urgent global health challenges for the new decade, namely stopping vaccine-preventable diseases.

Authors' Contributions

All authors contributed equally to the development, review, and submission of this manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Frequencies of responses and nonresponses by each smart assistant.
[PNG File , 55 KB - jmir_v22i8e19018_app1.PNG]

References

1. Hoy MB. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. Med Ref Serv Q 2018;37(1):81-88. [doi: 10.1080/02763869.2018.1404391] [Medline: 29327988]
2. Nearly half of Americans use digital voice assistants, mostly on their smartphones. Pew Research Center. 2017 Dec 12. URL: https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/ [accessed 2020-07-23]
3. Auxier B. 5 things to know about Americans and their smart speakers. Pew Research Center. 2019 Nov 21. URL: https://www.pewresearch.org/fact-tank/2019/11/21/5-things-to-know-about-americans-and-their-smart-speakers/ [accessed 2020-07-23]
4. Fox S, Duggan M. Health Online 2013. Pew Research Center. 2013 Jan 15. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2020-01-23]
5. Berdasco A, López G, Diaz I, Quesada L, Guerrero LA. User Experience Comparison of Intelligent Personal Assistants: Alexa, Google Assistant, Siri and Cortana. Proceedings 2019 Nov 20;31(1):51. [doi: 10.3390/proceedings2019031051]
6. López G, Quesada L, Guerrero L. Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces. 2018 Presented at: 8th International Conference on Applied Human Factors and Ergonomics (AFHE); 2017; Los Angeles, CA p. 241-250. [doi: 10.1007/978-3-319-60366-7_23]
7. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
8. Triantafyllidis AK, Tsanas A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. J Med Internet Res 2019 Apr 05;21(4):e12286 [FREE Full text] [doi: 10.2196/12286] [Medline: 30950797]
9. Zanaboni P, Ngangue P, Mbemba GI, Schopf TR, Bergmo TS, Gagnon M. Methods to Evaluate the Effects of Internet-Based Digital Health Interventions for Citizens: Systematic Review of Reviews. J Med Internet Res 2018 Jun 07;20(6):e10202 [FREE Full text] [doi: 10.2196/10202] [Medline: 29880470]
10. Boyd M, Wilson N. Just ask Siri? A pilot study comparing smartphone digital assistants and laptop Google searches for smoking cessation advice. PLoS One 2018;13(3):e0194811 [FREE Full text] [doi: 10.1371/journal.pone.0194811] [Medline: 29590168]
11. Senkomago V, Henley SJ, Thomas CC, Mix JM, Markowitz LE, Saraiya M. Human Papillomavirus-Attributable Cancers - United States, 2012-2016. MMWR Morb Mortal Wkly Rep 2019 Aug 23;68(33):724-728 [FREE Full text] [doi: 10.15585/mmwr.mm6833a3] [Medline: 31437140]
12. Drolet M, Bénard É, Pérez N, Brisson M, HPV Vaccination Impact Study Group. Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and meta-analysis. Lancet 2019 Aug 10;394(10197):497-509 [FREE Full text] [doi: 10.1016/S0140-6736(19)30298-3] [Medline: 31255301]
13. Bradshaw AS, Treise D, Shelton SS, Cretul M, Raisa A, Bajalia A, et al. Propagandizing anti-vaccination: Analysis of Vaccines Revealed documentary series. Vaccine 2020 Feb 18;38(8):2058-2069. [doi: 10.1016/j.vaccine.2019.12.027] [Medline: 31980194]
14. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, et al. The spreading of misinformation online. Proc Natl Acad Sci U S A 2016 Jan 19;113(3):554-559 [FREE Full text] [doi: 10.1073/pnas.1517441113] [Medline: 26729863]

15. Dubé E, Vivion M, MacDonald NE. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. Expert Rev Vaccines 2015 Jan;14(1):99-117. [doi: 10.1586/14760584.2015.964212] [Medline: 25373435]
16. Elkin LE, Pullon SR, Stubbe MH. 'Should I vaccinate my child?' comparing the displayed stances of vaccine information retrieved from Google, Facebook and YouTube. Vaccine 2020 Mar 17;38(13):2771-2778. [doi: 10.1016/j.vaccine.2020.02.041] [Medline: 32107061]
17. Hoffman BL, Felter EM, Chu K, Shensa A, Hermann C, Wolynn T, et al. It's not all about autism: The emerging landscape of anti-vaccination sentiment on Facebook. Vaccine 2019 Apr 10;37(16):2216-2223. [doi: 10.1016/j.vaccine.2019.03.003] [Medline: 30905530]
18. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm--an overview of tactics and tropes used online by the anti-vaccination movement. Vaccine 2012 May 28;30(25):3778-3789. [doi: 10.1016/j.vaccine.2011.11.112] [Medline: 22172504]
19. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. Vaccine 2010 Feb 17;28(7):1709-1716. [doi: 10.1016/j.vaccine.2009.12.022] [Medline: 20045099]
20. Southwell BG, Thorson EA, Sheble L. The Persistence and Peril of Misinformation. Am. Sci 2017;105(6):372. [doi: 10.1511/2017.105.6.372]
21. Cooper DL, Hernandez ND, Rollins L, Akintobi TH, McAllister C. HPV vaccine awareness and the association of trust in cancer information from physicians among males. Vaccine 2017 May 09;35(20):2661-2667 [FREE Full text] [doi: 10.1016/j.vaccine.2017.03.083] [Medline: 28396210]
22. Fu LY, Zimet GD, Latkin CA, Joseph JG. Associations of trust and healthcare provider advice with HPV vaccine acceptance among African American parents. Vaccine 2017 Feb 01;35(5):802-807 [FREE Full text] [doi: 10.1016/j.vaccine.2016.12.045] [Medline: 28063706]
23. Gidengil C, Chen C, Parker AM, Nowak S, Matthews L. Beliefs around childhood vaccines in the United States: A systematic review. Vaccine 2019 Oct 23;37(45):6793-6802. [doi: 10.1016/j.vaccine.2019.08.068] [Medline: 31562000]
24. LaCour M, Davis T. Vaccine skepticism reflects basic cognitive differences in mortality-related event frequency estimation. Vaccine 2020 May 06;38(21):3790-3799. [doi: 10.1016/j.vaccine.2020.02.052] [Medline: 32169393]
25. Thompson EL, Wheldon CW, Rosen BL, Maness SB, Kasting ML, Massey PM. Awareness and knowledge of HPV and HPV vaccination among adults ages 27-45 years. Vaccine 2020 Mar 30;38(15):3143-3148. [doi: 10.1016/j.vaccine.2020.01.053] [Medline: 32029321]
26. Urgent health challenges for the next decade. World Health Organization. 2020. URL: https://www.who.int/news-room/photo-story/photo-story-detail/urgent-health-challenges-for-the-next-decade [accessed 2020-01-23]
27. Paterson P, Meurice F, Stanberry LR, Glismann S, Rosenthal SL, Larson HJ. Vaccine hesitancy and healthcare providers. Vaccine 2016 Dec 20;34(52):6700-6706 [FREE Full text] [doi: 10.1016/j.vaccine.2016.10.042] [Medline: 27810314]
28. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
29. Ten health issues WHO will tackle this year. World Health Organization. 2019. URL: https://www.who.int/news-room/feature-stories/ten-threats-to-global-health-in-2019 [accessed 2020-07-23]
30. Dibble KE, Maksut JL, Siembida EJ, Hutchison M, Bellizzi KM. A Systematic Literature Review of HPV Vaccination Barriers Among Adolescent and Young Adult Males. J Adolesc Young Adult Oncol 2019 Oct;8(5):495-511. [doi: 10.1089/jayao.2019.0004] [Medline: 31090474]
31. Warner EL, Ding Q, Pappas L, Bodson J, Fowler B, Mooney R, et al. Health Care Providers' Knowledge of HPV Vaccination, Barriers, and Strategies in a State With Low HPV Vaccine Receipt: Mixed-Methods Study. JMIR Cancer 2017 Aug 11;3(2):e12 [FREE Full text] [doi: 10.2196/cancer.7345] [Medline: 28801303]
32. Olson C, Kemery K. 2019 Voice report: Consumer adoption of voice technology and digital assistants. Microsoft. 2019. URL: https://about.ads.microsoft.com/en-us/insights/2019-voice-report [accessed 2020-07-23]
33. Petrescu P, Ghita M, Loiz D. Google Organic CTR Study. Advanced Web Ranking. 2014. URL: https://www.advancedwebranking.com/ebooks/Google-Organic-CTR-Study-2014.pdf [accessed 2020-07-23]
34. Lincoln J. SEO and Intent 2020: New Study Reveals True State Of Search. Ignite Visibility. 2020. URL: https://ignitevisibility.com/seo-and-intent-2020-new-study-reveals-true-state-of-search/ [accessed 2020-07-23]
35. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]
36. Bode L, Vraga EK. See Something, Say Something: Correction of Global Health Misinformation on Social Media. Health Commun 2018 Sep;33(9):1131-1140. [doi: 10.1080/10410236.2017.1331312] [Medline: 28622038]
37. Goksel Canbek N, Mutlu ME. On the track of Artificial Intelligence: Learning with Intelligent Personal Assistants. HumanSciences 2016 Jan 28;13(1):592. [doi: 10.14687/ijhs.v13i1.3549]
38. Cimino J, Stefanacci R, Pakizegee M. Alexa, Can You Transform Healthcare? Managed Healthcare Executive. 2020 Feb 26. URL: https://www.managedhealthcareexecutive.com/article/alexa-can-you-transform-healthcare-0 [accessed 2020-07-23]

Abbreviations

HPV: human papillomavirus
WHO: World Health Organization

Edited by T Kool, G Eysenbach; submitted 01.04.20; peer-reviewed by M Boyd, A Benis, Y Rivera, R Dodd, A Benis; comments to author 22.04.20; revised version received 18.05.20; accepted 02.07.20; published 03.08.20.

Please cite as:
Ferrand J, Hockensmith R, Houghton RF, Walsh-Buhi ER
Evaluating Smart Assistant Responses for Accuracy and Misinformation Regarding Human Papillomavirus Vaccination: Content Analysis Study
J Med Internet Res 2020;22(8):e19018
URL: https://www.jmir.org/2020/8/e19018
doi:10.2196/19018
PMID:32744508

Original Paper

Understanding Time Series Patterns of Weight and Meal History Reports in Mobile Weight Loss Intervention Programs: Data-Driven Analysis

Junetae Kim[1,2,3*], PhD; Hye Jin Kam[4*], PhD; Youngin Kim[5,6], MD; Yura Lee[7], MD, PhD; Jae-Ho Lee[7,8], MD, PhD

[1]Graduate School of Cancer Science and Policy, National Cancer Center, Goyang-si, Republic of Korea

[2]Cancer Data Center, National Cancer Control Institute, National Cancer Center, Goyang-si, Republic of Korea

[3]Healthcare AI Team, Healthcare Platform Center, National Cancer Center, Goyang-si, Republic of Korea

[4]AIMMED Co, Ltd, Seoul, Republic of Korea

[5]Noom Inc., New York, NY, United States

[6]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[7]Department of Information Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[8]Department of Emergency Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jae-Ho Lee, MD, PhD
Department of Information Medicine
Asan Medical Center
University of Ulsan College of Medicine
88, Olympic-ro 43-gil, Songpa-gu
Seoul, 05505
Republic of Korea
Phone: 82 3010 3350
Fax: 82 3010 3360
Email: rufiji@gmail.com

Abstract

Background: Mobile apps for weight loss provide users with convenient features for recording lifestyle and health indicators; they have been widely used for weight loss recently. Previous studies in this field generally focused on the relationship between the cumulative nature of self-reported data and the results in weight loss at the end of the diet period. Therefore, we conducted an in-depth study to explore the relationships between adherence to self-reporting and weight loss outcomes during the weight reduction process.

Objective: We explored the relationship between adherence to self-reporting and weight loss outcomes during the time series weight reduction process with the following 3 research questions: “How does adherence to self-reporting of body weight and meal history change over time?”, “How do weight loss outcomes depend on weight changes over time?”, and “How does adherence to the weight loss intervention change over time by gender?”

Methods: We analyzed self-reported data collected weekly for 16 weeks (January 2017 to March 2018) from 684 Korean men and women who participated in a mobile weight loss intervention program provided by a mobile diet app called Noom. Analysis of variance (ANOVA) and chi-squared tests were employed to determine whether the baseline characteristics among the groups of weight loss results were different. Based on the ANOVA results and slope analysis of the trend indicating participant behavior along the time axis, we explored the relationship between adherence to self-reporting and weight loss results.

Results: Adherence to self-reporting levels decreased over time, as previous studies have found. BMI change patterns (ie, absolute BMI values and change in BMI values within a week) changed over time and were characterized in 3 time series periods. The relationships between the weight loss outcome and both meal history and self-reporting patterns were gender-dependent. There was no statistical association between adherence to self-reporting and weight loss outcomes in the male participants.

Conclusions: Although mobile technology has increased the convenience of self-reporting when dieting, it should be noted that technology itself is not the essence of weight loss. The in-depth understanding of the relationship between adherence to self-reporting

and weight loss outcome found in this study may contribute to the development of better weight loss interventions in mobile environments.

(J Med Internet Res 2020;22(8):e17521) doi:10.2196/17521

KEYWORDS

weight loss; self-reporting; adherence; mobile weight loss intervention; diet

Introduction

Mobile apps are widely used for tracking weight loss [1-3]. A major feature of these apps is providing users with self-reporting capabilities regarding their lifestyle and health-related measurements [1-3].

Self-reporting has been recognized as a very important behavioral treatment method for weight loss [4-9] because it provides diet participants with a continuous feedback loop that can modify behaviors to achieve goals through the process of observation, evaluation, and reinforcement [5,10]. Indeed, several empirical studies have shown a positive relationship between adherence to self-reporting and satisfactory weight loss outcomes [4,5,7]. For these reasons, considerable research has been conducted regarding adherence to self-reporting in the field of weight loss research [6,8,11]. These studies commonly suggest the expectation that the use of mobile technology will improve adherence to self-reporting [6,7,12].

Compared to traditional tools, mobile technology's portability has made it easier and more effective to monitor and control one's weight [4,9,13-15]. In addition, the use of mobile technology such as smartphones and wearable devices enables researchers to explore new topics by collecting data that could not previously be gathered. Specifically, a representative research question related to new data may demonstrate the relationship between characteristics of adherence to self-reporting and weight loss outcomes [4,15,16]. Another research question may explore the differences in adherence to self-reporting during the diet in terms of demographic characteristics such as gender and sociocultural aspects such as race [17-22]. These studies have found that adherence to self-reporting decreases over time [4,16,21,23,24] and is dependent on demographics [17-22].

However, an in-depth understanding is lacking regarding the relationship between weight loss outcomes and the dynamics of adherence to self-reporting according to specific groups. Attempts at an in-depth understanding of this relationship is very important for several reasons. First, given that demographic characteristics affect both adherence to self-reporting and dietary approaches [17–22], researchers may gain a fundamental understanding of how certain groups respond to weight loss interventions. Second, researchers may understand the mechanism of weight changes during the weight loss process that involves a series of weight gain, loss, and regain cycles [21-23]. Ultimately, gaining an in-depth understanding of these matters will contribute to designing better weight loss programs suitable for mobile apps.

Thus, this study aimed to analyze the relationship between weight loss outcomes and the dynamics of adherence to self-reporting. To that end, we analyzed self-reported data that were aggregated on a weekly basis for 16 weeks, from January 2017 to March 2018, from 684 Korean men and women who participated in a mobile weight loss intervention program provided by a commercial smartphone app.

The following 3 research questions were explored and answered: How does adherence to self-reporting on body weight and meal history change over time? How do weight loss outcomes depend on weight changes over time? How does adherence to the weight loss intervention change over time by gender?

Methods

Mobile Weight Loss Intervention

The mobile intervention program for this study was delivered through a commercially available mobile coaching program called Noom [25]. During the 16-week program, users were asked to log their meals, exercise, and weight data using this app [25]. In-app articles about healthy dietary intake and promotions to encourage physical activities were provided on a daily basis [25]. The articles were adapted by physicians, clinical dietitians, and clinical psychologists and based on studies provided by the American Diabetes Association, American Heart Association, and Centers for Disease Control and Prevention.

All coaches were registered dietitians, and they were supervised weekly by a clinical psychologist. They helped users set weekly goals and provided personalized feedback based on each user's lifestyle. Messages were sent to the users at least twice per week.

Measurements

Weight Entry Trends

The frequency of weight entries was analyzed. In most health care app research programs, user adherence to a certain intervention has often been measured by the frequency of self-reporting [4,24,26,27]. Therefore, the number of body weight entries per week served as a trustworthy variable in measuring the degree of adherence to the weight loss intervention in this research [4,24,26,27].

The study participants reported their weight through the app. After reporting, weight-related menu-usage logs were automatically stored on a server. The collected usage log data were summed at weekly intervals to generate data indicating adherence to weight-reporting activity over 16 weeks, for each subject.

BMI Trends

BMI is often used to determine whether a person is within a healthy weight range based on their height [4,14,28]. Since this

is calculated by dividing weight by height (kg/m^2), the BMI value represents relative weight with respect to height [29,30]. In other words, BMI values indicate whether a person is "underweight," "healthy," "overweight," or "obese" [4,14,28]. Because of its ease of interpretation and satisfactory explanatory power, BMI has been used instead of basic weight measurements for many diet-related studies on web and mobile weight tracking environments [4,14,25,31-33].

Participants reported their body weight daily through Noom. The daily BMI values for each participant were calculated by dividing the daily reported weights by their initially entered height values, and the daily results were averaged on a weekly basis.

BMI delta values were used to identify weekly weight fluctuation trends [26,34,35]. For this analysis, BMI delta was defined as the difference between the largest and smallest BMI values in a given week. As such, for each participant, the average weekly BMI values representing their overall obesity level and the weekly BMI deltas indicating their degree of weight fluctuation over 16 weeks were used for the statistical analysis.

Meal History Entry Trends

As with weight entries, the frequency of self-reported meal history entries has frequently been used in prior studies to address the level of adherence to weight loss intervention programs [4,32]. In this study, participants entered not only their main meals but also snacks in between meals. As participants recorded food items by searching for them within Noom's food database, their eating habits and selections could be grouped in detail by food types.

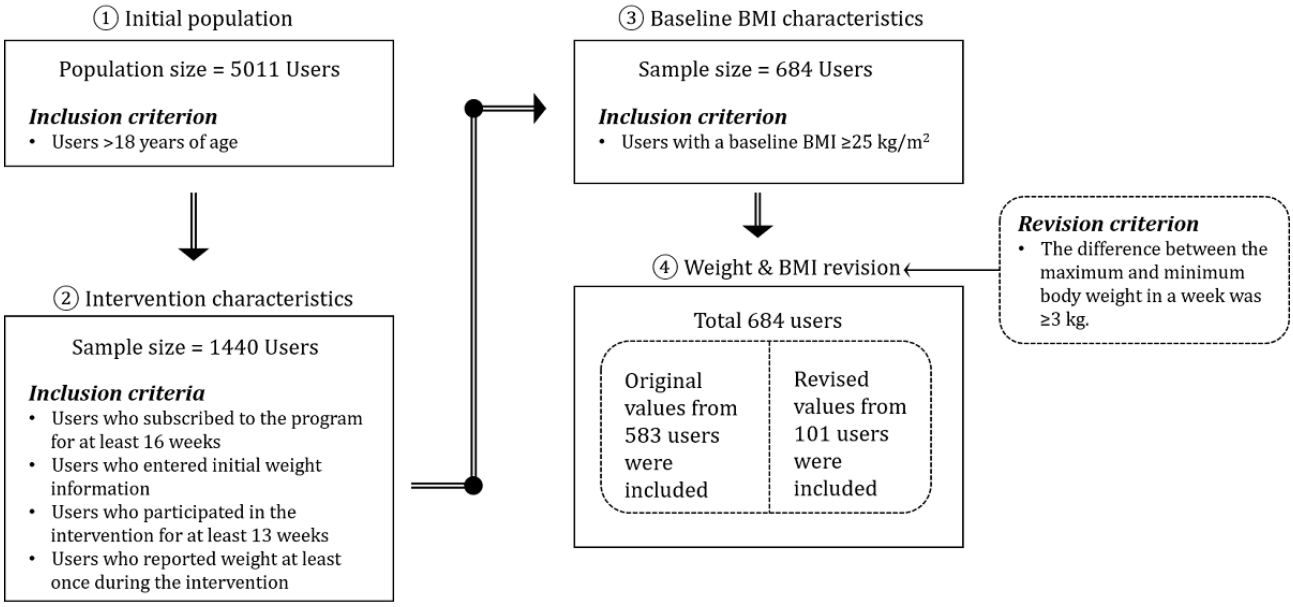
These food types were categorized as green, yellow, and red groups based on their caloric density, that is, calories per gram or milliliter [29,30,36]. In general, foods in the red group have very high caloric density and include foods high in fat and oil [29,30,36]. Foods categorized as yellow have moderate caloric density; they are less fatty and oily than the red group but still not optimal [29,30,36]. Foods in the green group have very low caloric density and are comprised mostly of natural fruits and vegetables [29,30,36]. The applied food type criteria according to caloric density are described in Multimedia Appendix 1.

Variables representing food intake by food type were generated based on the aforementioned criteria. Furthermore, food intake was reported 6 times per day (ie, breakfast, morning snacks, lunch, afternoon snacks, dinner, and evening snacks). This provided a substantial amount of data regarding the changes in food intake for the 3 food groups of each participant over 16 weeks.

Study Sample

This study was approved by the Institutional Review Board of the Asan Medical Center, Korea (No. 2017-1253). Figure 1 shows the study sample selection procedure. Initially, the study population comprised 5011 users over the age of 18 years who participated in the coaching program from January 1, 2017 to March 5, 2018 [4,37]. Of these, 1440 users were included in the study based on 4 intervention characteristics [4,9,38]. First, according to prior research, users who subscribed to the program for at least 16 weeks were included in the study [4,9,38]. Second, only users with initial weight information were included in the study, to analyze baseline weight. Third, only those who reported their weight at least once among those who participated in the intervention for at least 13 weeks were included in the study [4]. Finally, 684 overweight and obese users with BMIs >25 kg/m^2 were selected for analysis [4,36].

Figure 1. Study sample selection process.



Participants measured their weight on their scales and then reported the weight via the app. To prevent incorrect input that may occur when users report these weight values, certain reported values were corrected. As no previous studies provided any clear criteria for mitigating data errors, a group of three experts consisting of doctors and health care professionals set

the error criteria and reviewed the data. It was assumed that there could be potential errors if the difference between the maximum weight and minimum weight per week was ≥3 kg. The daily weight and BMI values of 101 cases corresponding to this assumption were not included in the weekly averages.

Statistical Analysis

To analyze the self-reported time-series pattern based on weight loss groups, subjects were divided into 3 groups according to recommendations from prior research [4]. The 3 groups consisted of those with weight losses <5%, of 5%-10%, and >10% [4].

Analysis of variance (ANOVA) and chi-squared tests were performed to analyze whether the baseline characteristics of the 3 groups were statistically different. Furthermore, ANOVAs along the time axis were performed to determine whether each measurement (ie, BMI, weight entry, and meal history entry trends) of the 3 groups (ie, <5%, 5%-10%, >10%) within gender were statistically different over 16 weeks. *P* values were adjusted with the false discovery rate using the Benjamini-Hochberg method, because multiple tests based on the week and groups may yield 5% false positives at the 5% significance level [39]. Finally, through slope analysis of each measurement calculated at the aggregated level, the overall time-series trends for the measurements were analyzed.

Results

Participant Characteristics

Table 1 shows the test results for the statistical differences in baseline characteristics between the 3 outcome groups. In the case of gender, although few men (19/218) and women (39/466) belonged to the group that lost >10% weight, there were no statistical associations among the 3 groups. For age, young participants (ie, mean 32.09 years old) seemed to lose the most weight; however, there were no statistical differences among the 3 groups. In the case of base weight status, the few obese (17/178) participants lost >10% of their weight; however, no statistical associations were found among the 3 groups.

Table 1. Participant characteristics for the total sample and according to outcome group.

Characteristics	Total sample	Outcome group according to level of weight loss			*P* value
		<5%	5%-10%	>10%	
Gender, n					
Male	218 (31.9)[a]	126	73	19	.39[b]
Female	466 (68.1)[a]	294	133	39	
Age (years), mean (SD)	33.31 (7.6)	33.7	32.87	32.09	.19[c]
Baseline weight status, n					
Obese	178 (26)[a]	112	49	17	.62[b]
Overweight	506 (74)[a]	308	157	41	

[a]n (%).

[b]Tested differences between the 3 groups using chi-squared tests.

[c]Tested differences between the 3 groups using analysis of variance (ANOVA).

Pattern Analysis

Weight Entries

Figure 2 shows the number of weight entries reported by user each week for the 16 weeks. The number of weight entries decreased over time in all 3 groups even though the decreases in the slopes were different (ie, −0.16 for the group who lost >10% of their starting weight, −0.13 for the group who lost 5%-10%, and −0.12 for the group who lost <5%). The members of the group with >10% weight loss recorded their weight more often in Noom than the other groups (ie, the groups who lost ≤10%). This trend at the aggregated level was statistically significant throughout the entire study period (Table 2). Moreover, the statistical significance of the trend was noticeable in women (Table 2).

Figure 2. The number of weight entries on a weekly basis over 16 weeks; the colored bands show the 95% CI.

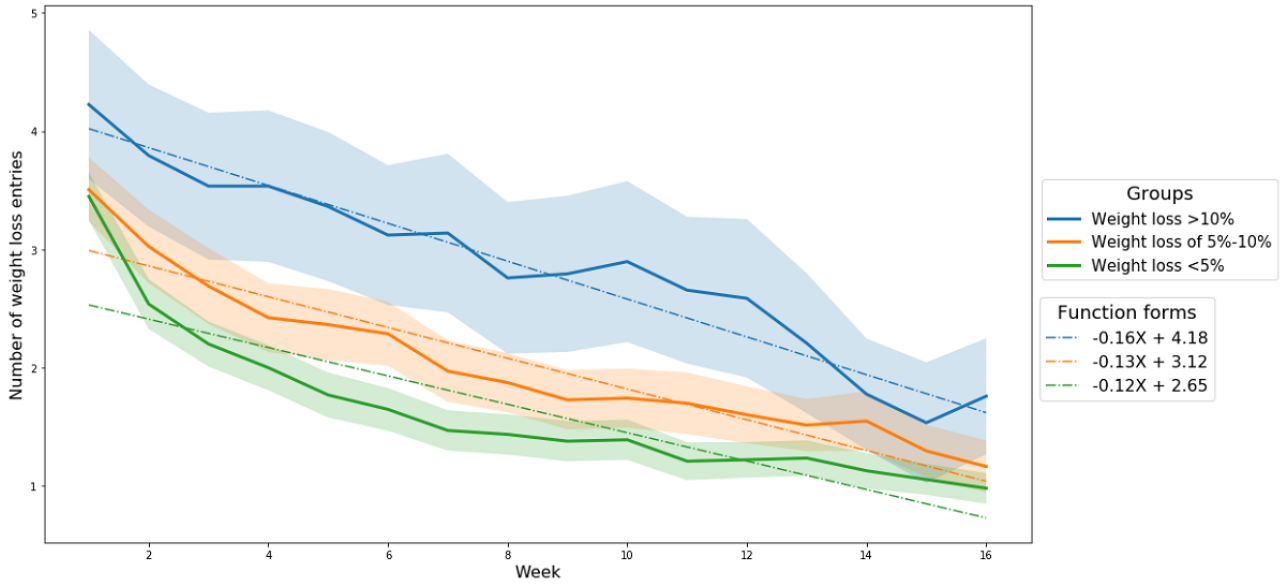


Table 2. Significance of analysis of variance (ANOVA) test results comparing the differences between groups for 16 weeks. *P* values were adjusted by the false discovery rate using the Benjamini-Hochberg method, and the actual *P* values are available in Multimedia Appendix 2.

Group	Week															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Weight entries																
Total[a]	**[b]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	**[b]	***[c]
Men[a]					**[b]	**[b]				*[d]						
Women[a]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]		***[c]
BMI values																
Total[a]					*[d]	*[d]	**[b]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]
Men[a]								*[d]	*[d]			**[b]	**[b]	**[b]	**[b]	*[d]
Women[a]					*[d]			***[c]	**[b]	**[b]	***[c]	***[c]	***[c]	***[c]	***[c]	***[c]
BMI delta[d]																
Total[a]	***[c]	***[c]	***[c]	*[d]	***[c]	***[c]	*[d]	**[b]	*[d]		***[c]	*[d]				
Men[a]		**[b]														
Women[a]	***[c]	***[c]	**[b]		***[c]	***[c]	*[d]	*[d]			***[c]	**[b]				
Meal history entries																
Total[a]		*[d]	*[d]	**[b]	*[d]	***[c]	***[c]	***[c]	***[c]	***[c]	**[b]	***[c]	*[d]	**[b]	*[d]	**[b]
Men[a]																
Women[a]				*[d]	*[d]	**[b]	***[c]	**[b]	**[b]	***[c]		***[c]	*[d]	**[b]	**[b]	**[b]

[a]When the null hypotheses were tested using ANOVA, the three groups (weight loss >10%, weight loss of 5%-10%, and weight loss <5%) had the same value.

[b]*P*<.05

[c]*P*<.01

[d]*P*<.1

[e]Defined as the difference between the maximum and minimum BMIs in a given week.

BMI Values

Figure 3 shows the trend in the BMI values for the 3 groups over time at the aggregated level. The differences in the BMI values between the 3 groups was statistically significant after week 5 (Table 2). Furthermore, the statistical difference was more noticeable in women than in men (Table 2).

In all 3 groups, the BMI delta values — representing the degree of weight fluctuation — were only high during the initial 3 weeks (Figure 4). The group that lost >10% weight had larger BMI delta values during the entire study period. The difference in BMI delta values among the 3 groups was the most statistically significant during the first several weeks. However, most of the differences in values became insignificant over time (Table 2). Moreover, the BMI delta was more statistically significant in women (Table 2).

Figure 3. BMIs over 16 weeks according to weight loss outcome groups; the colored bands show the 95% CI.

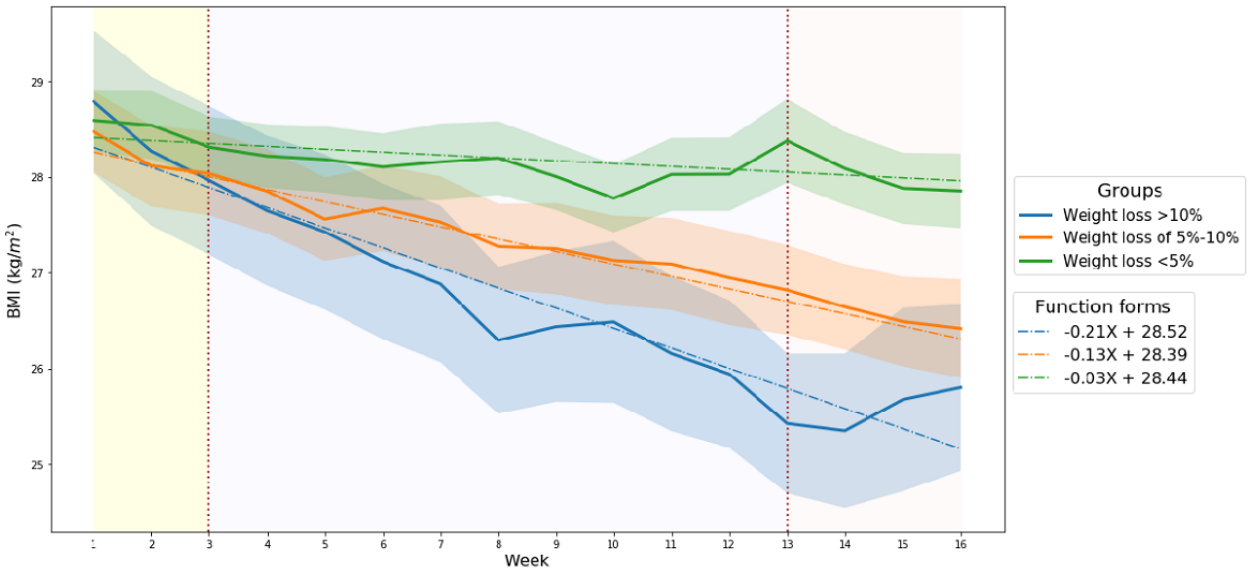
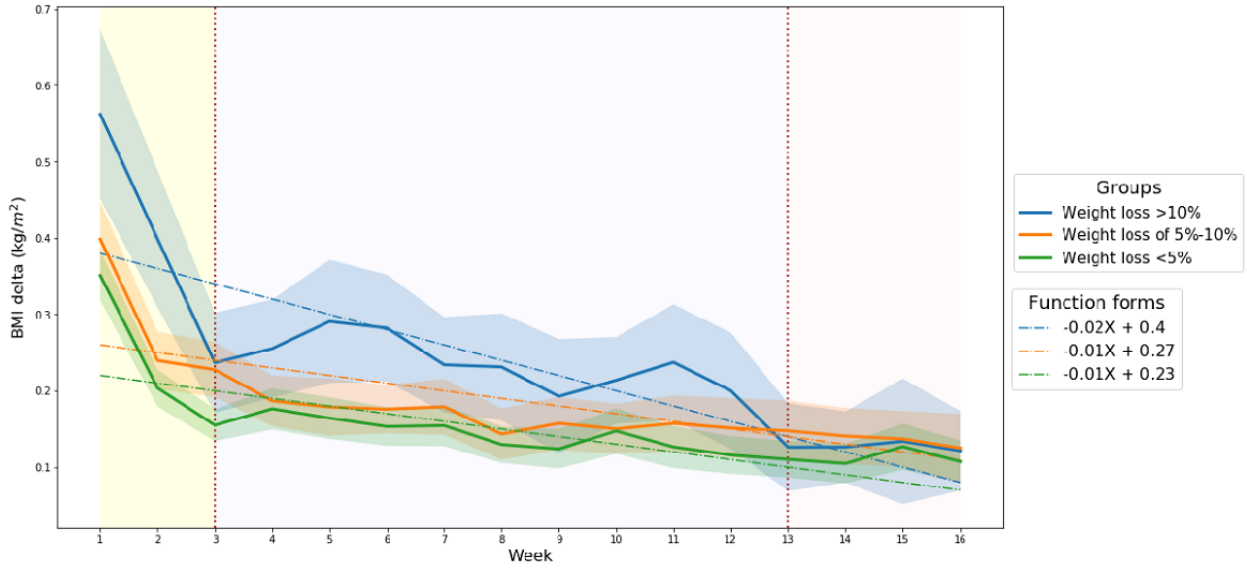


Figure 4. BMI delta over 16 weeks according to weight loss outcome groups; the colored bands show the 95% CI.



Meal History Entries

As seen in Figure 5, the number of meal history entries decreased over time in all 3 groups (slope values of −0.67, −0.77, and −0.79 for the groups with >10% weight loss, 5%-10% weight loss, and <5% weight loss, respectively). The 2 groups that reduced their weight by ≥5% had a higher total number of meal history entries than the less successful group, and the differences in total number of meal history entries among the 3 groups were statistically significant during most of the study period (Table 2). No significant differences in the total number of meal history entries were found for men (Table 2). On the other hand, for women, significant differences were found during most of the study period (Table 2).

Table 3 shows food intakes divided into the green, yellow, and red groups for 16 weeks for women. Women had many statistical differences between the yellow and red food groups, especially at lunch and dinner. In particular, when the intake of foods in

the red category was averaged over all study periods, statistical differences between the 3 groups were significant for all meal types (Table 3). However, for men, there were no statistical differences between weight loss outcome groups and food groups over the 16 weeks.

Figure 5. Number of meal history entries on a weekly basis over 16 weeks; the colored bands show the 95% CI.

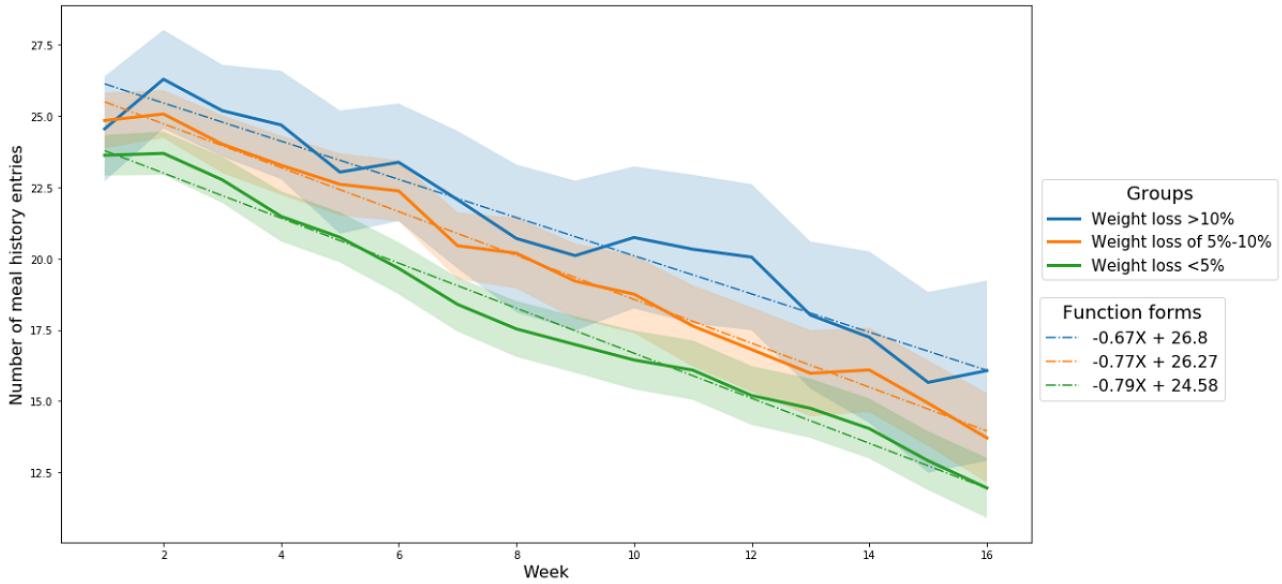


Table 3. Significance of analysis of variance (ANOVA) test results comparing the differences in intake amount by food group (green, yellow, and red) in women for 16 weeks. *P* values were adjusted with the false discovery rate using the Benjamini-Hochberg method, and the actual *P* values are available in Multimedia Appendix 3.

Food group	Week																Avg[a]
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Green																	
BRE[b]			**[c]														
MOR.S[d]																	
LUN[e]																	
AFT.S[f]																	
DIN[g]																	
EVE.S[h]																	
Yellow																	
BRE											**[c]						***[i]
MOR.S																	
LUN				**[c]					**[c]		**[c]	**[c]		**[c]			***[i]
AFT.S																	
DIN	**[c]			*[j]		**[c]	**[c]		*[j]		*[j]	**[c]			**[c]		***[i]
EVE.S																	
Red																	
BRE				**[c]	*[j]					*[j]	**[c]						***[i]
MOR.S						**[c]			*[j]								**[c]
LUN			**[c]	**[c]	*[j]	**[c]	*[j]	**[c]	*[j]	*[j]	*[j]	**[c]				**[c]	***[i]
AFT.S		**[c]	*[j]									**[c]					***[i]
DIN		**[c]	**[c]	**[c]	**[c]			*[j]	*[j]	**[c]	**[c]	*[j]	**[c]		**[c]		***[i]
EVE.S	*[j]						*[j]			**[c]							***[i]

[a]Avg: grouped average of food intake over 16 weeks; the *P* value for was not adjusted by the false discovery rate using the Benjamini-Hochbert method. When the null hypotheses were tested using ANOVA, the three groups (weight loss >10%, weight loss of 5%-10%, and weight loss <5%) had the same value.

[b]BRE: breakfast.

[c]*P*<.05

[d]MOR.S: morning snack.

[e]LUN: lunch.

[f]AFT.S: afternoon snack.

[g]DIN: dinner.

[h]EVE.S: evening snack.

[i]*P*<.01

[j]*P*<.1

Discussion

Principal Findings

Decreased Level of Adherence Over Time

Previous studies showed that the degree of adherence to self-reporting in mobile weight loss intervention programs decreases over time [4,6,23]. Going one step further, this study found that adherence decreases over time regardless of weight loss outcomes. Specifically, both the total weight and meal history entries decreased gradually over 16 weeks (Figures 2 and 5). Prior studies had assumed that mobile technology utilization could overcome the low adherence to self-reporting that occurs when utilizing rudimentary paper tools during weight loss interventions [6,13]. Although mobile technology clearly makes self-reporting easier, the results of this study revealed

that there is a natural tendency to lower one's adherence to self-reporting over time regardless of weight loss outcomes.

Among the many factors that hinder long-term adherence, the most relevant to this study may be the lack of serious awareness of current health conditions [37,40,41]. In general, most people are not likely to feel any ominous threats even if they do not lose their weight right away, and they are prone to stop dieting. Therefore, to promote long-term adherence, it is necessary to provide an environment in which diet participants recognize the critical importance of weight management [42,43]. Given the lessons learned from previous diet studies that have shown that digital push notifications have no effect on adherence, reminding participants regularly about the importance of dieting may be ineffective [44,45]. Instead, there seems to be a need for further research to provide divergent strategies for people with obesity.

Conversely, it may be entirely reasonable to simply admit that humans are less committed to interventions over time, especially for slow and arduous tasks like trying to lose weight. In other words, it may be wise to avoid long-term adherence issues that are unlikely to be solved. Rather, the best practice may be to focus on understanding the characteristics of adherence with latent patterns. Recent data science algorithms may be employed to understand specific patterns of adherence. Specifically, the use of time series decomposition algorithms that eliminate the inevitable long-term declines in adherence may be used for exploring repeated patterns of adherence [46,47]. Then, demonstrating the association between the pattern and weight loss outcomes can provide new insights.

Characteristics of BMI Change Over Time

According to the results of this empirical analysis, BMI changes can be characterized in 3 time series periods (Figures 3 and 4). The first interval is from week 1 to week 3, where the overall BMI delta value started high and fell rapidly over time and the overall BMI value was high in the initial stages. High values for both BMI and BMI delta may indicate substantial weight fluctuations with low weight losses within a week. This phenomenon may be due to the participants' biological responses to homeostasis (ie, steady weight) [27]. It takes time for the human body to adapt to a diet (ie, limiting caloric intake and increasing energy expenditure through exercise) [27,48,49]. During this adaptation process, human weight can fluctuate based on a variety of factors, including hormones, psychology, and physiology [21-23]. Although the period of adaptation may vary depending on certain factors (eg, race, diet style), several empirical studies have found that it takes human bodies approximately 2 to 4 weeks to adapt to a diet [48,49]. Therefore, participants in this study might have experienced weight fluctuations during the dietary adaptation process period.

The second interval is between weeks 3 and 13. This interval was characterized by differences in BMI that increased over time among the outcome groups. Further, most BMI deltas were statistically different among the groups (Figures 3 and 4, Table 2). In this period, the differences in weight loss between the successful and unsuccessful diet groups were noticeable. Specifically, the successful group had high BMI deltas and low BMI values, which may indicate that the successful participants lost weight. On the other hand, the unsuccessful group during the same period had low BMI deltas and high BMI values, thereby suggesting that the unsuccessful participants rarely lost any weight at all.

The last interval is between weeks 13 and 16. This interval was characterized by no differences in BMI deltas between the groups (Figure 4). Another feature during this period was the manifestation of weight increases in the group that lost the most weight (>10%). This may indicate that the group members stopped trying to lose weight, because people do not lose weight beyond their ideal weight. From a physiological point of view, the feature may be attributed to the widely known "yo-yo" effect that commonly occurs after rapid weight loss [45-47]. The yo-yo effect is defined as a rapid weight gain after a diet. Indeed, many previous studies have shown that the majority of diet participants regain their weight over the long term [45-47]. In this study, the yo-yo effect may have occurred in participants who had dietary laxity as the mobile weight loss intervention program was nearing its end in week 16.

Different Patterns for Female and Male Participants

Gender effects on the differences in self-reporting adherence have often been discussed in previous studies of weight loss interventions [4,6,22]. Going one step further, this study explored the time-dependent relationships among adherence to self-reporting, gender effects, and weight loss outcomes.

Unlike female participants, male participants had little differences in the number of weight entries among the outcome groups during the 16 weeks (Table 2). Furthermore, in terms of the number of meal history entries, there were no statistical differences in the values among the outcome groups for male participants (Table 2). However, statistical differences in BMI values among the 3 groups were found around the last several weeks (Table 2).

These results suggest that there was no statistical association between adherence to self-reporting and weight loss outcomes in the male population. This contradicts some previous studies that showed positive relationships between the two [6,7,50]. Inconsistencies in the results can be addressed as follows. Considering that the tendency to self-report is influenced by various factors, such as gender and race [6,22,51], Korean adult men may not be as inclined to self-report. In other words, they may have avoided timely self-reporting regardless of the success of the diet. As self-reporting itself has no direct effect on weight loss, male participants might have been able to lose weight without consistent self-reporting if they managed their diet well and consistently exercised.

Another interesting finding in this analysis is that the male participants, unlike the female participants, demonstrated no statistical difference in the intake of food types among the 3 weight loss groups. Previous studies found that men were insensitive to food intake during their diets, while women were very sensitive [52,53]. Therefore, the results may have captured the food intake characteristics of dieting men during mobile interventions. However, it should be recognized that these characteristics may also depend on other factors such as age or ethnicity [17–19]. Therefore, further research is needed in

different settings to identify the factors that may contribute to gender differences in self-reporting and weight loss outcomes.

Limitations

This study has limitations that may inspire future research. First, analysis results cannot be free from errors, as the data are self-reported. Particularly, the weight statistics used to divide participants into 3 weight loss groups were self-reported. Although reporting of actual information was encouraged in the research agreements, potential mistakes could have arisen for which participants could have been unaware while reporting their weight. Furthermore, scales that the participants used may not have been reliable. To prevent these potential problems, some weight values were modified by applying logic during preprocessing. However, the applied logic was empirical and somewhat arbitrary. In addition, all reported data other than weight information, such as gender and age, were assumed to be true and consistent throughout the study. No correction was made for misreported data. Further research should be conducted in more sophisticated environments by collecting data under rigorous verification by researchers or health professionals.

Second, there is a potential limitation related to the approach and data analyses in this study. First, the significance of the *P* value may have been underpowered. In particular, only 8.7% (19/218) of the male participants lost more than 10% of their weight (Table 1). Further, only 9.5% (17/178) of the participants who were obese before they started the experiment lost more than 10% of their weight (Table 1). The small sample size could have potentially undermined statistical significance [54]. Second, the interval division criteria applied to the analysis of BMI and BMI delta values are somehow arbitrary. Although methodologies exist for determining trend-change points, such methodologies may not provide adequate solutions in short-term multivariate time-series environments (ie, 16 weeks with 6 series) [55]. Thus, a more rigorous data analysis setting should be established by collecting more data and applying sophisticated approaches.

Third, there may be potential limitations in the experimental setting assumptions. For instance, although the registered dietitians in the program did not recommend the use of weight-reducing drugs like diuretics, some participants might have relied on some medicines that have a huge effect on weight loss. Furthermore, the study design did not include a control group, which might have undermined the ability to evaluate the efficacy of the mobile weight loss intervention. For example, the weights of the subjects who canceled the program prior to week 13 were not analyzed. By analyzing such data in a control group, the efficacy of the mobile intervention can be evaluated in a more rigorous manner. In addition, significant differences in food intake reporting by weight loss outcomes in female participants may have been caused by the Hawthorne effect [56]. In other words, the significant difference may be the result of a keen sense of being observed in the experiment rather than gender. Therefore, further experimentations with more sophisticated settings are essential.

Fourth, two factors may have undermined the representativeness of self-reported data for measurements. First, the number of self-reported entries may not have measured spontaneous adherence levels, because the app provides push notifications to remind users to report their data. Additionally, as the degree of self-reporting decreases over time, the data may be insufficient for representing weight status. Particularly, the average number of weight entries in the 16th week for the lowest weight loss group (<5%) was less than 1.5 per participant. When body weight is reported only once per week, potentially biased values may have been reported depending on the timing of the report (eg, after exercise or immediately after main meals, mornings versus evenings). Therefore, for studies requiring self-reporting, further discussion is required regarding data collection and analytical approaches.

Conclusions

Mobile technology has increased the convenience of self-reporting when dieting [6,13]. However, it should be noted that technology is not the essence of weight loss; rather, it provides a stimulus through simplified self-reporting that may have positive effects on weight loss. Therefore, further research should be conducted to determine ways to couple mobile technology with human nature to foster more effective dieting and consistently healthy lifestyles.

Acknowledgments

This study was supported by a grant from the Research and Development Project, Ministry of Trade, Industry, and Energy, Republic of Korea (no. 20004503).

Conflicts of Interest

YK is employed at Noom and has conflicts of interest. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Food-type criteria according to caloric density.
[PDF File (Adobe PDF File), 68 KB - jmir_v22i8e17521_app1.pdf]

Multimedia Appendix 2
P values from the ANOVA tests comparing the differences between groups for 16 weeks.
[PDF File (Adobe PDF File), 87 KB - jmir_v22i8e17521_app2.pdf]

Multimedia Appendix 3
P values from the ANOVA tests comparing the differences in intake amount by food group (green, yellow, and red) in women for 16 weeks.
[PDF File (Adobe PDF File), 176 KB - jmir_v22i8e17521_app3.pdf]

References

1. Lee W, Chae YM, Kim S, Ho SH, Choi I. Evaluation of a mobile phone-based diet game for weight control. J Telemed Telecare 2010 Jun;16(5):270-275. [doi: 10.1258/jtt.2010.090913] [Medline: 20558620]
2. Toro-Ramos T, Lee D, Kim Y, Michaelides A, Oh TJ, Kim KM, et al. Effectiveness of a Smartphone Application for the Management of Metabolic Syndrome Components Focusing on Weight Loss: A Preliminary Study. Metab Syndr Relat Disord 2017 Nov;15(9):465-473. [doi: 10.1089/met.2017.0062] [Medline: 29035677]
3. Jeon E, Park H. Development of a smartphone application for clinical-guideline-based obesity management. Healthc Inform Res 2015 Jan;21(1):10-20 [FREE Full text] [doi: 10.4258/hir.2015.21.1.10] [Medline: 25705553]
4. Painter SL, Ahmed R, Hill JO, Kushner RF, Lindquist R, Brunning S, et al. What Matters in Weight Loss? An In-Depth Analysis of Self-Monitoring. J Med Internet Res 2017 May 12;19(5):e160 [FREE Full text] [doi: 10.2196/jmir.7457] [Medline: 28500022]
5. Laitner MH, Minski SA, Perri MG. The role of self-monitoring in the maintenance of weight loss success. Eat Behav 2016 Apr;21:193-197 [FREE Full text] [doi: 10.1016/j.eatbeh.2016.03.005] [Medline: 26974582]
6. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
7. Baker RC, Kirschenbaum DS. Self-monitoring may be necessary for successful weight control. Behavior Therapy 1993;24(3):377-394. [doi: 10.1016/S0005-7894(05)80212-6]
8. Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity (Silver Spring) 2015 Feb 17;23(2):256-265 [FREE Full text] [doi: 10.1002/oby.20946] [Medline: 25521523]
9. Cadmus-Bertram L, Marcus BH, Patterson RE, Parker BA, Morey BL. Use of the Fitbit to Measure Adherence to a Physical Activity Intervention Among Overweight or Obese, Postmenopausal Women: Self-Monitoring Trajectory During 16 Weeks. JMIR Mhealth Uhealth 2015 Nov 19;3(4):e96 [FREE Full text] [doi: 10.2196/mhealth.4229] [Medline: 26586418]
10. Kanfer FH. Self-monitoring: Methodological limitations and clinical applications. Journal of Consulting and Clinical Psychology 1970;35(2):148-152. [doi: 10.1037/h0029874]
11. Burke LE, Warziski M, Starrett T, Choo J, Music E, Sereika S, et al. Self-monitoring dietary intake: current and future practices. J Ren Nutr 2005 Jul;15(3):281-290. [doi: 10.1016/j.jrn.2005.04.002] [Medline: 16007557]
12. Helsel DL, Jakicic JM, Otto AD. Comparison of techniques for self-monitoring eating and exercise behaviors on weight loss in a correspondence-based intervention. J Am Diet Assoc 2007 Oct;107(10):1807-1810. [doi: 10.1016/j.jada.2007.07.014] [Medline: 17904942]
13. Beasley J. The pros and cons of using PDAs for dietary self-monitoring. J Am Diet Assoc 2007 May;107(5):739; author reply 739-739; author reply 740. [doi: 10.1016/j.jada.2007.03.023] [Medline: 17467366]
14. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
15. Helander EE, Vuorinen A, Wansink B, Korhonen IKJ. Are breaks in daily self-weighing associated with weight gain? PLoS One 2014 Nov;9(11):e113164 [FREE Full text] [doi: 10.1371/journal.pone.0113164] [Medline: 25397613]
16. Webber KH, Tate DF, Ward DS, Bowling JM. Motivation and its relationship to adherence to self-monitoring and weight loss in a 16-week Internet behavioral weight loss intervention. J Nutr Educ Behav 2010 May;42(3):161-167. [doi: 10.1016/j.jneb.2009.03.001] [Medline: 20138583]
17. Li W, Youssef G, Procter-Gray E, Olendzki B, Cornish T, Hayes R, et al. Racial Differences in Eating Patterns and Food Purchasing Behaviors among Urban Older Women. J Nutr Health Aging 2017 Nov 5;21(10):1190-1199 [FREE Full text] [doi: 10.1007/s12603-016-0834-7] [Medline: 29188879]
18. Wang Y, Chen X. How much of racial/ethnic disparities in dietary intakes, exercise, and weight status can be explained by nutrition- and health-related psychosocial factors and socioeconomic status among US adults? J Am Diet Assoc 2011 Dec;111(12):1904-1911 [FREE Full text] [doi: 10.1016/j.jada.2011.09.036] [Medline: 22117667]
19. Wirtz JG, Kulpavaropas S. The effects of narrative and message framing on engagement and eating intention among a sample of adult Hispanics. J Nutr Educ Behav 2014 Sep;46(5):396-400. [doi: 10.1016/j.jneb.2013.12.005] [Medline: 24517968]
20. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo J, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov 07;6(1):34563 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]
21. Comulada WS, Swendeman D, Koussa MK, Mindry D, Medich M, Estrin D, et al. Adherence to self-monitoring healthy lifestyle behaviours through mobile phone-based ecological momentary assessments and photographic food records over

6 months in mostly ethnic minority mothers. Public Health Nutr 2017 Dec 04;21(4):679-688. [doi: 10.1017/s1368980017003044]
22. Hollis JF, Gullion CM, Stevens VJ, Brantley PJ, Appel LJ, Ard JD, Weight Loss Maintenance Trial Research Group. Weight loss during the intensive intervention phase of the weight-loss maintenance trial. Am J Prev Med 2008 Aug;35(2):118-126 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.013] [Medline: 18617080]
23. Glover M, Nolte M, Wagemakers A, McRobbie H, Kruger R, Breier BH, et al. Adherence to daily dietary and activity goals set within a Māori and Pacific weight loss competition. BMC Obes 2019 Mar 4;6(1):6 [FREE Full text] [doi: 10.1186/s40608-019-0228-6] [Medline: 30867932]
24. Asbjørnsen RA, Smedsrød ML, Solberg Nes L, Wentzel J, Varsi C, Hjelmesæth J, et al. Persuasive System Design Principles and Behavior Change Techniques to Stimulate Motivation and Adherence in Electronic Health Interventions to Support Weight Loss Maintenance: Scoping Review. J Med Internet Res 2019 Jun 21;21(6):e14265 [FREE Full text] [doi: 10.2196/14265] [Medline: 31228174]
25. Kim Y, Oh B, Shin H. Effect of mHealth With Offline Antiobesity Treatment in a Community-Based Weight Management Program: Cross-Sectional Study. JMIR Mhealth Uhealth 2020 Jan 21;8(1):e13273 [FREE Full text] [doi: 10.2196/13273] [Medline: 31961335]
26. Foreyt JP, Brunner RL, Goodrick GK, Cutter G, Brownell KD, St. Jeor ST. Psychological correlates of weight fluctuation. Int. J. Eat. Disord 1995 Apr;17(3):263-275. [doi: 10.1002/1098-108x(199504)17:3<263::aid-eat2260170307>3.0.co;2-n]
27. Maclean PS, Bergouignan A, Cornier M, Jackman MR. Biology's response to dieting: the impetus for weight regain. Am J Physiol Regul Integr Comp Physiol 2011 Sep;301(3):R581-R600 [FREE Full text] [doi: 10.1152/ajpregu.00755.2010] [Medline: 21677272]
28. Voelker R. Improved use of BMI needed to screen children for overweight. JAMA 2007 Jun 27;297(24):2684-2685. [doi: 10.1001/jama.297.24.2684] [Medline: 17595264]
29. Beleigoli AM, Andrade AQ, Cançado AG, Paulo MN, Diniz MDFH, Ribeiro AL. Web-Based Digital Health Interventions for Weight Loss and Lifestyle Habit Changes in Overweight and Obese Adults: Systematic Review and Meta-Analysis. J Med Internet Res 2019 Jan 08;21(1):e298 [FREE Full text] [doi: 10.2196/jmir.9609] [Medline: 30622090]
30. Gray DS, Fujioka K. Use of relative weight and body mass index for the determination of adiposity. Journal of Clinical Epidemiology 1991 Jan;44(6):545-550. [doi: 10.1016/0895-4356(91)90218-x]
31. Doucet E, St-Pierre S, Alméras N, Tremblay A. Relation between appetite ratings before and after a standard meal and estimates of daily energy intake in obese and reduced obese individuals. Appetite 2003 Feb;40(2):137-143. [doi: 10.1016/s0195-6663(02)00143-5]
32. Nagpal TS, Prapavessis H, Campbell C, Mottola MF. Measuring Adherence to a Nutrition and Exercise Lifestyle Intervention: Is Program Adherence Related to Excessive Gestational Weight Gain? Behav Anal Pract 2017 Dec 17;10(4):347-354 [FREE Full text] [doi: 10.1007/s40617-017-0189-5] [Medline: 29214130]
33. Kim J, Lim S, Min YH, Shin Y, Lee B, Sohn G, et al. Depression Screening Using Daily Mental-Health Ratings from a Smartphone Application for Breast Cancer Patients. J Med Internet Res 2016 Aug 04;18(8):e216 [FREE Full text] [doi: 10.2196/jmir.5598] [Medline: 27492880]
34. Kimokoti R, Newby P, Gona P, Zhu L, Jasuja G, Pencina M, et al. Diet quality, physical activity, smoking status, and weight fluctuation are associated with weight change in women and men. J Nutr 2010 Jul;140(7):1287-1293 [FREE Full text] [doi: 10.3945/jn.109.120808] [Medline: 20484553]
35. Heatherton TF, Polivy J, Herman CP. Restraint, weight loss, and variability of body weight. Journal of Abnormal Psychology 1991;100(1):78-83. [doi: 10.1037/0021-843x.100.1.78]
36. Ma C, Avenell A, Bolland M, Hudson J, Stewart F, Robertson C, et al. Effects of weight loss interventions for adults who are obese on mortality, cardiovascular disease, and cancer: systematic review and meta-analysis. BMJ 2017 Nov 14;359:j4849 [FREE Full text] [doi: 10.1136/bmj.j4849] [Medline: 29138133]
37. Poobalan AS, Aucott LS, Precious E, Crombie IK, Smith WCS. Weight loss interventions in young people (18 to 25 year olds): a systematic review. Obes Rev 2010 Aug;11(8):580-592. [doi: 10.1111/j.1467-789X.2009.00673.x] [Medline: 19874531]
38. Volpp KG, John LK, Troxel AB, Norton L, Fassbender J, Loewenstein G. Financial incentive-based approaches for weight loss: a randomized trial. JAMA 2008 Dec 10;300(22):2631-2637 [FREE Full text] [doi: 10.1001/jama.2008.804] [Medline: 19066383]
39. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
40. RAYNOR HA, KILANOWSKI CK, ESTERLIS I, EPSTEIN LH. A cost-analysis of adopting a healthful diet in a family-based obesity treatment program. Journal of the American Dietetic Association 2002 May;102(5):645-656. [doi: 10.1016/s0002-8223(02)90148-3]
41. Epstein LH, Paluch RA, Beecher MD, Roemmich JN. Increasing healthy eating vs. reducing high energy-dense foods to treat pediatric obesity. Obesity (Silver Spring) 2008 Feb;16(2):318-326 [FREE Full text] [doi: 10.1038/oby.2007.61] [Medline: 18239639]

42. Jin J, Sklar GE, Min Sen Oh V, Chuen Li S. Factors affecting therapeutic compliance: A review from the patient's perspective. Ther Clin Risk Manag 2008 Feb;4(1):269-286 [FREE Full text] [doi: 10.2147/tcrm.s1458] [Medline: 18728716]
43. Atinga RA, Yarney L, Gavu NM. Factors influencing long-term medication non-adherence among diabetes and hypertensive patients in Ghana: A qualitative investigation. PLoS One 2018;13(3):e0193995 [FREE Full text] [doi: 10.1371/journal.pone.0193995] [Medline: 29590156]
44. Gill S, Panda S. A Smartphone App Reveals Erratic Diurnal Eating Patterns in Humans that Can Be Modulated for Health Benefits. Cell Metab 2015 Nov 03;22(5):789-798 [FREE Full text] [doi: 10.1016/j.cmet.2015.09.005] [Medline: 26411343]
45. Freyne J, Yin J, Brindal E, Hendrie GA, Berkovsky S, Noakes M. Push Notifications in Diet Apps: Influencing Engagement Times and Tasks. International Journal of Human–Computer Interaction 2017 Feb 06;33(10):833-845. [doi: 10.1080/10447318.2017.1289725]
46. Verbesselt J, Hyndman R, Newnham G, Culvenor D. Detecting trend and seasonal changes in satellite image time series. Remote Sensing of Environment 2010 Jan;114(1):106-115. [doi: 10.1016/j.rse.2009.08.014]
47. Bergmeir C, Hyndman R, Benítez J. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. International Journal of Forecasting 2016 Apr;32(2):303-312. [doi: 10.1016/j.ijforecast.2015.07.002]
48. Lambert EV, Speechly DP, Dennis SC, Noakes TD. Enhanced endurance in trained cyclists during moderate intensity exercise following 2 weeks adaptation to a high fat diet. Eur J Appl Physiol 1994 Oct;69(4):287-293. [doi: 10.1007/bf00392032]
49. Hall KD, Chen KY, Guo J, Lam YY, Leibel RL, Mayer LE, et al. Energy expenditure and body composition changes after an isocaloric ketogenic diet in overweight and obese men. Am J Clin Nutr 2016 Aug 06;104(2):324-333 [FREE Full text] [doi: 10.3945/ajcn.116.133561] [Medline: 27385608]
50. Bray G, Bouchard C. Handbook of obesity: clinical applications. New York: Marcel Dekker; 2004. URL: http://www.crcnetbase.com/isbn/9780824758622 [accessed 2019-10-03]
51. Burke LE, Swigart V, Warziski Turk M, Derro N, Ewing LJ. Experiences of self-monitoring: successes and struggles during treatment for weight loss. Qual Health Res 2009 Jun;19(6):815-828 [FREE Full text] [doi: 10.1177/1049732309335395] [Medline: 19365099]
52. Kiefer I, Rathmanner T, Kunze M. Eating and dieting differences in men and women. The Journal of Men's Health & Gender 2005 Jun;2(2):194-201. [doi: 10.1016/j.jmhg.2005.04.010]
53. Rolls BJ, Fedoroff IC, Guthrie JF. Gender differences in eating behavior and body weight regulation. Health Psychology 1991;10(2):133-142. [doi: 10.1037//0278-6133.10.2.133] [Medline: 2055211]
54. Biau D, Kernéis S, Porcher R. Statistics in brief: the importance of sample size in the planning and interpretation of medical research. In: Clin Orthop Relat Res. Statistics in brief: the importance of sample size in the planning and interpretation of medical research. Clin Orthop Springer; Sep 2008:2282-2288.
55. James NA, Matteson DS. ecp : An Package for Nonparametric Multiple Change Point Analysis of Multivariate Data. J. Stat. Soft 2014;62(7):1-25. [doi: 10.18637/jss.v062.i07]
56. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. BMC Med Res Methodol 2007 Jul 03;7(1):30 [FREE Full text] [doi: 10.1186/1471-2288-7-30] [Medline: 17608932]

Abbreviations

ANOVA: analysis of variance.

Edited by G Eysenbach; submitted 17.12.19; peer-reviewed by P Pantoja Bustillos, A Leichtle, S Comulada; comments to author 01.03.20; revised version received 02.04.20; accepted 03.06.20; published 11.08.20.

Please cite as:
Kim J, Kam HJ, Kim Y, Lee Y, Lee JH
Understanding Time Series Patterns of Weight and Meal History Reports in Mobile Weight Loss Intervention Programs: Data-Driven Analysis
J Med Internet Res 2020;22(8):e17521
URL: https://www.jmir.org/2020/8/e17521
doi:10.2196/17521
PMID:32780028

Original Paper

Role of Assistive Robots in the Care of Older People: Survey Study Among Medical and Nursing Students

Sylwia Łukasik[1], PhD; Sławomir Tobis[2], PhD; Sylwia Kropińska[3], PhD; Aleksandra Suwalska[4], PhD

[1]Institute of Human Biology and Evolution, Faculty of Biology, Adam Mickiewicz University in Poznań, Poznań, Poland

[2]Occupational Therapy Unit, Chair of Geriatric Medicine and Gerontology, Poznan University of Medical Sciences, Poznań, Poland

[3]Geriatrics Unit, Chair and Department of Palliative Medicine, Poznan University of Medical Sciences, Poznań, Poland

[4]Department of Mental Health, Chair of Psychiatry, Poznan University of Medical Sciences, Poznań, Poland

Corresponding Author:
Sylwia Łukasik, PhD
Institute of Human Biology and Evolution, Faculty of Biology
Adam Mickiewicz University in Poznań
Uniwersytetu Poznańskiego 6
Poznań, 61-614
Poland
Phone: 48 61 829 56 26
Email: lukasik@amu.edu.pl

Abstract

Background: Populations are aging at an alarming rate in many countries around the world. There has been not only a decrease in the number of births and an increase in the percentage of older people, but also an increase in the number of people living alone. There is growing demand for specialist medical care and daily care with the number of people who can act as caregivers reducing. The use of assistive robots can, at least partially, solve these problems.

Objective: The purpose of this study was to examine the opinions of future health care professionals (medical and nursing students) regarding the use of assistive robots in the care of older people.

Methods: The study was conducted with a group of 178 students from Poznan University of Medical Sciences, Poznań, Poland (110 nursing students and 68 medical students), using the Users' Needs, Requirements, and Abilities Questionnaire.

Results: The participants of this study believed that assistive robots should, first of all, remind older people to take medication regularly, ensure their safety, monitor their health status and environment, provide cognitive training, and encourage them to maintain physical activity. In the students' opinion, the robot should not be an older person's companion but only act as an assistant. Nursing students had significantly higher scores than medical students in several statements concerning everyday use of robots, including reminding about meals (P=.03), monitoring the environment (P=.001), providing advice about a healthy diet (P=.04), monitoring the intake of food and fluids (P=.02), and automatic "switch on" function (P=.02). Nursing students were more focused on the social functions of robots, including encouraging contact with friends (P=.003) and reducing the sense of loneliness and improving mood (P=.008). Medical students were more aware of privacy issues in the statement concerning the possibility of switching off the robot in specific situations (P=.01).

Conclusions: Our study revealed a generally positive attitude of future doctors and nurses toward assistive robots, which can have an impact on their acceptance by older adults. In the future, medical professionals could help their patients to choose the right robots (and necessary functions) that are best suited to their needs. However, this would require expanding the curriculum to include the issues of gerontechnology.

(J Med Internet Res 2020;22(8):e18003) doi:10.2196/18003

KEYWORDS

assistive robots; older adults; acceptance; medical students; nursing students

Introduction

In recent years, populations are aging at an alarming rate in many countries around the world. There has been not only a decrease in the number of births and an increase in the percentage of older people, but also an increase in the number of people living alone [1]. Europe is currently one of the oldest regions in the world [2]. According to Eurostat data, in 2014, people aged over 65 years accounted for almost 20% of the population of the European Union [3]. It is estimated that this value will increase to around 30% before 2060 [4]. The aging of the European society is now one of the biggest social and economic challenges faced by the European Union [3].

Since people are living longer and older people are constituting an increasing proportion of the population, there is progressively insufficient availability of specialized caregivers [5]. One possible form of support that has the potential to solve the problems of the aging of European societies, at least partially, is the use of assistive robots in the care of older people. Such robots can make it easier for older people to remain self-efficient and independent for longer while also reducing the burden on the family and formal caregivers [6]. Additionally, robots can not only help older adults in everyday life, but also be used in medical care (eg, for remote monitoring of patient health), which can additionally contribute to reducing costs for public services or care-assurance budgets [7].

Thus far, several models of robots supporting older people, with quite a variety of uses, have been developed. Robots can be used as aids in preparing [8] and consuming meals [9,10], daily toileting [11], doing housework [12], and monitoring the user's state of health [2], among others. In addition, these devices can also provide older users with company (eg, as a chess companion) and encourage them to do cognitive training, as some studies have suggested the positive effects of these devices on cognitive function in older people [13]. Social robot interventions have been reported to improve mood and reduce stress levels in elderly users [14].

Advances in technology have enabled the development of robots that are closely adjusted to the needs of their users. However, this requires a thorough understanding of the needs and expectations of older people regarding these devices [15,16]. While the research conducted so far has concentrated mainly on the acceptance of existing robots [5,17,18], understanding why older people accept or reject assistive robots and what expectations they have of them is essential to not only improve the design of these robots but also develop effective strategies for placing them on the market [7].

While designing a new robot, the opinion of the end user has a central place, followed by the views of their family members and caregivers. As demonstrated by Sorri and Leinonen, the preferences of these two groups can be different [19]. Our previous study of occupational therapy students' perceptions on the use of robots in the care of older adults indicated that students were rather skeptical regarding the abilities of older people to operate robots [20]. Furthermore, we found that older subjects themselves indicated a need for competent pretraining to be able to cope with a social robot [21]. The approach of older people to new technologies, such as assistive robots, depends on many factors, including gender, education level, and previous experience with electronic devices [22,23]. Research to date has revealed that men are more open to new technologies than women [24], and people who have previously used various electronic devices are more likely to get acquainted with emerging technological innovations [19]. Still, it should be kept in mind that the acceptance of robots is a complex process, and these are certainly not the only factors that affect the views and approaches of older people to such devices.

The study by Schwartz et al [25] showed that consumers' decisions regarding the use of services largely depend on the advice given to them by professional experts, such as doctors, financial advisors, and accountants. A series of psychological experiments revealed that people, who have established a long-term relationship with an expert, are quite reluctant to seek additional advice on the services offered to them [26]. The results of this study suggest that the opinions of older people on the use of assistive robots can fundamentally depend on the point of view of health care professionals who take care of them (eg, doctors and nurses). If the attending physicians or formal caregivers of older people (whom they trust and whose knowledge they value) have a positive attitude toward new technologies (eg, assistive robots), it is quite likely that the older people will be more willing to become familiar with such a device and to use it at home. Therefore, it seems important to check what future doctors and nurses think about assistive robots that can support older people.

The aim of this study was to collect the opinions of future health care professionals (medical and nursing students) about the use of robots in the care of older people.

Methods

The ENRICHME Project

The study was conducted as part of the ENRICHME (ENabling Robot and Assisted Living Environment for Independent Care and Health Monitoring of the Elderly) project funded by the European Union under the Horizon 2020 framework (project number: 643691), with consent from the Bioethical Committee of the Poznan University of Medical Sciences (consent number: 389/17). The project evaluated the possibility of supporting patients with mild cognitive impairment using an assistive robot in their home environment [27].

The Studied Group

The study was performed in a group of 178 students from the Poznan University of Medical Sciences, Poznan, Poland (110 nursing students and 68 medical students). The study included only students who completed the second year of study (ie, those who can be reasonably expected to have knowledge about the care of older people). None of the participants indicated previous experience with robotics.

The Users' Needs, Requirements, and Abilities Questionnaire

The research was conducted using the Users' Needs, Requirements, and Abilities Questionnaire (UNRAQ) developed

jointly by ENRICHME project partners [21,28]. The UNRAQ is composed of three main parts. The first of these involves demographic data, with questions about age, gender, education, and computer skills. The second part involves opinions about robots (ie, use of robots, roles of robots, social aspects, and assistant role). For each statement in this part of the questionnaire, the respondent had the opportunity to choose the most appropriate answer on a 5-point Likert scale (1, I totally disagree; 2, I partially disagree; 3, I neither agree nor disagree; 4, I partially agree; 5, I totally disagree). Answers 4 and 5 were treated as positive, answer 3 was treated as neutral, and answers 1 and 2 were treated as negative. The third part of the questionnaire involves the so-called creativity box (ie, a place where respondents can present their ideas about robots and their functions). Prior to the interview, the participants were presented with a photograph of the Kompaï robot (Robosoft).

Statistical Analysis

Data are presented as mean (SD). Statistical calculations were performed using the Statistica 13 software (StatSoft). The Mann-Whitney *U* test was used to assess the statistical significance of differences between the studied groups, where a *P* value <.05 was considered statistically significant.

Results

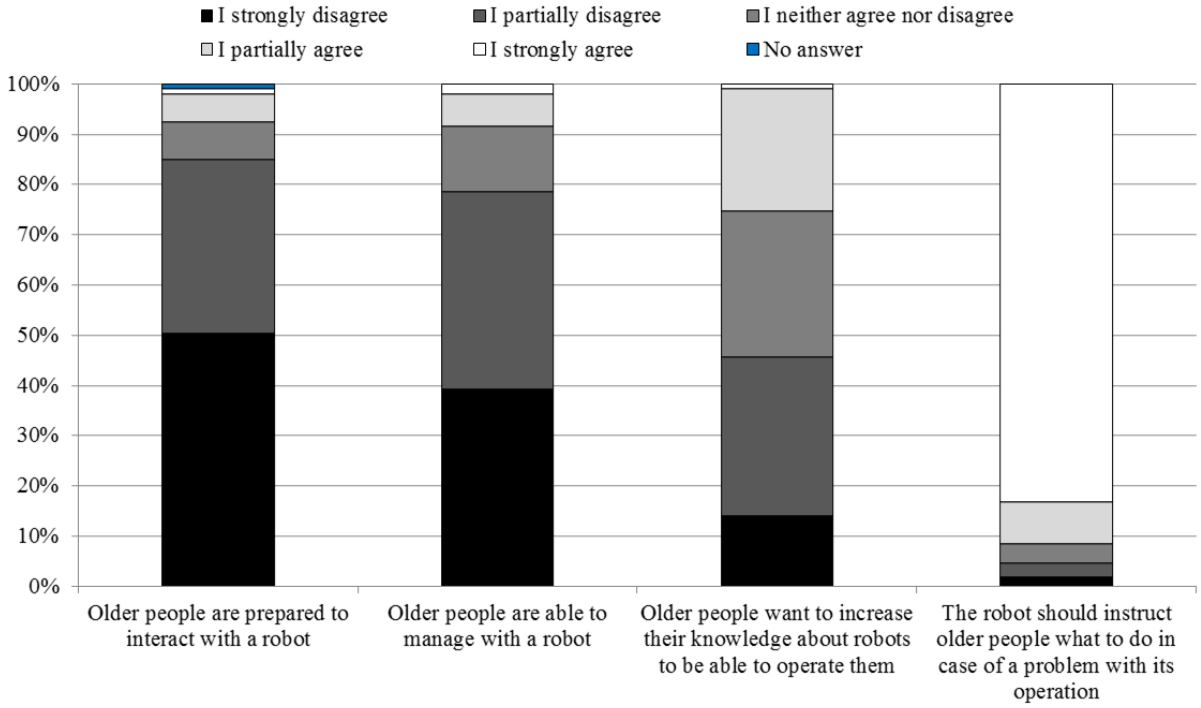
Characteristics of the Studied Groups

The mean age of the participants was 21.6 (SD 2.4) years among nursing students and 23.2 (SD 2.0) years among medical students. All study participants declared that they were able to operate a computer, and the vast majority of them claimed that they were at ease in using electronic devices (148/170, 87.1%; 96 nursing students and 52 medical students). Out of 178 respondents, 23 looked after a family member (15 nursing students and eight medical students).

Opinions of Students About the Readiness of Older People for Operating a Robot

The majority of students (151/176, 85.8%) participating in the study thought that older people are currently not prepared for the use of assistive robots. Among these students, 90 strongly agreed with this statement and 61 partially agreed (Figure 1). Only 10% (7/68) of medical students positively assessed the readiness of older people to use a robot. Notably, this opinion was shared by only one nursing student (1/108, 0.9%). The differences observed between the groups were not relevant. The surveyed students believed that older people would like to broaden their knowledge about robots in order to be able to use them, and 30.9% (21/68) of medical students and 17.4% (19/109) of nursing students agreed with this statement (*P*=.07). Respondents in their statements emphasized that there are no forms of education in this field for older adults in Poland and that not all older people can use electronic devices; therefore, robot training should be conducted. The vast majority of students believed that an assistive robot should give the user instructions on its use (164/177, 92.7%).

Figure 1. Opinions of the whole group of students about the readiness of older people to use a robot.



Students' Opinions About the Role of a Robot

A great majority of nursing and medical students (167/178, 93.8%) considered a robot as a useful device for older people. Additionally, majority of students (152/178, 85.4%) considered it as an assistant for older people. Only about 55% of students (96/175, 54.9%) considered a robot as a companion for older people (Table 1). Respondents unanimously emphasized that a robot will not replace relationships with real people and that for older people, it should be only an assistant and not a companion.

Table 1. Opinions of nursing and medical students about the general role of a robot in the care of older people.

Statement	Total (N=178)		Nursing students (N=110)		Medical students (N=68)		*P* value
	n[a] (%[b])	Mean (SD)	n[a] (%[b])	Mean (SD)	n[a] (%[b])	Mean (SD)	
A robot is a useful device for older people (something to be used when needed, with no other interaction)[c]	167 (93.8)	4.6 (0.8)	105 (95.5)	4.6 (0.7)	62 (91.2)	4.6 (0.9)	.25
A robot is an assistant for older people	152 (85.4)	4.2 (0.9)	92 (83.6)	4.2 (1.0)	60 (88.2)	4.3 (0.9)	.30
A robot is a companion for older people	96 (54.9)	3.5 (1.2)	59 (54.6)	3.5 (1.2)	37 (55.2)	3.4 (1.2)	.53

[a]Number of students who agreed with the statement (strongly or partially).

[b]Percentage of students who agreed with the statement (strongly or partially).

[c]Data computed for 175 students owing to missing data in three questionnaires.

The answers regarding the use of a robot and its specific tasks in the care of older people were grouped according to the degree of their acceptance by students participating in the study (Table 2). Nine perfect match statements (over 90% positive answers) defined the most important use of an assistive robot concerning mainly functions, such as reminding about medications, ensuring security at home (calling for help if needed), monitoring the user's health status and the environment, preventing memory deterioration (eg, through cognitive training), and encouraging the user to maintain physical activity. A very good match (71%-90% positive answers) was obtained for 11 statements related, among others, to monitoring the amount of food and fluid intake, as well as providing entertainment to the older user. A good match (51%-70% positive answers) was observed for seven statements, including encouraging users to contact other people and the ability to customize the robot's functions to the individual needs of the user. However, students emphasized that older people should not be allowed to change the robot's settings themselves, as this may result in the needed functions being turned off.

Table 2. Opinions of nursing and medical students related to statements about the specific functions of a robot in the care of older people, arranged by the percentage of students who agreed with the statement.

Group, Statement	Total (N=178)		Nursing students (N=110)		Medical students (N=68)		*P* value
	n[a] (%[b])	Mean (SD)	n[a] (%[b])	Mean (SD)	n[a] (%[b])	Mean (SD)	
Perfect match							
The robot reminds about medications.	169 (94.9)	4.7 (0.6)	102 (92.7)	4.7 (0.6)	67 (98.5)	4.7 (0.6)	.52
The robot helps to preserve memory function (eg, plays memory games).	167 (94.9)	4.6 (0.7)	105 (96.3)	4.6 (0.6)	62 (92.5)	4.4 (0.9)	.13
The robot encourages and guides the owner to perform physical exercises.	167 (93.8)	4.5 (0.7)	100 (90.9)	4.6 (0.7)	67 (98.5)	4.5 (0.6)	.08
The robot increases the safety of the person: calls for help when needed and monitors health parameters (blood pressure, heart rate, body temperature, respiration rate, etc).	166 (93.3)	4.6 (0.8)	103 (93.6)	4.6 (0.8)	63 (92.6)	4.6 (0.8)	.85
The robot increases the safety of the home (eg, locks doors, detects leaking gas, etc).	165 (92.7)	4.6 (0.8)	101 (91.8)	4.6 (0.8)	64 (94.1)	4.6 (0.7)	.99
The robot reminds about meal times and drink times.	165 (92.7)	4.5 (0.8)	102 (93.6)	4.6 (0.7)	62 (91.2)	4.4 (0.9)	.03
The robot reminds about appointments.	163 (91.6)	4.5 (0.7)	99 (90.0)	4.5 (0.7)	64 (94.1)	4.4 (0.8)	.23
The robot monitors the environment (temperature and humidity) and suggests air conditioning adjustment or opening windows.[c]	162 (91.5)	4.5 (0.8)	102 (93.6)	4.6 (0.8)	60 (88.2)	4.3 (0.9)	.001
The robot should be customizable (adjustable to individual user preferences and needs).[c]	160 (90.4)	4.6 (0.8)	96 (88.1)	4.5 (0.8)	64 (94.1)	4.6 (0.8)	.52
Very good match							
The robot provides advice about a healthy diet.	160 (89.9)	4.5 (0.8)	98 (89.1)	4.5 (0.8)	62 (91.2)	4.4 (0.8)	.04
The robot observes the behavior of the older person to detect falls or changes due to illness.	158 (88.8)	4.5 (0.8)	97 (88.2)	4.5 (0.7)	61 (89.7)	4.5 (0.9)	.96
The robot monitors the amount of food and fluid intake of the owner.	155 (87.1)	4.3 (0.9)	99 (90.0)	4.4 (0.8)	56 (82.4)	4.1 (1.0)	.02
The robot informs a family member or caregiver about the older person's behavior/health problem.	152 (85.4)	4.3 (0.9)	92 (83.6)	4.4 (0.9)	60 (88.2)	4.2 (0.8)	.01
The robot helps the owner to find lost objects (eg, glasses and keys).	150 (84.3)	4.4 (0.8)	91 (83.6)	4.4 (0.8)	58 (85.3)	4.3 (0.8)	.23
The robot has entertainment functions (eg, gaming partner, reading aloud, and playing music).	147 (82.6)	4.2 (0.9)	91 (82.7)	4.2 (0.9)	56 (82.4)	4.2 (0.8)	.69
The robot is able to make the life of older people easier.	146 (82.0)	4.2 (0.8)	88 (80.0)	4.2 (0.8)	58 (85.3)	4.3 (0.8)	.37
The older person has control over the robot.[c]	134 (75.7)	4.1 (0.9)	79 (72.5)	4.1 (0.9)	55 (80.9)	4.2 (0.9)	.30
The robot can automatically reactivate after being switched off.	131 (73.6)	4.0 (1.2)	84 (76.4)	4.1 (1.2)	47 (69.1)	3.8 (1.1)	.02
The older person is able to choose the required functions of the robot and disable other functions.	126 (70.8)	3.9 (1.0)	78 (70.9)	3.9 (1.0)	48 (70.6)	4.0 (1.0)	.31
The robot has much information about the user (social, medical, and others).[d]	126 (71.6)	3.9 (0.9)	75 (69.4)	3.9 (0.9)	51 (75.0)	3.9 (1.0)	.89
Good match							
The robot initiates contact with others (eg, calls friends and initiates skype conversations).	123 (69.1)	4.0 (1.0)	87 (70.9)	4.0 (1.0)	45 (66.2)	3.9 (1.1)	.35

Group, Statement	Total (N=178)		Nursing students (N=110)		Medical students (N=68)		*P* value
	n[a] (%[b])	Mean (SD)	n[a] (%[b])	Mean (SD)	n[a] (%[b])	Mean (SD)	
The robot encourages to enhance contact with friends.	123 (69.1)	4.0 (1.0)	83 (75.5)	4.1 (0.9)	40 (58.8)	3.7 (1.0)	.003
The robot accompanies the owner in everyday activities (eg, watching TV and preparing meals).	123 (69.1)	3.8 (1.1)	76 (69.1)	3.9 (1.2)	47 (69.1)	3.8 (1.0)	.32
The older person is able to send the robot to its place/docking station and keep it staying there.	121 (68.4)	3.9 (1.1)	69 (63.3)	3.8 (1.1)	52 (76.5)	4.0 (1.0)	.10
The older person is able to switch off the robot in specific situations (friends' visits, privacy reasons, etc).[c]	121 (68.0)	3.9 (1.2)	68 (61.8)	3.7 (1.2)	53 (77.9)	4.2 (1.0)	.01
The robot detects the owner's mood (facial expression).	121 (68.0)	3.9 (1.1)	74 (67.3)	3.9 (1.1)	47 (69.1)	3.8 (1.1)	.54
The robot decreases the sense of loneliness and improves the mood of the older person.	118 (66.3)	3.7 (1.2)	76 (69.1)	3.9 (1.1)	42 (61.8)	3.4 (1.2)	.008

[a]Number of students who agreed with the statement (strongly or partially).

[b]Percentage of students who agreed with the statement (strongly or partially).

[c]Data computed for 177 students owing to missing data in one questionnaire.

[d]Data computed for 176 students owing to missing data in two questionnaires.

Nursing students scored significantly higher than medical students in the case of eight statements concerning everyday use of robots (reminding about meals [*P*=.03], monitoring the environment [*P*=.001], providing advice about a healthy diet [*P*=.04], monitoring the intake of food and fluids [*P*=.02], and automatic "switch on" function [*P*=.02]). Nursing students were more focused on the social functions of robots (encouraging to enhance contact with friends and reducing the sense of loneliness and improving mood). On the other hand, medical students were more aware of privacy issues in the statement concerning the possibility of switching off the robot in specific situations.

Discussion

The Attitude of Future Medical Professionals to Assistive Robots

Over the past several years, we have been experiencing the rapid development of technology that penetrates almost all aspects of life (including the care of older adults). Assistive robots can not only help older people remain independent for longer, but also support and facilitate the work of doctors and formal caregivers [29]. Our research concentrated on the approach of future health care professionals to the use of robots in the care of older people and what roles they think such devices should play. In general, the results of our analyses indicate a positive attitude of medical and nursing students to socially assistive robots. The vast majority of participants saw high potential in such devices, similar to the findings in the study by van Kemenade et al [30] for companion robots. These results are also consistent with the results obtained in focus group studies of potential end users, and younger and older caregivers of older people as part of the Domeo project [31], as well as surveys conducted by Faucounau et al [32] and Cylkowska-Nowak et al [21].

However, it should be emphasized that the surveyed students had some reservations about the use of robots by older people. They suggested that older people in Poland might not yet be ready to use such devices owing to difficulties in handling these devices. In addition, the problem may concern not only the operation of the robots but also the selection of the most suitable model or the setting of functions appropriate for a given user. According to the students, older people often do not have sufficient knowledge of the use of electronic devices or their suitability for potential users' needs and requirements. The participants of our study pointed to the necessity to provide specific training to older people on the use of robots. This corresponds with the observation of Flandorfer that the more experienced people are in using new technologies and the smarter the devices, the higher is the desire to use such devices when needed [18]. Johansson-Pajala et al also observed that the attitudes of study participants to care robots improved as their knowledge increased [33], which was interpreted as a general need for an improved orientation within the field.

The Role of Robots According to Future Medical Professionals

Our study revealed that the most important roles of assistive robots relate to functions such as reminding about taking medications, ensuring the safety of older people, preventing deterioration of their memory, and encouraging them to maintain physical activity. Future doctors and nurses were most critical of using a robot as a companion of an older person. Although the results of previous studies indicated that older people can benefit from such interactions [34], students involved in our research believed that robots should never replace contact with other people, but rather should encourage such contact.

The study participants also believed that assistive robots should be personalized (ie, it should be possible to select appropriate functions in relation to the individual needs of a given user,

depending, among others, on the state of health). Notably, surveyed students believed that older people should not be able to change the robot settings themselves. They argued that an older person could disable the functions that are most needed (eg, those for monitoring their health and detecting a threat), thereby possibly exposing themselves to danger.

Differences in Opinions Among Future Medical Staff About Assistive Robots

We noticed slight differences between the studied groups. Nursing students were more open to robots and saw more opportunities for their use in the care of older people compared with medical students. This is reflected in their higher average results regarding positive responses and acceptance of a larger spectrum of robot functions (very good and perfect match). They were also more aware of the problems related to the aging process, such as restricted social life, which can cause negative social effects, as well as undernourishment or drinking too little fluid, which may have health consequences (eg, lead to cognitive decline) [35].

The Worries of Future Medical Staff About Assistive Robots

Although the development of robotics may raise some ethical issues [36], none of the students participating in our study showed any ethical concerns about who could decide to turn off the robot or set it up, similar to the results in other studies [31]. The surveyed students believed that older people should not be able to turn off the robot completely for their safety.

However, it should be kept in mind that, when creating such devices, it is necessary to not only improve the lives of older people by enabling them to live in their homes longer, but also protect their individual rights and their physical and mental well-being [37,38].

Among future medical staff, we did not observe a fear of losing their jobs from the deployment of assistive robots. The students very clearly emphasized in their answers that robots will never replace contact with another person (social issues). These devices should only act as an assistant to formal caregivers and doctors, helping them to care for older people and monitor their health remotely. Importantly, students showed concerns about older caretakers, indicating the possibility of an emotional bond with the device and the negative consequences it could have in the event of damage or failure of the robot.

Doctors and Nurses as Professional Experts

The study by Schwartz et al [25] proved that the opinions and approaches of people to use certain services might be influenced by professional advisers, such as doctors and nurses. This effect is particularly evident in the case of a long-term relationship between the expert and the client, as well as when the client values the expert's knowledge and experience. Therefore, the positive attitude of future medical personnel to assistive robots observed in our study can presumably have an impact on the acceptance of robots by older people. It is also worth noting that doctors and nurses as formal caregivers could additionally facilitate the introduction of robots into the homes of their patients. The research conducted so far proves that it is easier for younger generations to use new technologies with which they have grown and that it is easier for them to make a more accurate choice of the right device [39]. Therefore, medical staff (from younger generations than their patients) could certainly help older people in choosing devices that would best suit their needs. Importantly, as doctors and nurses know the needs of older people, they should actively cooperate with engineers at as early a stage as possible in robot technology design. Such collaboration allows designing better robots, which could consequently maximize the independence of older people while reducing health care costs. The study by Hoenig et al [40] showed that people who needed assistance in their daily activities and who did not use assistive devices needed an average of 4 hours a week of additional professional care compared with people who used such devices. This seems particularly important if we take into account the fact that the world population is aging, and thus, fewer people can act as formal caregivers.

Assistive Robots: Changes in the Training of Medical Staff

Our study findings indicate that future doctors and nurses are profoundly interested in the functioning of older adults and are happy to comment on the possibilities of using assistive technologies. However, it appears that changes in the curriculum of courses aimed at gerontechnology are desirable, and such classes would be particularly important in terms of doctors and nurses acting as experts or advisers for older people and their cooperation with designers on the necessary functions and features of assistive robots. Skiba [41] emphasized that lecturers for nursing students should introduce gerontechnology to the curriculum to reflect important trends that occur in the society (technology development), and this is especially so, as older people are willing to learn how to use new technologies to be able to stay at home longer, which is extremely important for them [6,42]. Following the suggestion of van Kemenade et al, there should be a place to address the ethical concerns related to the use of robots in care [30]. Curricula at medical universities should include basic engineering concepts as well, which will allow doctors and nurses to evaluate and select the right device (thus helping clients) and will facilitate the solution of possible handling problems arising during its use [43]. As the implementation of socially assistive robots in care requires a partnership among academic institutions, clinicians, and the industry [30], interprofessional collaboration involving highly skilled nurses and medical doctors is indispensable.

Limitations of Our Study and Future Research Directions

We are aware that the relatively small number of participating students is a limitation of our research. In addition, the students did not have the opportunity to observe the use of an assistive robot in the care of older people in practice, which could have affected the results of the study. Broadbent et al [44] found that attitudes toward a robot improve after interacting with it. This suggests that conducting such research after giving the participants an opportunity to use an assistive robot in practice could result in even higher acceptance by future doctors and nurses. In future studies, it would be worth investigating to what

extent such interaction with a robot increases its acceptance. Additionally, differences in the attitudes of students from different countries and regions (eg, Japan vs Europe) are worth exploring [45]. Moreover, the inclusion of potential end users from older generations as well as other stakeholders (family members, professional caregivers, etc) might yield noteworthy results, as indicated in a range of previous studies [20,28,31,33].

Conclusions

Our research revealed a generally positive attitude of future doctors and nurses toward assistive robots. According to the participants of this study, an assistive robot should primarily remind older people to take medications regularly, ensure their safety, monitor their health status and environment, counteract the deterioration of their memory, and encourage them to maintain physical activity. In the opinion of the students, such a robot should not be an older person's companion but should rather act as an assistant. The positive attitude of future medical personnel (professional experts) toward assistive robots, as demonstrated by us, can have a relevant impact on the acceptance of such devices by older adults. In the future, doctors and formal caregivers could help their patients choose the right robots (and the necessary functions) that are best suited to their needs. However, this would require expanding the scope of university education to include questions on gerontechnology.

Acknowledgments

The study was partly funded by the European Union H2020 Program (grant agreement no.: 643691C; project ENRICHME [ENabling Robot and assisted living environment for Independent Care and Health Monitoring of the Elderly]).

Conflicts of Interest

None declared.

References

1. Yamazaki K, Ueda R, Nozawa S, Kojima M, Okada K, Matsumoto K, et al. Home-Assistant Robot for an Aging Society. Proc. IEEE 2012 Aug;100(8):2429-2441. [doi: 10.1109/jproc.2012.2200563]
2. Plaza I, Martín L, Martin S, Medrano C. Mobile applications in an aging society: Status and trends. Journal of Systems and Software 2011 Nov;84(11):1977-1988. [doi: 10.1016/j.jss.2011.05.035]
3. Eurostat. Eurostat Regional Yearbook. Luxembourg: Publications Office of the European Union; 2016:1-264.
4. Creighton H. Europe's Ageing Demography. ILC-UK 2014 EU Factpack. London: International Longevity Centre; 2014:1-20.
5. Heerink M, Kröse B, Evers V, Wielinga B. Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. Int J of Soc Robotics 2010 Sep 4;2(4):361-375. [doi: 10.1007/s12369-010-0068-5]
6. Mitzner TL, Boron JB, Fausset CB, Adams AE, Charness N, Czaja SJ, et al. Older Adults Talk Technology: Technology Usage and Attitudes. Comput Human Behav 2010 Nov 01;26(6):1710-1721 [FREE Full text] [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]
7. Wu Y, Wrobel J, Cornuet M, Kerhervé H, Damnée S, Rigaud A. Acceptance of an assistive robot in older adults: a mixed-method study of human-robot interaction over a 1-month period in the Living Lab setting. Clin Interv Aging 2014 May 8;9:801-811 [FREE Full text] [doi: 10.2147/CIA.S56435] [Medline: 24855349]
8. Bovbel P, Nejat G. Casper: An Assistive Kitchen Robot to Promote Aging in Place. Journal of Medical Devices 2014;8(3):030945 [FREE Full text] [doi: 10.1115/1.4027113]
9. Soyama R, Ishii S, Fukase A. Selectable operating interfaces of the meal-assistance device My Spoon. In: Bien Z, Stefanov D, editors. Advances in Rehabilitation Robotics. Berlin/Heidelberg, Germany: Springer; 2004:155-163.
10. Graf B, Hans M, Schraft RD. Care-O-bot II—Development of a Next Generation Robotic Home Assistant. Autonomous Robots 2004 Mar;16(2):193-205. [doi: 10.1023/b:auro.0000016865.35796.e9]
11. Brooke J. Japan seeks robotic help in caring for the aged. Caring 2004 Jul;23(7):56, 59. [Medline: 15341304]
12. Dario P, Guglielmelli E, Laschi C, Teti G. MOVAID: a personal robot in everyday life of disabled and elderly people. TAD 1999 Jun 01;10(2):77-93. [doi: 10.3233/tad-1999-10202]
13. Esposito R, Fiorini L, Limosani R, Bonaccorsi M, Manzi A, Cavallo F. Supporting Active and Healthy Aging with Advanced Robotics Integrated in Smart Environment. In: Morsi YS, Shukl A, Rathore CP, editors. Optimizing Assistive Technologies for Aging Populations. Hershey: IGI Global; 2016:46-77.
14. Scoglio AA, Reilly ED, Gorman JA, Drebing CE. Use of Social Robots in Mental Health and Well-Being Research: Systematic Review. J Med Internet Res 2019 Jul 24;21(7):e13322 [FREE Full text] [doi: 10.2196/13322] [Medline: 31342908]
15. Łukasik S, Tobis S, Wieczorowska-Tobis K, Suwalska A. Could Robots Help Older People with Age-Related Nutritional Problems? Opinions of Potential Users. Int J Environ Res Public Health 2018 Nov 12;15(11):2535 [FREE Full text] [doi: 10.3390/ijerph15112535] [Medline: 30424561]
16. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013 Nov;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]

17. Broadbent E, Stafford R, MacDonald B. Acceptance of Healthcare Robots for the Older Population: Review and Future Directions. Int J of Soc Robotics 2009 Oct 3;1(4):319-330. [doi: 10.1007/s12369-009-0030-6]
18. Flandorfer P. Population Ageing and Socially Assistive Robots for Elderly Persons: The Importance of Sociodemographic Factors for User Acceptance. International Journal of Population Research 2012 Apr 22;2012:1-13. [doi: 10.1155/2012/829835]
19. Sorri L, Leinonen E. Technology that Persuades the Elderly. In: Oinas-Kukkonen H, Hasle P, Harjumaa M, Segerståhl K, Øhrstrøm P, editors. Persuasive Technology. PERSUASIVE 2008. Lecture Notes in Computer Science, vol 5033. Berlin, Heidelberg: Springer; 2008:270-273.
20. Tobis S, Cylkowska-Nowak M, Wieczorowska-Tobis K, Pawlaczyk M, Suwalska A. Occupational Therapy Students' Perceptions of the Role of Robots in the Care for Older People Living in the Community. Occup Ther Int 2017 Feb 7;2017:9592405 [FREE Full text] [doi: 10.1155/2017/9592405] [Medline: 29097983]
21. Cylkowska-Nowak MT, Salatino C, Tapus A, Suwalska A. The robots in elderly care. 2015 Presented at: 2nd International Multidisciplinary Scientific Conferences on Social Sciences and Arts - SGEM; August 26-September 01, 2015; Albena, Bulgaria URL: https://sgemsocial.org/ssgemlib/spip.php?article1347&lang=en [doi: 10.5593/sgemsocial2015/b11/s2.130]
22. Heerink M. Exploring the influence of age, gender, education and computer experience on robot acceptance by older adults. 2011 Presented at: 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI); March 8-11, 2011; Lausanne, Switzerland URL: https://ieeexplore.ieee.org/document/6281268 [doi: 10.1145/1957656.1957704]
23. Abolhassani N, Santos-Eggimann B, Chiolero A, Santschi V, Henchoz Y. Readiness to accept health information and communication technologies: A population-based survey of community-dwelling older adults. Int J Med Inform 2019 Oct;130:103950. [doi: 10.1016/j.ijmedinf.2019.08.010] [Medline: 31446357]
24. Scopelliti M, Giuliani MV, Fornara F. Robots in a domestic setting: a psychological approach. Univ Access Inf Soc 2005 Jul 19;4(2):146-155. [doi: 10.1007/s10209-005-0118-1]
25. Schwartz J, Luce MF, Ariely D. Are Consumers Too Trusting? The Effects of Relationships with Expert Advisers. Journal of Marketing Research 2019 Mar 06;48(SPL):S163-S174. [doi: 10.1509/jmkr.48.spl.s163]
26. Ariely D. The (honest) truth about dishonesty: how we lie to everyone - especially ourselves. New York: Harper Collins Publishers; 2012.
27. Salatino C, Gower V, Ghrissi M, Tapus A, Wieczorowska-Tobis K, Suwalska A. The EnrichMe project: A robotic solution for independence and active aging of elderly people with MCI. 2016 Presented at: 15th International Conference, ICCHP 2016; July 13-15, 2016; Linz, Austria.
28. Tobis S, Salatino C, Tapus A, Suwalska A, Wieczorowska-Tobis K. Opinions about robots in care for older people. 2017 Presented at: 4th International Multidisciplinary Scientific Conference on Social Sciences and Arts SGEM; August 24-30, 2017; Albena, Bulgaria URL: https://www.sgemsocial.org/index.php/elibrary?view=publication&task=show&id=2768 [doi: 10.5593/sgemsocial2017/33/s12.056]
29. Gross H, Schroeter C, Mueller S, Volkhardt M, Einhorn E, Bley A, et al. Further progress towards a home robot companion for people with mild cognitive impairment. 2012 Presented at: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC); October 14-17, 2012; Seoul, South Korea. [doi: 10.1109/icsmc.2012.6377798]
30. van Kemenade M, Hoorn J, Konijn E. Healthcare Students' Ethical Considerations of Care Robots in The Netherlands. Applied Sciences 2018 Sep 20;8(10):1712. [doi: 10.3390/app8101712]
31. Zsiga K, Edelmayer G, Rumeau P, Péter O, Tóth A, Fazekas G. Home care robot for socially supporting the elderly: focus group studies in three European countries to screen user attitudes and requirements. Int J Rehabil Res 2013 Dec;36(4):375-378. [doi: 10.1097/MRR.0b013e3283643d26] [Medline: 24189106]
32. Faucounau V, Wu Y, Boulay M, Maestrutti M, Rigaud A, QuoVADis project. Caregivers' requirements for in-home robotic agent for supporting community-living elderly subjects with cognitive impairment. Technol Health Care 2009;17(1):33-40. [doi: 10.3233/THC-2009-0537] [Medline: 19478403]
33. Johansson-Pajala R, Thommes K, Hoppe J, Tuisku O, Hennala L, Pekkarinen S. Improved Knowledge Changes the Mindset: Older Adults? Perceptions of Care Robots. In: Zhou J, Salvendy G, editors. Human Aspects of IT for the Aged Population. Design for the Elderly and Technology Acceptance. HCII 2019. Lecture Notes in Computer Science, vol 11592. Cham: Springer; 2019:212-227.
34. Beer J, Takayama L. Mobile remote presence systems for older adults: Acceptance, Benefits, and Concerns. 2011 Presented at: 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI); March 8-11, 2011; Lausanne, Switzerland URL: https://ieeexplore.ieee.org/document/6281295 [doi: 10.1145/1957656.1957665]
35. Wilson MG, Morley JE. Impaired cognitive function and mental performance in mild dehydration. Eur J Clin Nutr 2003 Dec 18;57 Suppl 2:S24-S29. [doi: 10.1038/sj.ejcn.1601898] [Medline: 14681710]
36. Robinson NL, Cottier TV, Kavanagh DJ. Psychosocial Health Interventions by Social Robots: Systematic Review of Randomized Controlled Trials. J Med Internet Res 2019 May 10;21(5):e13203 [FREE Full text] [doi: 10.2196/13203] [Medline: 31094357]
37. Sharkey A, Sharkey N. Granny and the robots: ethical issues in robot care for the elderly. Ethics Inf Technol 2010 Jul 3;14(1):27-40. [doi: 10.1007/s10676-010-9234-6]

38. Kim TK, Choi M. Older adults' willingness to share their personal and health information when adopting healthcare technology and services. Int J Med Inform 2019 Jun;126:86-94. [doi: 10.1016/j.ijmedinf.2019.03.010] [Medline: 31029268]
39. Vaportzis E, Clausen MG, Gow AJ. Older Adults Perceptions of Technology and Barriers to Interacting with Tablet Computers: A Focus Group Study. Front Psychol 2017 Oct 04;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
40. Hoenig H, Taylor DH, Sloan FA. Does assistive technology substitute for personal assistance among the disabled elderly? Am J Public Health 2003 Feb;93(2):330-337. [doi: 10.2105/ajph.93.2.330] [Medline: 12554595]
41. Skiba DJ. The connected age: big data & data visualization. Nurs Educ Perspect 2014;35(4):267-268. [Medline: 25158424]
42. Mattimore TJ, Wenger NS, Desbiens NA, Teno JM, Hamel MB, Liu H, et al. Surrogate and physician understanding of patients' preferences for living permanently in a nursing home. J Am Geriatr Soc 1997 Jul;45(7):818-824. [doi: 10.1111/j.1532-5415.1997.tb01508.x] [Medline: 9215332]
43. Sharts-Hopko NC. The coming revolution in personal care robotics: what does it mean for nurses? Nurs Adm Q 2014 ;38(1):5-12. [doi: 10.1097/NAQ.0000000000000000] [Medline: 24317027]
44. Broadbent E, Kuo IH, Lee YI, Rabindran J, Kerse N, Stafford R, et al. Attitudes and reactions to a healthcare robot. Telemed J E Health 2010 Jun;16(5):608-613. [doi: 10.1089/tmj.2009.0171] [Medline: 20575729]
45. Nomura T, Suzuki T, Kanda T, Han J, Shin N, Burke J, et al. What People Assume about Humanoid and Animal-Type Robots: Cross-Cultural Analysis between Japan, Korea, and the United States. Int J Human Robot 2008;05(01):25-46. [doi: 10.1142/s0219843608001297]

Abbreviations

ENRICHME: ENabling Robot and Assisted Living Environment for Independent Care and Health Monitoring of the Elderly
UNRAQ: Users' Needs, Requirements, and Abilities Questionnaire

Edited by G Eysenbach; submitted 28.01.20; peer-reviewed by KH Tan, A van Maris; comments to author 07.04.20; revised version received 21.04.20; accepted 14.05.20; published 12.08.20.

Please cite as:
Łukasik S, Tobis S, Kropińska S, Suwalska A
Role of Assistive Robots in the Care of Older People: Survey Study Among Medical and Nursing Students
J Med Internet Res 2020;22(8):e18003
URL: https://www.jmir.org/2020/8/e18003
doi:10.2196/18003
PMID:32784187

Original Paper

Developing a Mobile App for Monitoring Medical Record Changes Using Blockchain: Development and Usability Study

MinDong Sung[1], MD; SungJun Park[1], BS; Sungjae Jung[2], MS; Eunsol Lee[2], MD; Jaehoon Lee[3,4], PhD; Yu Rang Park[1], PhD

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Medibloc Inc, Seoul, Republic of Korea

[3]Intermountain Healthcare, Salt Lake City, UT, United States

[4]Department of Biomedical informatics, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical Systems Informatics
Yonsei University College of Medicine
50-1 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 22228 2493
Email: yurangpark@yuhs.ac

Abstract

Background: Although we are living in an era of transparency, medical documents are often still difficult to access. Blockchain technology allows records to be both immutable and transparent.

Objective: Using blockchain technology, the aim of this study was to develop a medical document monitoring system that informs patients of changes to their medical documents. We then examined whether patients can effectively verify the monitoring of their primary care clinical medical records in a system based on blockchain technology.

Methods: We enrolled participants who visited two primary care clinics in Korea. Three substudies were performed: (1) a survey of the recognition of blockchain medical records changes and the digital literacy of participants; (2) an observational study on participants using the blockchain-based mobile alert app; and (3) a usability survey study. The participants' medical documents were profiled with HL7 Fast Healthcare Interoperability Resources, hashed, and transacted to the blockchain. The app checked the changes in the documents by querying the blockchain.

Results: A total of 70 participants were enrolled in this study. Considering their recognition of changes to their medical records, participants tended to not allow these changes. Participants also generally expressed a desire for a medical record monitoring system. Concerning digital literacy, most questions were answered with "good," indicating fair digital literacy. In the second survey, only 44 participants—those who logged into the app more than once and used the app for more than 28 days—were included in the analysis to determine whether they exhibited usage patterns. The app was accessed a mean of 5.1 (SD 2.6) times for 33.6 (SD 10.0) days. The mean System Usability Scale score was 63.21 (SD 25.06), which indicated satisfactory usability.

Conclusions: Patients showed great interest in a blockchain-based system to monitor changes in their medical records. The blockchain system is useful for informing patients of changes in their records via the app without uploading the medical record itself to the network. This ensures the transparency of medical records as well as patient empowerment.

(J Med Internet Res 2020;22(8):e19657) doi:10.2196/19657

KEYWORDS

blockchain; monitoring app; clinical documents

Introduction

We currently live in an era of data management and often pursue goals of open access and transparency, which means that anyone can usually find information promptly based on their specific needs. One exception to such transparency is medical data [1]; although medical records entered by clinicians and stored in clinical information systems legally belong to patients [2,3], many patients realistically find it difficult to gain full and transparent access to their own medical records.

Patient empowerment has long been emphasized and was recently highlighted in the "2020-2025 Federal Health IT Strategic Plan" from the Office of the National Coordinator for Health Information Technology (ONC). To improve patient empowerment, electronic health records (EHRs) should be shared with patients. However, there are several potential threats to EHRs that could undermine trust in data on these systems. First, records can be altered or lost, either accidentally or intentionally, such as through hacking. Even though redundancy exists in database systems, these redundancies are often obscure to outside observers. Second, data can be fabricated or manipulated by medical staff intent on committing fraud. A possible solution to overcoming these dilemmas is blockchain technology, which uses distributed and cryptographically secure ledgers to ensure immutability, transparency, and decentralization. Bitcoin is a well-known example of blockchain in the field of cryptocurrency [4]. Blockchain also provides logs of when data are created, changed, or deleted. Thus, providing all data logs can overcome the two primary threats to EHRs.

Some previous studies have reported the implementation of blockchain to health care [5]. Most of these approaches focus on storing and sharing institutional medical data between EHRs [6-9] and personal health records (PHRs) [10-12]. In addition, blockchain has been implemented for sharing and storing clinical trial data [13,14]. Most of these existing works proposed a well-organized architecture or frameworks and a few demonstrated the performance of the prototype developed during the study. However, no study has yet revealed the actual benefits of developing an EHR system using blockchain. We could consider that such a system using blockchain, which features characteristics of transparency and immutability, would be transparent and immutable.

Therefore, in this study, we used blockchain technology to develop a medical document monitoring system that notifies patients of changes in their medical records. The system was then tested with simulation data and the proof-of-concept study was performed in primary care clinics in Korea.

Methods

Study Design

This is a proof-of-concept study consisting of three substudies: (1) a survey of the recognition of blockchain medical records changes and the digital literacy of participants; (2) an observational study on participants using the blockchain-based mobile alert app; and (3) a usability survey study. The study was approved by the Institutional Review Board of Yonsei University Health System (Y-2019-0127) and all participants provided informed consent.

Before proceeding with the design and development of the mobile app, since the EHR systems used in each hospital setting differ, the documents used in the EHR systems were profiled using HL7 Fast Healthcare Interoperability Resources (FHIR; see Multimedia Appendix 1) [15,16]. Although it is reasonable that patients are notified of any changes in medical records, patients might be overwhelmed by too many notifications whenever any change occurs. Moreover, to improve the usability of the mobile app, changes should be summarized. Therefore, medical documents were divided into three types according to their importance and impact, and the mobile app was designed so that patients are notified of only high-risk changes in documents according to the following three risk levels: risk 1, medical information and other critical items that should not be changed; risk 2, medical information and other items that are allowed to be changed; and risk 3, nonmedical information. These risk levels were also considered in the profiles created.

The mobile app used in this study leverages the blockchain network MediBloc Panacea [17] that was developed based on the Tendermint blockchain [18]. The blockchain uses the delegated proof-of-stake method implemented by the practical Byzantine fault tolerance algorithm to create blocks. In this system, "delegated" refers to delegated nodes that perform and validate transactions and blocks. Validators are selected through voting. Normally, one block is generated per second, and all transaction history transmitted over the network is stored in the generated block. Further, similar to other blockchains, once created, blocks cannot be reversed. We used four different properties in the blockchain transaction: (1) writer of the transaction, which can only be specific clinics; (2) topic that is assigned per individual patient; (3) key; and (4) value. Key and value have the following five attributes: (1) hash value for medical documents, (2) document URL for the FHIR profile in which the document was transformed, (3) hash value before the medical document was changed, (4) risk and number of documents with changes, and (5) date to represent when the document was created. As soon as the document is created, the health information system spontaneously hashes and transacts the metadata of the document to the blockchain network (Figure 1). The mobile app provides users with logs of changes in medical records. The app has been available on the Google Play Store [19] since October 23, 2019.

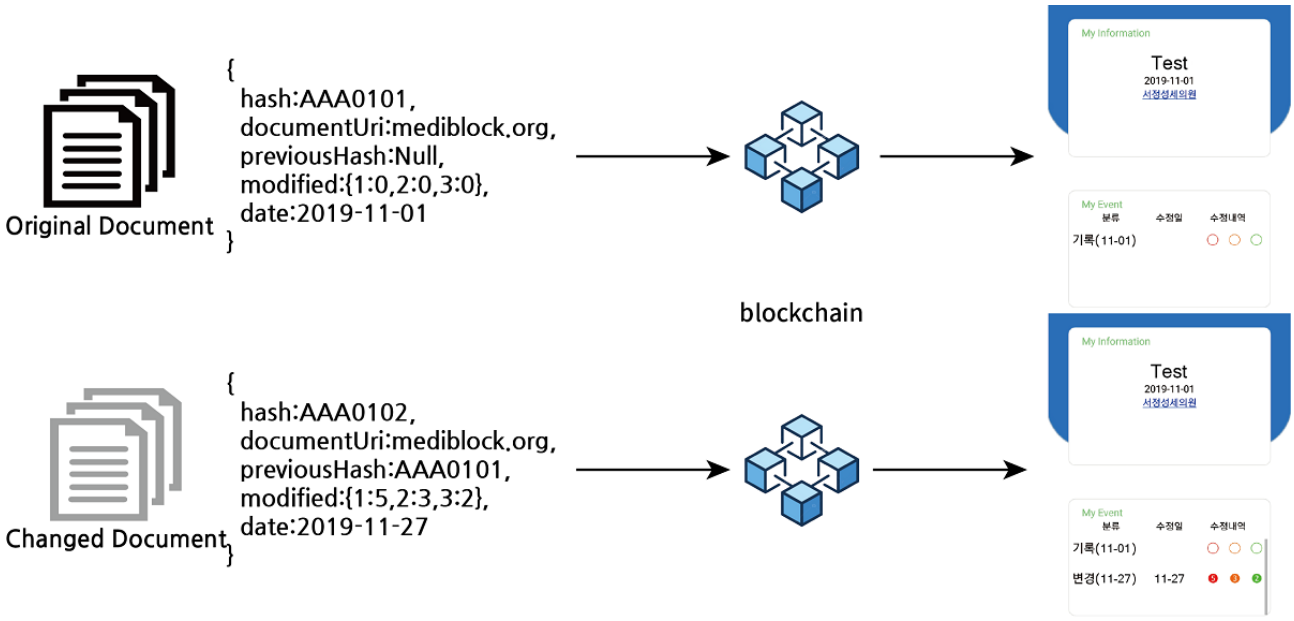
Before the proof-of-concept study was deployed, we simulated the app to evaluate how it captures fake medical records caused by fraudulent actions. We created five fake medical documents for the purpose of simulation. All except for one dataset were assumed to have changes in the documents. The risk levels of the changes were set differently for each dataset. We transacted the datasets and checked how the mobile app worked.

The study was conducted at two primary care clinics specializing in pediatrics in Korea. Only outpatients of these two clinics were enrolled in the study. Anyone who visited the selected clinics was eligible to participate in the study. Because young patients were not interested in their own medical records,

whereas their guardians were more interested, the guardians of patients enrolled in the study on their behalf.

When patients launch the mobile app from their devices, the user is authenticated according to a username and password. After login, the patient can see their own profile and events, which include when the records were created as well as which and how many items at high risk were changed. This information was obtained by querying blockchain using Owner and Topic values that were stored in the backend server of the mobile app.

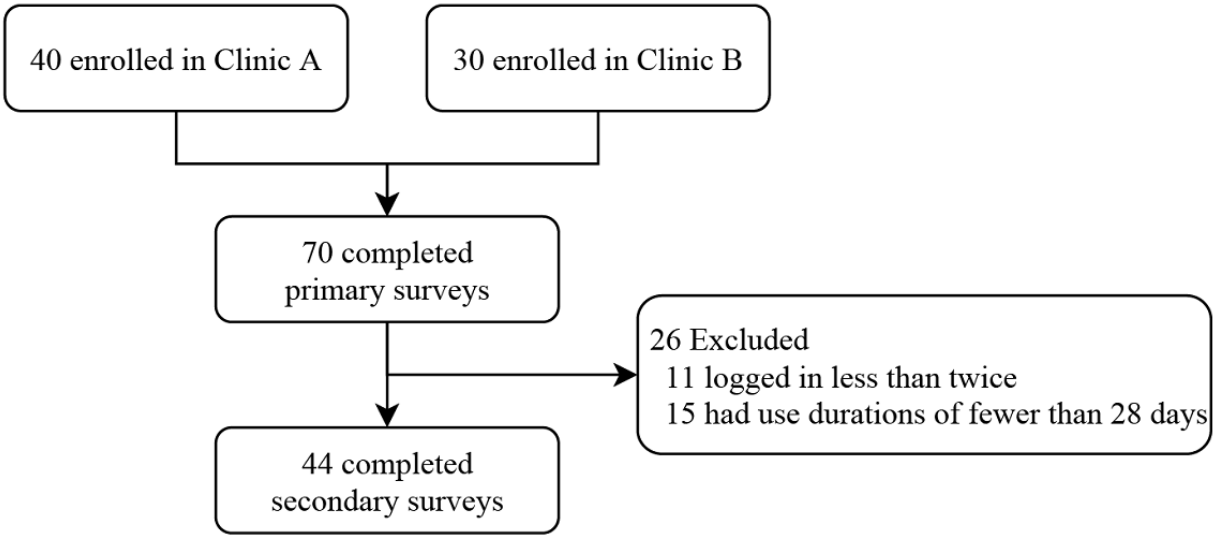
Figure 1. Structure of the medical document monitoring system.



Substudies

To investigate participants' level of recognition of blockchain, medical records changes, and their digital literacy, questionnaires were completed by all participants (Figure 2). The primary survey consisted of items regarding the recognition of blockchain, thoughts and experiences of changes in medical records, and digital literacy. There were six questions that measured blockchain recognition; these were based on virtual currency questions from a 2014 Survey of Consumer Payment Choice [20]. The second part of the questionnaires consisted of four questions. The first two questions were related to the recognition of medical record changes and the need for a monitoring system for medical record changes. These two questions were answered on a 5-point response scale. The other two questions addressed the participants' experience of medical record changes. The third part, digital literacy, consisted of 10 questions answered on a 5-point response scale. These questions were taken from other digital literacy questions [21], modified for the mobile app, and condensed to 10 questions. A higher score represents a higher level of digital health literacy, except in the case of two questions. The questionnaire ended with demographic characteristics (age, sex, and occupation). All 5-point questions were answered, with responses ranging from very negative to very positive (ranging between 1 and 5).

Figure 2. Flow diagram of study participants.



Usage patterns of the app were measured by the number of logins, duration that indicated the difference in days between the first and the last logins, and an event log, which indicated how many items were changed and how important they were. These usage patterns were observed in participants who logged in more than once and for a duration of more than 28 days.

After 28 days of using the app, participants were invited via the app to take part in the secondary survey, which included a System Usability Scale (SUS) survey. For the evaluation of learnability and usability, we used the modified SUS, a reliable, low-cost usability scale that can be used for global assessments of systems usability (see Multimedia Appendix 2) [22,23]. Bangor et al [23] described the adjective ratings associated with SUS scores as follows: worst possible (mean SUS score 25), poor (39.17), satisfactory (52.01), good (72.75), excellent (85.58), and best possible (100).

Statistical Analysis

Categorical variables are presented as numbers and percentages and were compared using the Chi-square test and the Fisher exact test. Continuous variables are expressed as mean (SD) and were compared using the Student unpaired *t*-test, analysis of variance, Wilcoxon signed-rank test, or Mann-Whitney *U* test as appropriate. Surveys answered by a Likert-type scale and scores are expressed as the mode as well as numbers and percentages. All statistical analyses were performed using R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) and 2-tailed tests. Results with $P<.05$ were considered to be statistically significant.

Results

Before the proof-of-concept study was performed, metadata from the 5 original documents and the 4 changed documents were transacted into the blockchain. The simulation metadata and transaction metadata are described in Multimedia Appendix 3. App screenshots are provided in Multimedia Appendix 4.

A total of 70 patients were enrolled in this study. Patient characteristics are shown in Table 1. There were more female patients, and the majority of patients were in their thirties. Thirty-two participants were guardians of the patients, and the occupation of nearly half of the participants was housework or parenting (Table 1).

Table 1. Demographic characteristics of participants.

Characteristic	Hospital A (n=40)	Hospital B (n=30)	Total (N=70)	*P* value
Sex, n (%)				.29
Male	9 (23)	3 (10)	12 (17)	
Female	31 (77)	27 (90)	58 (83)	
Respondent, n (%)				.47
Guardian	19 (47)	18 (60)	37 (53)	
Patient	21 (53)	12 (40)	33 (47)	
Age groups (years), n (%)				.40
19-20	0 (0)	1 (3)	1 (1)	
20-29	5 (12)	1 (3)	6 (9)	
30-39	22 (55)	17 (57)	39 (56)	
40-49	12 (30)	10 (33)	22 (31)	
50-59	1 (3)	0 (0)	1 (1)	
>60	0 (0)	1 (3)	1 (1)	
Occupation, n (%)				.24
Information technology	1 (3)	0 (0)	1 (1)	
Office job	5 (12)	5 (17)	10 (14)	
Management	2 (5)	0 (0)	2 (3)	
Professional	4 (10)	3 (10)	7 (10)	
Housework/parenting	13 (32)	17 (57)	30 (43)	
Sales	2 (5)	0 (0)	2 (3)	
Student	0 (0)	1 (3)	1 (1)	
Unemployed	2 (5)	1 (3)	3 (4)	
Other	11 (28)	3 (10)	14 (20)	

Recognition of blockchain concepts differed depending on the type of question (ie, whether it refers to bitcoin or blockchain). The majority of participants stated that they were aware of bitcoin, whereas less than half were aware of blockchain.

Similarly, respondents were more familiar with bitcoin than with blockchain. However, respondents had less trust in bitcoin than in blockchain. In terms of the recognition of medical record changes, participants tended to not allow changes in their medical records. Subsequently, participants stated a need for a medical record monitoring system. There was only one participant who reported having experienced medical document changes in the first questionnaire. The medical records were changed to correct the wrong information entered previously. In terms of digital literacy, most of the questions were answered as "good" digital literacy (Table 2).

Table 2. Distribution, n (%), of primary survey responses and modes of Likert scale scores (N=70).

Survey question	1	2	3	4	5	Mode
Awareness of blockchain						
Have you heard of bitcoin?	63 (90)	7 (10)	0 (0)	0 (0)	0 (0)	1
How familiar are you with bitcoin and how it works?	17 (24)	21 (30)	24 (34)	6 (8.6)	2 (2.9)	3
How much do you trust bitcoin?	10 (14)	29 (41)	26 (37)	3 (4.3)	2 (2.9)	2
Have you heard of blockchain?	32 (46)	38 (54)	0 (0)	0 (0)	0 (0)	2
How familiar are you with blockchain and how it works?	37 (53)	14 (20)	10 (14)	7 (10.0)	2 (2.9)	1
How much do you trust blockchain?	19 (27)	11 (16)	29 (41)	7 (10.0)	4 (5.7)	3
Recognition of medical document changes						
Do you think changing medical records should be allowed?	42 (60)	7 (10)	11 (16)	9 (13)	1 (1)	1
Do you think we need a medical records falsification monitoring system?	1 (1)	2 (3)	6 (9)	18 (26)	43 (61)	5
Digital literacy[a]						
Can you use the internet?	1 (1)	5 (7)	18 (26)	29 (41)	17 (24)	4
Can you use digital technology?	2 (3)	8 (11)	29 (41)	20 (29)	11 (16)	3
Can you use the app well on your phone?	1 (1)	3 (4)	21 (30)	22 (31)	23 (33)	5
Can you use the camera well on your phone?	1 (1)	1 (1)	6 (9)	15 (21)	47 (67)	5
Can you download and install apps from your phone?	1 (1)	1 (1)	9 (13)	16 (23)	43 (61)	5
I feel comfortable using digital technology	0 (0)	3 (4)	10 (14)	30 (43)	27 (39)	4
I am active in learning digital technology	3 (4)	13 (19)	21 (30)	24 (34)	9 (13)	4
I feel threatened when others talk about digital technology[b]	18 (26)	30 (43)	16 (23)	4 (6)	2 (3)	2
I feel behind other people my age in terms of digital technology[b]	6 (9)	23 (33)	32 (46)	7 (10)	2 (3)	3
I believe that it is important for me to learn how to use digital technology	0 (0)	3 (4)	17 (24)	20 (29)	30 (43)	5

[a]A high score represents a high level of digital literacy, unless otherwise indicated.

[b]A low score represents a high level of digital health literacy.

Only 44 participants were included in the analysis of usage patterns and the secondary survey. During the study period, the app had been accessed a mean of approximately 5 times. The duration of app use, which was indicated by the difference

between the first login and the last login, was approximately 34 days. However, there were no medical document changes during this period. The mean SUS score indicated "satisfactory" usability. The number of logins and the duration were significantly higher in hospital A than in hospital B. Moreover, the duration was significantly different according to occupations. There were no document changes during the study period (Table 3).

Table 3. Usage patterns and System Usability Scale (SUS) scores of app users.

Group	Participants, n (%)	Number of logins, mean (SD)	Duration (days), mean (SD)	SUS score, mean (SD)
Total	44 (100)	5.10 (2.60)	33.60 (10.00)	64.60 (16.00)
Hospital				
Hospital A	24 (55)	5.88 (2.71)	34.63 (11.14)	67.50 (16.32)
Hospital B	20 (45)	4.10 (2.07)	32.34 (8.55)	61.12 (15.36)
P value[a]		.02	.05	.16
Gender				
Male	7 (16)	4.57 (2.94)	31.12 (3.09)	68.21 (19.08)
Female	37 (84)	5.16 (2.53)	34.06 (10.8)	63.92 (15.6)
P value		.45	.98	.54
Age group (years)				
20-30	6 (14)	6.50 (2.88)	36.95 (13.98)	67.08 (15.61)
30-40	23 (52)	4.87 (2.85)	34.52 (11.21)	63.48 (17.4)
40-50	14 (32)	4.86 (1.96)	30.86 (5.2)	65.54 (15.42)
50-60[b]	1 (2)	4.00	30.09	62.50
P value		.49	.38	.90
Respondent				
Guardian	27 (61)	4.70 (2.61)	35.29 (12.33)	67.13 (17.29)
Patient	17 (39)	5.65 (2.47)	30.89 (3.03)	60.59 (13.3)
P value		.15	.57	.19[c]
Occupation				
Housework/parenting	23(52)	4.96 (2.69)	31.44 (7.84)	63.15 (16.52)
Professional	4 (9)	4.25 (2.22)	29.44 (1.17)	54.38 (2.39)
Sales	2 (5)	3.00 (1.41)	33.46 (5.01)	57.50 (21.21)
Office job	7 (16)	5.57 (2.51)	41.30 (14.53)	74.29 (19.02)
Unemployed	2 (5)	6.00 (4.24)	30.04 (0.10)	72.50 (3.54)
Other	6 (14)	5.83 (2.64)	36.82 (13.95)	65.42 (14.44)
P value		.70	.03	.40[d]

[a]Mann-Whitney *U* test and Kruskal-Wallis rank-sum test were used to calculate *P* values between two groups and three groups, respectively, unless otherwise indicated.

[b]SD not available since there is only one value.

[c]Calculated using the *t* test.

[d]Calculated using analysis of variance.

Discussion

Principal Findings

This proof-of-concept study applied a blockchain-based medical document monitoring system in two primary care clinics. Although there was a lack of recognition of the concept of blockchain (compared with bitcoin), participants' trust level in blockchain was higher than that in bitcoin. Moreover, although there were few people who had experienced changes in their medical documents, the participants considered that medical records should not be changed, and therefore that this monitoring system is necessary.

There have been some blockchain-based implementations for managing medical records. Ariel C Ekblaw designed MedRec—a decentralized record management system for EHRs—and implemented the pilot system in the Beth Israel Deaconess Medical Center [24]. Zhang et al [9] designed the architecture of DApp, named FHIRChain, based on 5 key requirements provided by the ONC interoperability roadmap and demonstrated the prototype. Roehrs et al [11,12] designed a distributed architecture model to integrate PHRs, which was called OmniPHR, and showed the performance of the prototypes. However, most of these models focus on data storage and sharing, whereas our study focused on monitoring changes in medical documents themselves.

There are some concerns about creating blockchain-based EHR systems because sensitive data—such as medical records—are protected by the General Data Protection Regulation legislation [25], which ensures that full control of data are given to data owners. Four rule-of-thumb principles that entrepreneurs and innovators can consider when designing blockchain-based apps have been proposed. Among these principles, the second states that personal data should be avoided on blockchain using data obfuscation, encryption, and aggregation techniques [26]. To comply with this principle, we used the hash value of the document using SHA-256—which is considered to be one of the most popular hashing algorithms in the world—to make it difficult to reverse the original.

Some blockchain systems ensure privacy and confidentiality using zero knowledge proofs (ZKPs) such as Zerocash [27]. ZKPs allow data to be verified without revealing the data. Although the app does not have access to the exact medical document contents, it can show whether the medical document has been changed or not. Our system therefore satisfies the concept of ZKPs because the app only identifies whether the item was changed or not, without knowledge of the item's content.

Medical records may be often falsified to hide medical accidents [28] or to claim insurance by fraud [29]. In Korea, the Medical Service Act states that "Where any medical personnel or the founder of a medical institution makes an addition or revision to electronic medical records, he/she shall separately keep the access logs thereof, as prescribed by Ordinance of the Ministry of Health and Welfare" [30]. This act was implemented in September 2018 and the situation is likely to be similar in other countries. Most of the tertiary hospitals in Korea have well-integrated EHR systems that can manage changes in medical documents. However, primary care clinics use vendor-dependent EHR systems, which are not equipped to track changes in documents. Therefore, we chose a primary care clinic to trace changes in records.

Limitations

This study has several limitations that should be noted. All logs in the system indicate that there were no changes to medical documents during the study period owing to the short duration and small sample size. In fact, there were not many documents in primary clinics to work with compared with those available at tertiary hospitals; the fewer the documents in primary clinics, the fewer the changes in documents. Nevertheless, this study demonstrated that the app functions as designed when using simulation data. Additionally, the selection of the hospitals was biased toward pediatric clinics. Finally, the SUS scores in this study were low. Because there were no changes in medical documents, the participants barely had a chance to realize the value of the app.

Conclusions

This study introduced an app that notifies patients of changes in medical records using blockchain technology. Blockchain helps the app to inform patients of changes in their documents without uploading the medical record itself to the network. Therefore, blockchain can help ensure the transparency of medical records and advance patient empowerment.

Acknowledgments

This study was supported by the Foundational Technology Development Program (NRF2019M3E5D406468212) of the Ministry of Science and ICT, Republic of Korea, and a grant of the Korea Health Technology Research and Development Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (KHIDIHI19C1015010020).

Authors' Contributions

MS, EL, and YP all had a role in the study concept and design. MS, SP, and SJ had a role in acquisition, analysis, or interpretation of data. MS, EL, JL, and YP drafted the manuscript, and MS, EL, JL, and YP critically revised the manuscript. YP and EL obtained funding. SP and SJ provided technical support. All authors had access to the data and together had the final responsibility regarding the decision to submit for publication.

Conflicts of Interest

EL is the employer of Medibloc, Inc, and SJ is an employee of Medibloc, Inc. JL provided informatics consultation services to the Yonsei University School of Medicine and MediBloc Inc since 2018.

Multimedia Appendix 1
FHIR Profiles of EHR Data.
[PNG File , 258 KB - jmir_v22i8e19657_app1.png]

Multimedia Appendix 2
Modified System Usability Scale (SUS).
[DOCX File , 14 KB - jmir_v22i8e19657_app2.docx]

Multimedia Appendix 3
Transaction metadata with simulation data.
[DOCX File , 13 KB - jmir_v22i8e19657_app3.docx]

Multimedia Appendix 4
Screenshots of medical document changes app tested with simulation data.
[DOCX File , 115 KB - jmir_v22i8e19657_app4.docx]

References

1. Scheurer D. Hospitalists Should Embrace Advances, Transparency in Health Record Technology. The Hospitalist. 2015 Jul. URL: https://www.the-hospitalist.org/hospitalist/article/122299/hospitalists-should-embrace-advances-transparency-health-record [accessed 2020-07-30]
2. The Health Insurance Portability and Accountability Act (HIPAA). United States Department of Labor. URL: http://purl.fdlp.gov/GPO/gpo10291 [accessed 2020-07-30]
3. Overview of the national laws on electronic health records in the EU Member States and their interaction with the provision of cross-border eHealth services: Final report and recommendations. Health Programme of the European Union. URL: https://ec.europa.eu/health/sites/health/files/ehealth/docs/laws_report_recommendations_en.pdf [accessed 2020-07-30]
4. Wright CS. Bitcoin: A Peer-to-Peer Electronic Cash System. SSRN 2019 Oct 19:1-10 [FREE Full text] [doi: 10.2139/ssrn.3440802]
5. Hussien HM, Yasin SM, Udzir SNI, Zaidan AA, Zaidan BB. A Systematic Review for Enabling of Develop a Blockchain Technology in Healthcare Application: Taxonomy, Substantially Analysis, Motivations, Challenges, Recommendations and Future Direction. J Med Syst 2019 Sep 14;43(10):320. [doi: 10.1007/s10916-019-1445-8] [Medline: 31522262]
6. Patel V. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. Health Informatics J 2019 Dec 25;25(4):1398-1411. [doi: 10.1177/1460458218769699] [Medline: 29692204]
7. McFarlane C, Beer M, Brown J, Prendergast N. Patientory: A Healthcare Peer-to-Peer EMR Storage Network v1. Entrust Inc. 2017 Apr. URL: https://patientory.com/wp-content/uploads/2017/04/Final-Patientory_Whitepaper.pdf [accessed 2020-07-30]
8. Li H, Zhu L, Shen M, Gao F, Tao X, Liu S. Blockchain-Based Data Preservation System for Medical Data. J Med Syst 2018 Jun 28;42(8):141-112. [doi: 10.1007/s10916-018-0997-3] [Medline: 29956058]
9. Zhang P, White J, Schmidt D, Lenz G, Rosenbloom S. FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data. Comput Struct Biotechnol J 2018;16:267-278 [FREE Full text] [doi: 10.1016/j.csbj.2018.07.004] [Medline: 30108685]
10. Rahmadika S, Rhee K. Blockchain technology for providing an architecture model of decentralized personal health information. Int J Engin Bus Manage 2018 Aug;10:184797901879058. [doi: 10.1177/1847979018790589]
11. Roehrs A, da Costa CA, da Rosa Righi R. OmniPHR: A distributed architecture model to integrate personal health records. J Biomed Inform 2017 Jul;71:70-81 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.012] [Medline: 28545835]
12. Roehrs A, da Costa CA, da Rosa Righi R, da Silva VF, Goldim JR, Schmidt DC. Analyzing the performance of a blockchain-based personal health record implementation. J Biomed Inform 2019 Apr;92:103140 [FREE Full text] [doi: 10.1016/j.jbi.2019.103140] [Medline: 30844481]
13. Maslove DM, Klein J, Brohman K, Martin P. Using Blockchain Technology to Manage Clinical Trials Data: A Proof-of-Concept Study. JMIR Med Inform 2018 Dec 21;6(4):e11949 [FREE Full text] [doi: 10.2196/11949] [Medline: 30578196]
14. Wong DR, Bhattacharya S, Butte AJ. Prototype of running clinical trials in an untrustworthy environment using blockchain. Nat Commun 2019 Feb 22;10(1):917 [FREE Full text] [doi: 10.1038/s41467-019-08874-y] [Medline: 30796226]
15. Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. : IEEE; 2013 Presented at: Proceedings of the 26th IEEE international symposium on computer-based medical systems; June 20-22, 2013; Porto, Portugal. [doi: 10.1109/cbms.2013.6627810]
16. Lee J, Park YR. Hospital Information System - MediBloc. Simplifier.Net. URL: https://simplifier.net/hospitalinformationsystem-medibloc2 [accessed 2020-07-30]
17. MediBloc Panacea-Core. GitHub. URL: https://github.com/medibloc/panacea-core [accessed 2020-07-30]
18. Tendermint Core (BFT Consensus) in Go. GitHub. URL: https://github.com/tendermint/tendermint [accessed 2020-07-30]
19. Medical record change log app. Google Play Store. URL: https://play.google.com/store/apps/details?id=org.medibloc.panacea [accessed 2020-07-25]

20. Schuh S, Shy O. US consumers adoption and use of Bitcoin and other virtual currencies. 2016 Apr 21 Presented at: DNB Payments Conference 2016 - Retail Payments: Mapping Out The Road Ahead; April 21-22, 2016; Amsterdam.
21. Son J, Robb T, Charismiadji I. Computer literacy and competency: a survey of Indonesian teachers of English as a foreign language. Comput Lang Learn Electron J 2011;12(1):26-42.
22. Brooke J. SUS-A quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. Boca Raton, FL: CRC Press; 1996.
23. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. Int J Hum-Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
24. Ekblaw A. MedRec: blockchain for medical data access, permission management and trend analysis. Massachusetts Institute of Technology, Thesis. Cambridge: Massachusetts Institute of Technology; 2017. URL: https://dspace.mit.edu/handle/1721.1/109658 [accessed 2020-07-30]
25. General Data Protection Regulation (GDPR). Intersoft Consulting. 2018. URL: https://gdpr-info.eu/ [accessed 2020-07-30]
26. Overview of the national laws on electronic health records in the EU Member States (2016). European Commission. 2014. URL: https://ec.europa.eu/health/ehealth/projects/nationallaws_electronichealthrecords_en [accessed 2020-07-30]
27. Sasson EB, Chiesa A, Garman C, Green M, Miers I, Tromer E, et al. Decentralized anonymous payments from bitcoin. 2014 Presented at: 2014 IEEE Symposium on Security and Privacy; May 18-21, 2014; San Jose, CA, USA p. 459. [doi: 10.1109/sp.2014.36]
28. Alteration of Medical Records. Miller & Zois. URL: https://www.millerandzois.com/doctor-alter-records.html [accessed 2020-07-30]
29. Health Insurance Fraud. PA Insurance Fraud Prevention Authority. URL: https://www.helpstopfraud.org/Types-of-Insurance-Fraud/Health [accessed 2020-07-20]
30. Medical Service Act. Korea Legislation Research Institute. URL: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=40970&lang=ENG [accessed 2020-07-30]

Abbreviations

EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
ONC: Office of the National Coordinator for Health Information Technology
PHR: personal health record
SUS: System Usability Scale
ZKP: zero knowledge proof

Edited by G Eysenbach; submitted 27.04.20; peer-reviewed by MG Kim, CM Choi; comments to author 12.06.20; revised version received 25.06.20; accepted 25.06.20; published 14.08.20.

Please cite as:
Sung M, Park S, Jung S, Lee E, Lee J, Park YR
Developing a Mobile App for Monitoring Medical Record Changes Using Blockchain: Development and Usability Study
J Med Internet Res 2020;22(8):e19657
URL: http://www.jmir.org/2020/8/e19657/
doi:10.2196/19657
PMID:32795988

Original Paper

Blockchain in Health Care Innovation: Literature Review and Case Study From a Business Ecosystem Perspective

Shuchih Ernest Chang[1], PhD; YiChian Chen[1], PhD

Graduate Institute of Technology Management, National Chung Hsing University, Taichung, Taiwan

Corresponding Author:
YiChian Chen, PhD
Graduate Institute of Technology Management
National Chung Hsing University
145 Xingda Rd.
South Dist.
Taichung, 40227
Taiwan
Phone: 886 4 22859465
Fax: 886 4 22859497
Email: cycx1000@gmail.com

Abstract

Background: Blockchain technology is leveraging its innovative potential in various sectors and its transformation of business-related processes has drawn much attention. Topics of research interest have focused on medical and health care applications, while research implications have generally concluded in system design, literature reviews, and case studies. However, a general overview and knowledge about the impact on the health care ecosystem is limited.

Objective: This paper explores a potential paradigm shift and ecosystem evolution in health care utilizing blockchain technology.

Methods: A literature review with a case study on a pioneering initiative was conducted. With a systematic life cycle analysis, this study sheds light on the evolutionary development of blockchain in health care scenarios and its interactive relationship among stakeholders.

Results: Four stages—birth, expansion, leadership, and self-renewal or death—in the life cycle of the business ecosystem were explored to elucidate the evolving trajectories of blockchain-based health care implementation. Focused impacts on the traditional health care industry are highlighted within each stage to further support the potential health care paradigm shift in the future.

Conclusions: This paper enriches the existing body of literature in this field by illustrating the potential of blockchain in fulfilling stakeholders' needs and elucidating the phenomenon of coevolution within the health care ecosystem. Blockchain not only catalyzes the interactions among players but also facilitates the formation of the ecosystem life cycle. The collaborative network linked by blockchain may play a critical role on value creation, transfer, and sharing among the health care community. Future efforts may focus on empirical or case studies to validate the proposed evolution of the health care ecosystem.

(J Med Internet Res 2020;22(8):e19480) doi:10.2196/19480

KEYWORDS

blockchain; health care industry; business ecosystem; smart contract; paradigm shift

Introduction

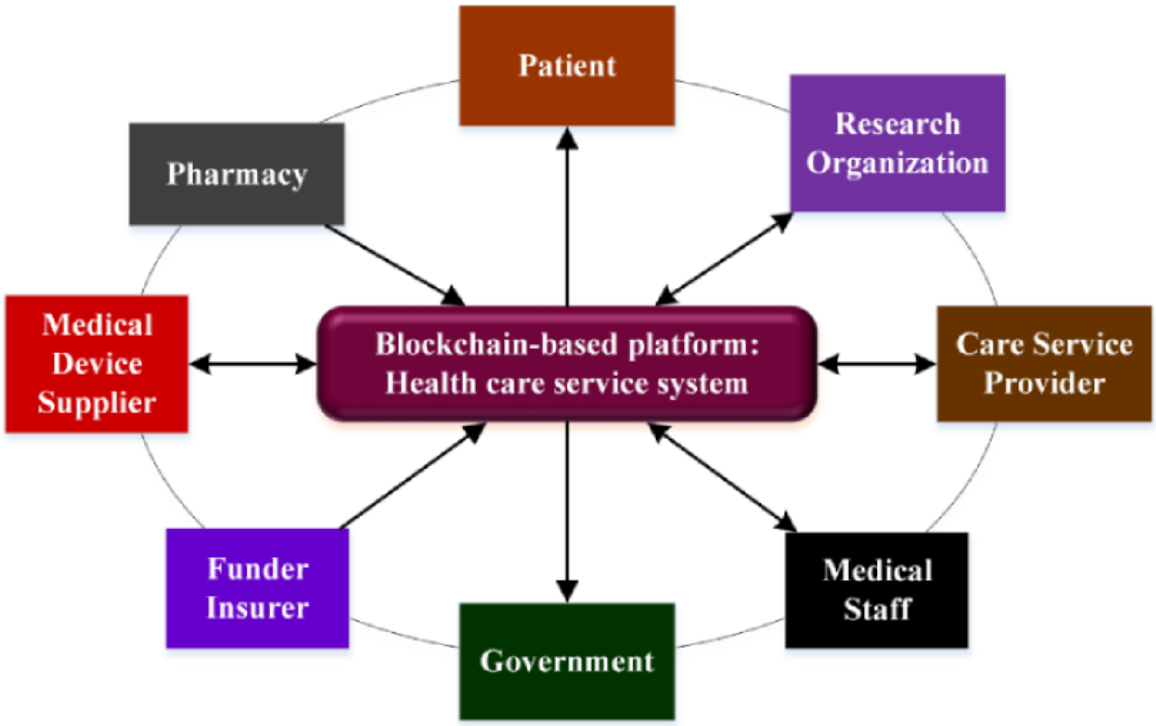
Background

In the last decade, blockchain technology has gained growing attention from both academia and practitioners in a range of industries, including banking, insurance, trade, and medicine. Blockchain has potential in various industries, including in financial applications, supply chains [1], the insurance industry [2], and even medical health care records [3-5]. Through maintaining an immutable, tamper-proof, consecutive list of transactional data in a distributed network, blockchain has created several disruptions in incumbent business processes with its unique features. Having a promising capability to improve information flow, sharing, and transmission among participating nodes (ie, partners in the real system), blockchain is expected to transform legacy operations with innovative service delivery and ownership transfer [6]. Blockchain adoption and pioneer pilots in different sectors have shown its power in transforming traditional working paradigms.

Blockchain, as a kind of distributed ledger technology, enables data storage, sharing, and verification under a distributed peer-to-peer network [7]. Participating nodes (ie, entities) may cooperatively maintain the common shared ledger by contributing efforts to data verification via cryptography. Blockchain can be viewed as a consecutive list of transactions that are chronologically appended to the previous ones. Updates of any part need to be verified and then recorded on the chain. This process is achieved by participating nodes' contributions to solving the cryptographical puzzle, which in turn increases the difficulty of malicious tampering and alterations. In this sense, all transactions are visible and immutable for all parties, thus providing audit trails and data integrity. In addition, its affiliated technology, smart contracts, can be deployed on blockchain-based platforms to activate or enforce specific desired processes. Smart contracts are computer protocols that aim to execute terms of a contract or agreements [8]. In real practice, smart contracts can be coded with computer languages to interact with one another and be triggered by events in the real world [9]. These attributes, when deployed on blockchain system, may facilitate business logic and process automation.

Recent publications, including technical reports, research articles [10,11], and consulting papers [12], have addressed blockchain's potential to reshape the complex operations in the field of health care. Blockchain applications in the realm of health care may be promising; however, the compositions and interactions among major health care stakeholders, such as patients, care service providers, pharmacies, funders and insurers, medical device suppliers, and research organizations, are rather complex (see Figure 1). Extant research topics on how these stakeholders may achieve benefits by the use of blockchain technology have been addressed from the perspective of a single industry. Comprehensive discussions on the development and potential evolution of blockchain-based health care have been discussed less. It is noted that activities and interactions among stakeholders may have crossed a variety of industries. As Moore [13] has suggested, a careful systematic approach to business strategy needs to consider firms in the scope of a larger ecosystem rather than a member of a single industry. To better elucidate the evolution of a health care ecosystem utilizing blockchain innovation, stakeholders must address cooperative and competitive issues when attempting to deliver tangible and intangible values to meet customer needs.

Figure 1. Typical health care ecosystem.



Through unique distributed schemes and immutable shared ledgers, blockchain allows better transparency, security, privacy, traceability, and trust-free environment among players [14]. This implies that blockchain connects not only individual siloed databases via decentralized governance but also the ecosystem surrounding health care stakeholders. However, this may lead to more complex supply-and-demand relationships and interactions among actors who operate their businesses in an original centralized manner. Therefore, this study attempts to shed light on driving inertia from a *business ecosystem* perspective rather than through a traditional supply chain vision. Moore [15] defined the business ecosystem as an economic community loosely connected by a group of interacting organizations and individuals who share common values and who coevolve with one another. Researchers also extended this argument by addressing cross-industry collaboration rather than disparate interactions among directly connected counterparties [16,17]. This concept provides broader visions when blockchain interplays, connects, and disintermediates the dynamic

relationships among connected medical communities, service providers, and end customers.

However, there is very limited research on blockchain-based health care ecosystems in the extant literature. Previous research efforts on blockchain mainly focused on technological potential [18], individual applications [19], medical record accessibility [20], and general influence. Others discussed the proof-of-concept of system design [21,22], adoption attitudes [23], governance, challenges, and opportunities in future research [24-26]. Few extant articles in the literature have addressed dynamic relationships among medical stakeholders with an overview of the blockchain ecosystem. Therefore, this research aims to investigate how blockchain can lead to a coevolving health care ecosystem by collating overviews of potential evolutions of blockchain-enabled health care applications from recent literature from a perspective of the business ecosystem. In this study, we address two research questions:

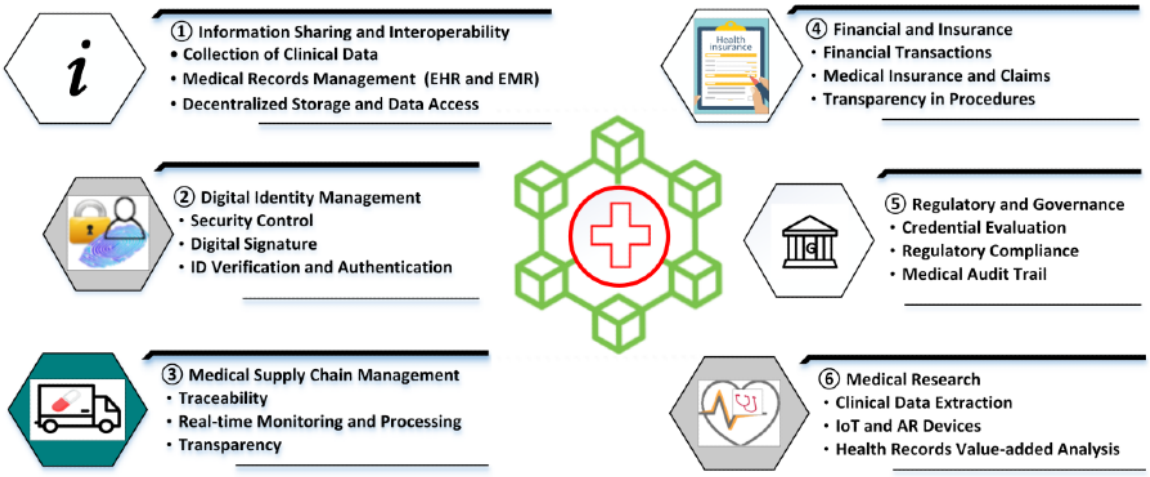
1. Research question 1: What kind of potential effects from recent innovations and applications make use of blockchain in the health care industry?
2. Research question 2: How do health care stakeholders participate, interact, and evolve in the blockchain-based ecosystem and how do they collaboratively contribute to a potential paradigm shift?

To shed light on blockchain's influence on value creation and capture of medical stakeholders, we examine and address these research questions from a perspective of the business ecosystem, with an aim to contribute to the body of knowledge in health care.

Existing Service Process and Blockchain Roles

Traditionally, medical information is located at disconnected databases in clinics, labs, or medical institutions. Aggregating health data from disparate sources and gaining a holistic view of patient treatment history have been difficult and costly. As blockchain can store transaction logs among participants, better transparency and completeness of treatment history could be achieved. Blockchain may drive the digital transformation of legacy information sharing [27]. Traditional paper-based processes and manual processing could be reduced and better interoperability among disconnected health systems is feasible. In addition, traditional medical supply chains have suffered from poor traceability and invisible provenance. Blockchain may provide solutions to improve transparency and real-time monitoring from manufacturing to delivery. Other focused areas also include secure identity management [28], audit and governance, and facilitation for medical research (see Figure 2).

Figure 2. Blockchain's role in improving the health care service system. AR: augmented reality; EHR: electronic health record; EMR: electronic medical record; IoT: internet of things.



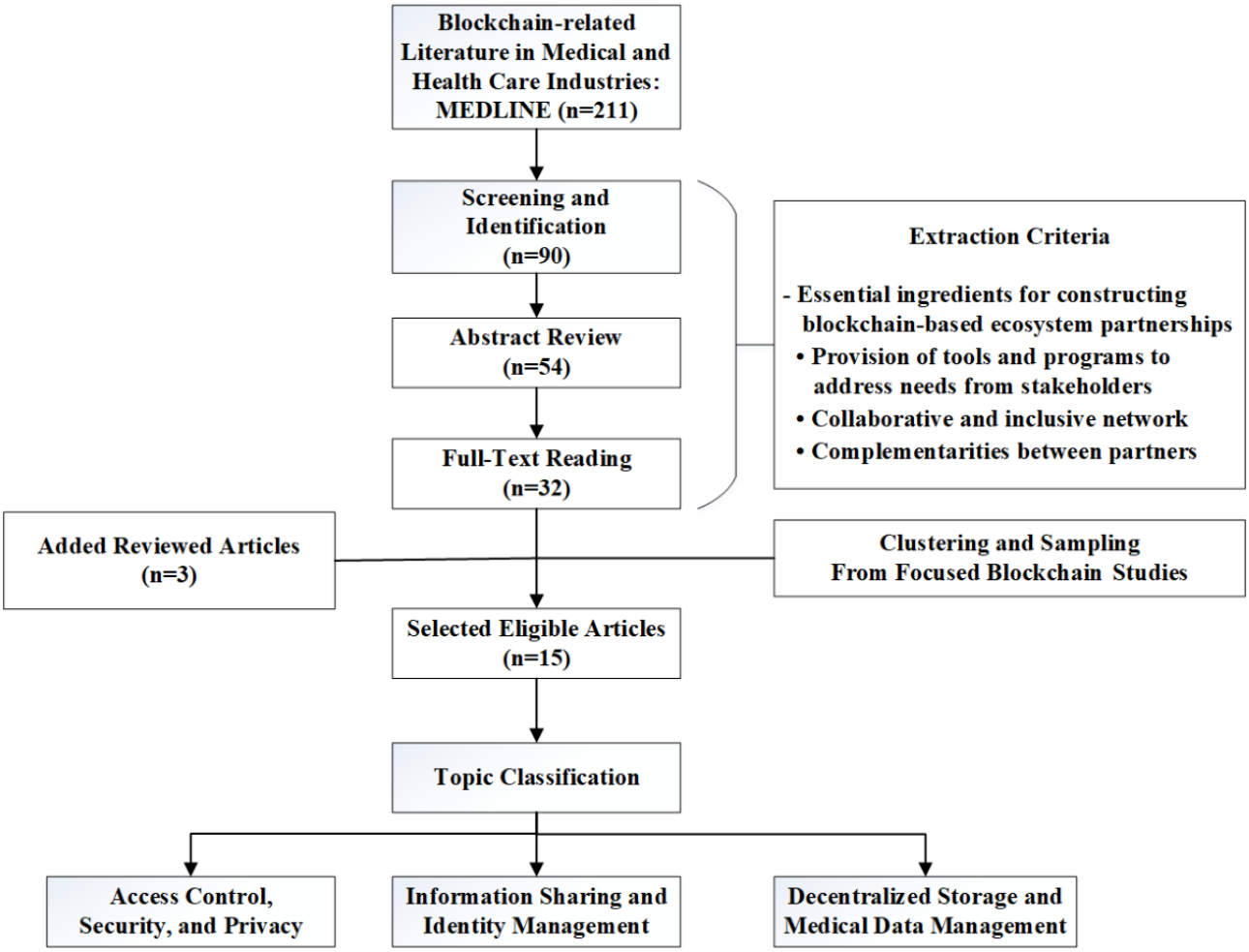
Methods

Literature Review

To answer research question 1, we conducted a literature survey to find the current state and potential of blockchain applications in the health care field. Other than using a systematic approach, we focused on specific applications that may be enabled by blockchain to transform the interaction and manipulation of a health care ecosystem. Some review articles in the recently published literature were also selected to help understand the potential evolution among health care stakeholders.

Figure 3 illustrates the search and review process of the focused literature. We searched for blockchain studies in medical and health care fields and conducted subsequent article screening and identification; abstract and text reviewing were conducted to select focused literature. The numbers in the flowchart boxes in Figure 3 denote articles that were available after the respective procedural steps. From the ecosystem perspective, the literature selection and extraction criteria paid attention to the capabilities that relevant studies highlight and that elucidate essential ingredients for constructing blockchain-based ecosystem partnerships. Some review articles were added to give a general overview of blockchain-based health care studies. Sampled articles were extracted from the filtered corpus to highlight focused topics, such as data management, information sharing, access control, security, and privacy.

Figure 3. The procedural framework for the focused literature review. MEDLINE: Medical Literature Analysis and Retrieval System Online.



Case Study

In this study, recent blockchain-based health care projects were examined to shed light on the disruption to health care practice. The case study, a qualitative method, refers to a systematic analysis of a specific target from a wide perspective and enables comprehension throughout the exploration process [29]. Applying this approach, researchers suggested that consideration of research objectives, contexts, and representativeness needs to be stressed [30,31], followed by inclusion of selection protocols suggested by previous literature articles, secondary resources from news archives, consulting reports, company websites, and academic articles [32]. Case study results were then collated to answer research question 2 and to elucidate the understanding of cooperative and competitive strategies and operational business schemes in the health care context. We selected the IBM blockchain–health care initiative [33] as the body of the target case and combined the concept of the business ecosystem with the health care context to analyze the interactions, cooperative or competitive, among species (ie, ecosystem members). Furthermore, major players' roles and influences in the blockchain-based health care ecosystem were analyzed to give research implications.

Business Ecosystem Perspective

Overview

This study analyzed the potential evolutionary path of blockchain-based health care innovation from a business ecosystem perspective. Moore [13] proposed the life cycle of a business ecosystem and divided it into four stages: birth, expansion, leadership, and self-renewal or death. We identified four major development stages within which health care stakeholders interact with each other and evolve chronologically with their roles and cooperative and competitive challenges. Iansiti and Levien [34,35] extended Moore's concepts by defining the roles of actors and argued that these roles were formed by large, loosely connected networks of entities. They further classified three categories of the actors' roles as keystone, dominator, and niche player. The business ecosystem is comprised of diverse participants across various industries. The overall health of the ecosystem depends on the positive interactions and operations among stakeholders.

Birth: Pioneering

During this stage, entrepreneurs focus on the value creation or proposition that meets customers' needs. The product or service needs to be presented in its best form to draw potential customers' attention and effectively deliver its value. Leaders in the ecosystem aggregate suitable suppliers to take part in the

environment and attempt to incorporate business partners' capabilities to optimize the value package to customers.

Expansion

The ecosystem grows and expands its territories. The business ecosystem faces competition to increase market share against its rivals. Firms may devote a large number of endeavors to marketing activities for increasing sales. Meanwhile, to improve overall performance, issues regarding large-scale adoption and distribution are stressed. In this stage, while incomplete ecosystems are likely to be expelled from competition, superior ones may integrate community members to complete sound supply chains, thus achieving ecosystem stability. Required conditions in the expansion stage include value-oriented business concepts and the corresponding potential to broaden scalability.

Leadership: Authority

Following the expansion, the leader or integrator needs to guide the direction of investment and technology standards. As innovation is a crucial factor for evolving ecosystems, stakeholders may find their positions and revenue models through the leader's guidance. While the bargaining power of suppliers increases during this stage, the system integrator needs to enhance the supply chain management with alternative options to assure the stability of production and distribution. How firms constantly create values to maintain their importance in the ecosystem is critical to the overall health and continued improvement of the ecosystem.

Self-Renewal or Death

This stage occurs when firms face external threats, for example, changes in regulations or the rise of new ecosystems and innovations such as emerging technologies. Original business communities may undergo different levels of change and fluctuation. The altered environment may challenge the survival of original members. How leaders detect the potential changes and new incoming elements and threats, thereby correspondingly reacting to these alterations, may decide the future outcome of the ecosystem. When facing obsolescence, either self-renewal by incorporating new innovative ideas or stepping toward death depends on the capabilities to enable system transformations.

Results

Reviewed Literature

Overview

The pursuit of building a sustainable and healthy ecosystem is essential for participating stakeholders. When the requirements for building a health care ecosystem are considered, we found issues that are being addressed by extant research studies. We selected a number of articles to elucidate the recent research foci. Table 1 summarizes several related articles regarding blockchain in health care; these articles were published in academically rigorous peer-reviewed scientific journals. Focused topics in the blockchain–health care ecosystem are briefly collated in the following sections.

Decentralized Storage and Medical Data Management

Centralization of health data from disparate sources has long been a major pain point for further medical usage. Generally, disconnected data sources could be utilized to increase the integration and aggregation of medical data. Based on the distributed nature of blockchain, researchers have stressed that there are data storage and management issues in clinical trials [22], insurance [2], and personal health scenarios [36].

Information Sharing and Identity Management

Based on the immutable and distributed features of blockchain, a common shared ledger may facilitate health information exchange (HIE). Some proof-of-concept studies have covered the potential and major contributions to these topics; for example, Ali et al [37] focused on remote health monitoring, Hau et al [23] surveyed stakeholders' attitudes, and Esmaeilzadeh and Mirzaei [18] conducted an experimental study to understand patients' perceptions of various exchange mechanisms. In addition, while several researchers conducted literature reviews to shed light on potential strengths and limitations of blockchain applications [38,39], others reviewed potential identity management solutions [28] and developed evaluation frameworks for assessing performance of blockchain initiatives [40].

Access Control, Security, and Privacy

As access control and authentication are major security requirements for managing health care and medical data, researchers have proposed some blockchain-based prototypes to provide solutions to current health systems [41,42]. Digitization of electronic medical records (EMRs) may introduce cyberattack risks on data security and privacy when stakeholders, such as providers, payers, and researchers, attempt to interact with patient data. Blockchain-enabled solutions may maintain patient-sensitive data through a friendly approach [43-45].

Table 1. Overview of blockchain-based health care applications in the research literature.

Article, author (year)	Research method	Focus	Key findings
Maslove et al (2018) [22]	Proof-of-concept	Clinical trials data management	A web-based interface, BlockTrial, allowed patients to grant researchers access to their data and allowed researchers to submit queries for data that are stored off-chain. The proposed system increased the trustworthiness of data collected during clinical research, with benefits to researchers, regulators, and drug companies alike.
Zhou et al (2018) [2]	Proof-of-concept	Medical insurance storage system	A blockchain-based medical insurance system, MIStore, deployed on Ethereum was proposed to serve as a medium for accountable record keeping. Medical insurance data were better managed in a distributed way.
Chen et al (2018) [36]	System design	Secure medical data storage and service framework	A storage scheme to manage personal medical data based on blockchain and cloud storage was proposed without third parties. No single party would have absolute power to affect the processing; better and more secure medical storage could be achieved.
Ali et al (2020) [37]	Proof-of-concept	Remote health monitoring and data sharing	A solution for patients to share biomedical data with their doctors was proposed without manipulation by trusted third parties. In various health monitoring scenarios, three use cases—cardiac monitoring, sleep apnea testing, and electroencephalogram following epileptic seizures—were tested for system feasibility.
Hau et al (2019) [23]	Survey	Attitudes on information sharing	Medical doctors reported significantly more negative attitudes than patients. Furthermore, self-employed doctors reported more negative attitudes than employed doctors and university professors.
Esmaeilzadeh and Mirzaei (2019) [18]	Experimental study	Medical information exchange	Significant differences existed in patients' perceptions of various exchange mechanisms with regard to patient privacy concern, trust in competency and integrity, opt-in intention, and willingness to share information. Participants held a favorable attitude toward the implementation of blockchain-based exchange mechanisms for privacy protection, coordination, and information exchange purposes. This study proposed the potential strengths and limitations of a blockchain-based attempt within a health information exchange context.
McGhin et al (2019) [38]	Literature survey and case study	Research challenges and opportunities	The survey presented a careful examination of specific blockchain issues and how they affect the health care industry. Health care industry requirements and blockchain potential effects in supporting these requirements were discussed.
Vazirani et al (2019) [39]	Systematic review	Blockchain implementation	Of the 71 included studies, the majority discussed potential benefits and limitations without evaluation of their effectiveness, although some systems were tested on live data.
Bouras et al (2020) [28]	Literature review	Identity management	This study presented state-of-art decentralized identity management using blockchain and highlighted the possible opportunities for future adoption. Decentralized models and pilot projects were presented to give implications.
Zhang et al (2019) [40]	Framework construction	Development of balanced scorecard evaluation framework	A framework was proposed to holistically assess the performance of blockchain initiatives in providing value-based care. By extending the concept of existing balanced scorecard evaluation, both the financial and nonfinancial benefits of blockchain initiatives were evaluated.
Shuaib et al (2019) [41]	Literature review	Blockchain potential in improving secured digitized medicine	The digital ledger technology could be used to improve current systems. Data are distributed and decentralized, preventing loss and allowing recovery in the event of an attack. Audit trails keep track of what transactions and modifications are made to patient records, while notifying all users on the network. Patients will be given more control over who has access to their data by selecting who carries the cryptographic keys required to decrypt and view them. In addition, issues such as scalability need more research efforts.
Guo et al (2018) [42]	System design	Secure signature authentication	An attribute-based signature scheme with multiple authorities, in which a patient endorses a message according to the attribute while disclosing no information other than the attested evidence, was proposed. By sharing the secret pseudorandom function seeds among authorities, this protocol resists collusion attack out of N from N–1 corrupt authorities.
Kadam et al (2019) [43]	System design	Patient data privacy	Patient data were secured by applying the Secure Hash Algorithm for the generation of hash values and the Paillier algorithm to re-encrypt the same information regarding patient data that is divided among a number of different servers. This approach increases the difficulty of hacker access and attack and maintains the security principles (ie, availability, integrity, and confidentiality).

Article, author (year)	Research method	Focus	Key findings
Al Omar et al (2019) [44]	System design	Health care data privacy	A patient-centric health care data management system was proposed by using blockchain technology for storage, which helped to attain privacy. Cryptographic functions were used to encrypt patients' data and to ensure pseudonymity.
Yue et al (2016) [45]	System design	Health care data privacy	The Healthcare Data Gateway architecture, using blockchain, enabled patient-centric data management (ie, own, control, and share patient data) in a secure way without violating privacy, which improves the intelligence of health care systems. The proposed access model ensures better manipulation of health care data and enables untrusted third parties to conduct computation with patient data without violating privacy.

Case Study of the IBM Blockchain–Health Care Initiative

On January 24, 2019, IBM announced its collaborative blockchain initiative with major health care players, including Aetna (acquired by pharmacy and health plan provider CVS Health), Anthem (health plan provider), Health Care Service Corporation (the largest customer-owned health insurance provider in the United States), and PNC Bank [46]. IBM has been searching for new opportunities by leveraging the potential of blockchain and attempting to build up a special networked health care ecosystem. In the last few months, health organizations, health care providers, start-ups, and technology companies joined in this initiative to grow the Health Utility Network, of which Cigna and Sentara Healthcare are participants. The aim is to drive digital transformation by providing better transparency and interoperability. Participants may reap benefits from building, sharing, and deploying solutions to incumbent challenges in the health care context. Major issues and potential blockchain use cases are enumerated as follows:

1. Provenance and traceability of pharmaceutical supply chain: fake and counterfeit drugs could be troublesome and dangerous issues as drug provenance is difficult to track in a cross-border setting. A large number of handovers from manufacturers, shippers, distributors, retailers, and pharmacies may cause inaccuracies and disputes in medical delivery operations. Counterfeit drugs with improper ingredients and dosages may jeopardize the health of patients and even cause legal disputes among manufacturers, suppliers, and customers. With immutable, tamper-proof, and trackable characteristics, blockchain may provide solutions to authenticity and traceability of transferred assets along with auditable and secure transaction records among stakeholders. For example, in a private drug blockchain, drug registration by pharmaceutical companies may grant a higher level of trustworthiness and authentic proof. Also, these companies, acting as dominators, could assign the roles of the actors; some of them may have the rights for registration while others may conduct verification of transactions. The provenance of drugs can be assured via verification processes with related manufacturing or identity information when appended on-chain, making it easy to be tracked.
2. Data management during clinical trials: when clinical trials are implemented, numerous data are produced by different devices via the operation of medical staff. How these data are stored, transmitted, shared, and utilized for medical therapy or operations is critical to existing manual systems. Errors and fraud during clinical trials operations could be generated via malicious alterations or unintentional mistakes. Typical flaws could occur when trial procedures are inaccurately designed by biased intentions from actors or inconsistent records and responses from patients' evolutionary medical reports. Blockchain in this case could provide proof-of-existence for any form of documentation. The information needs to be verified via the consent of the participating nodes and not under a single entity's control. Modifying or changing information would be cryptographically difficult to conduct among a majority of network players, thus making documentation highly trusted.
3. EMR and electronic health record (EHR) management: where patient or medical records are concerned, a challenge is that individual medical data are not easily accessed by different medical institutions or clinics. While the medical information is stored disparately in various databases or systems, it is difficult to deliver proper medication and care service in a personalized context. Sensitive data can also hinder the transmission efficiency among medical organizations. How to access, share, and utilize a holistic medical treatment history in a secure way remains a challenging issue in centralized EMR systems. However, with the help of distributed ledger technology, blockchain may have potential regarding the manipulation and access control of such EHR and EMR systems. Blockchain platforms can be combined with existing EHR and EMR systems, either in the cloud computing environment or otherwise, through the use of Oracle and data gateways. Patients can share their medical records, with or without permission, to registered users or stakeholders on a medical blockchain. Patients may decide the level of information disclosure through smart contract settings to specific users, thus receiving rewards from the blockchain system, accordingly. As described above, blockchain could facilitate the sharing and management of EHRs and EMRs among supply and demand entities. Related data analysis and rewards from sharing could potentially promote the participation of the medical community and, consequently, leverage a network effect.

In health care, major inefficiencies can arise from clinical operations, administrative processing, and frictions among disparate systems. These pain points have decreased the overall performance and have led to poor customer experiences in regard to incumbent medical and health care systems and

services. The act of incorporating major players through blockchain-based systems and services in health care may help to develop a healthy, open-networked, and collaborative ecosystem. The blockchain-enabled collaboration aims to address the aforementioned challenges by pursuing reduced administrative error, mitigated system frictions, streamlined claims and payment transactions, and efficient information exchange. Iansiti and Levien expanded Moore's ecosystem view and proposed the strategies that firms might adopt to position themselves in the business ecosystem. The strategic roles include keystone, niche player, and physical dominator. The keystone in the business ecosystem provides a platform to which niche players add value and build offerings. Niche players account for the bulk proportion of the ecosystem and are responsible for value creation and innovation. The physical dominator directly controls the majority of a network via horizontal or vertical integration. In an IBM blockchain ecosystem, the major players' roles and corresponding functions are shown in Table 2 and are summarized as follows:

1. IBM: keystone—blockchain platform provider and coordinator.
2. Aetna of CVS: niche player—improves data accuracy and optimization of health care system operation.
3. Anthem: niche player—medical information exchange.
4. Health Care Service Corporation: physical dominator—reduces information fragmentation and improves claims procedures and health care system connection.
5. PNC Bank: niche player—facilitates payment transactions and supports medical finance.

Table 2. Major players' roles and influences in a blockchain-based health care ecosystem.

Major player type	Major players	Roles	Influences on the ecosystem
Keystone	IBM	Platform provider in the ecosystem Aim to create opportunities for niche players and support the operation of the whole system	Enable the establishment of a healthy environment, which leads to an organization's survival and prosperity Convene followers to achieve diversity
Physical dominator	Health Care Service Corporation or other health care service providers	Integrators in the ecosystem Integrators directly own and manage a large proportion of a network by using vertical or horizontal measures	Provide most products and services to meet customers' needs Exploit their positions to take over the network and extract the created value
Niche player	Aetna of CVS, Anthem, or PNC Bank	Value creators and innovators in the ecosystem Focus all potential endeavors on enhancing their narrow domain of expertise	Leverage complementary resources from others to create differentiated value Competition and cooperation of niche players support the coevolution of the ecosystem

Embracing of blockchain technology is not the privilege of this initiative only. Competitors making similar efforts, such as Change Healthcare, Hashed Health, Guardtime, Gem, and SimplyVital Health, have also teamed up to launch a blockchain pilot—Intelligent Healthcare Network with Blockchain Processes—in the realm of health care. Other competing projects with a more-or-less different focus have also led to consortia competition. Prominent examples include Synaptic Health Alliance, targeting provider directories and data reconciliation, and ProCredEx, focusing on storage and sharing of medical credentials. PNC Bank, acting as a partner of interdisciplinary alliance, stands in a public position and contributes its edge to facilitate transactions among patients, payers, and providers in both domestic and cross-border contexts.

Business Ecosystem With Evolutionary Life Cycle

Blockchain, as an emerging technological innovation, has provided opportunities for incumbent health care stakeholders. As for the IBM case, a collaboration of health care partners has resulted in a new ecosystem. Its potential evolutionary stages have formed a business ecosystem lens; these stages are summarized in Table 3.

At the *birth* stage, the IBM blockchain–health care pilot faces consortia competitions from other allies. Even though the focused markets might be slightly different from pilot to pilot, similar efforts and common objectives for driving digital transformation in the health care industry are the same. IBM, as a recognized leading enterprise blockchain provider, possesses an advantageous edge against competitors. When entering into the *expansion* stage, the key focus is to bring new innovations to market to increase the market share. This could be carried out by optimizing platform functionality, absorbing complementary health care members, and addressing the changing demands for customers. In addition, to outperform rival ecosystems, it is essential to build up technical or industrial standards in terms of competitive strategy [47]. During the *leadership* stage, the leading ecosystem may focus on future prospects for followers. This could be implemented by compelling suppliers and customers to complete sound visions; for example, integration with other disruptive technologies, such as machine learning, artificial intelligence, mobile and ubiquitous health, wearables, and internet of things (IoT). Inversely, to prevent pressure from increased bargaining power, actions such as using backward integration, searching multiple suppliers, increasing profile, and conducting market education are needed. At the last stage, the blockchain–health care ecosystem may step toward *self-renewal or death*. This may depend highly on capabilities that the existing ecosystem may possess; it can either innovate or be replaced with alternative ecosystems or paradigms.

Table 3. The evolutionary path of a blockchain–health care ecosystem: the IBM case.

Stage	Cooperative challenges	Competitive challenges
Birth	Stakeholders create new value propositions of blockchain-based ecosystems and define their roles when working with suppliers and customers Players seize opportunities Example: IBM blockchain–health care ecosystem	Protect ideas against competitors with similar offerings Pilot cases with similar features Examples: Change Healthcare's Intelligent Healthcare Network with blockchain processes, Synaptic Health Alliance, and ProCredEx
Expansion	Bring new innovations (ie, products or services) to market to increase the market share or coverage Strategy: optimize platform functionality, absorb complementary health care members, and identify and address changing demands from customers	Compete with and defeat rival implementations Expand market share by establishing market or technical standards Strategy: build up technical or industrial standards and expand the adoption of blockchain-based applications
Leadership	Make future prospects and encourage partners to step forward Measure: integrate with other disrupting technologies (eg, machine learning, artificial intelligence, mobile and ubiquitous health, wearables, and internet of things)	Maintain bargaining power against ecosystem players Measures: keep customers satisfied and strengthen the customer relationship management; use backward integration, search multiple suppliers, increase profile, and conduct market education
Self-renewal or death	Cope with innovators to generate or seize new opportunities or be replaced by alternative paradigms	Build high levels of entry barriers and customer switching costs to prevent being replaced by alternative ecosystems

Discussion

Comparative Analysis of the Existing System and the Future Ecosystem

Blockchain applications in the health sector have been receiving increased attention and prospects. We have summarized the current health care service pain points and highlighted the potential of blockchain in reshaping traditional practice and operations. Researchers have conducted literature reviews to report on the current challenges [48,49]. The major issues with the corresponding potential effects of blockchain are listed in Table 4.

Table 4. Health care service pain points and the potential effects of blockchain in the health care ecosystem.

Issue	Health care service pain points	Potential effects of blockchain leverage
Medical data storage	Highly disparate data sources across individual clinics or health care–related institutions	Decentralized data storage allows duplicate and immutable health records in the health network
Fraud and authenticity	Malicious attempts or human processing errors may cause fraud, alterations, or medical disputes Authorities are required for trust building among stakeholders Major issues include drug counterfeiting and provenance	Keeping critical items (ie, medical transactions or records) on blocks and permanently recording operations on-chain Mitigating the tampering issue via the verification and consensus architecture
Document type	Paper-based and manual processing causes difficulties in data aggregation	Supporting digitalized health documents deployed on secured shared ledger
Interoperability	Siloed data structures hinder interoperations across different databases	Blockchain-based networks enable interactions among health care stakeholders
Health claims and transactions	Inefficiencies that exist in clinical and administrative procedures and frictions among respective health systems have caused poor operations	Process automation facilitated by blockchain-based smart contracts enables streamlined claims and transaction procedures
Research data access and monetization	Challenges in aggregating, recruiting, and retaining data among medical parties and difficulties in monetization	Enabling of clinical trial data sharing and value-added analysis to create data use and monetization
Information sharing and transmission	Manual processing increases operational costs and expenditures Vulnerability and uncertainties from cyberattacks or system malfunction	Blockchain's distributed attributes allow shared information in the health care network Consensus mechanism with tamper-proof features could reduce security and privacy concerns
Medical supply chain traceability	Uncertainties during handovers among participating parties Poor control in tracking user identities, ownership, and delivery status	Common shared ledger system allows for better transparency and monitoring on supply chain traceability Smart contracts can facilitate notifications of state changes

Blockchain Impacts and the Changing Paradigm on the Health Care Ecosystem

Overview

This study collated blockchain-related literature in the health care industry. While many research efforts highlighted the potential effects of blockchain from a viewpoint of a single firm or industry, we attempted to shed light on its power from a more holistic manner, which focuses on the inclusive health care ecosystem. This changing and evolving paradigm may go through complicated cooperative and competitive challenges with the participating stakeholders. Therefore, from the illustrative case—the IBM blockchain–health care initiative—we elucidated and discussed the potential impacts and complex interactions during the lifecycle of component species or players. Five critical issues, when coevolving with blockchain adoption, are discussed to provide implications for researchers and practitioners.

Health Information Exchange With Interoperability and Integrity

HIE has long been a critical issue when data interoperability is considered; only with an effective information exchange scheme could the true value of health care information be unleashed [50]. Recently, a proliferation of publications and pilots have addressed the issue of medical records and health records. A decentralized scheme using a commonly shared ledger for information sharing offers innovators opportunities to disrupt traditional practice [51]. Health care data has granted blockchain-enabled applications great penetration points into the health care industry. Blockchain-enabled health care information exchange may unleash the power of blockchain to reduce frictions among siloed databases as well the costs from intermediaries [12]. To facilitate information exchange among disparate data systems across individual organizations, the transmission protocols or standards need to be addressed to provide data integrity. In so doing, an important part is the integration of transmission protocols, which mitigates effects of potential missing information and avoids incompatible situations. In addition, blockchain's distributed framework may support cross-system health information usage. However, due to current technological limitations in designing blockchain applications, limited block size could become an issue for extended scalability. Therefore, only critical transactions will be appended on-chain and supporting data access schemes will be necessary for data manipulation. While blockchain could allow interoperability among health systems, incentives for individual stakeholders may become essential when creating beneficial models and supporting sustainable ecosystems are considered. In this regard, blockchain may unlock the true value of interoperability and achieve a higher level of disintermediation.

Digital Identity Management

Traditional identity management has been subject to the limitations of a centralized mechanism, such as security, privacy, and scalability. Centralized identity management is vulnerable to malicious attacks and alterations, thus being prone to theft, counterfeit, and fraud risks [28]. In addition, credentials required to request registration or access to health care services are also prone to misuse or to causing privacy disclosure. Distributed identity management may provide solutions to these limitations with its capabilities of ensuring data integrity and information sharing across different health care systems if deployed on an immutable and distributed network. The distributed model may also solve the duplicate and multi-version identity issues in health care use cases. Due to these features, identity owners may have full control of their unique digital identities and, in turn, enjoy benefits as the stakeholders in a valuable health care ecosystem. This implies that users have become the owners of their health data without the intermediation supported by traditional identity management systems. A higher degree of freedom to access, release, or share medical records has become possible. The blockchain-enabled digital identity is also useful for managing health care supply chain activities, such as the ownership transfer of specific assets. After all, as health care data are normally sensitive and confidential in nature, blockchain identity may leverage its characteristics to grant better security and privacy by reducing manual intervention and operational failures.

Health Care Supply Chain Management

Blockchain's immutable and tamper-proof attributes have granted disruptive innovation to supply chain management. In a health care ecosystem, records of goods, such as drugs, and service flows could be recorded on-chain to provide better logistics visibility and timeliness. The integration of blockchain and medical IoT devices may be the next evolution of blockchain technology in the realm of supply chain management. A large amount of medical data generated by medical devices may be stored across different stakeholder systems. With the aid of blockchain, patient-generated data can be stored off-chain but accessed with permissions preset by blockchain-based smart contracts. In this regard, HIE can become more streamlined without intermediation. Another blockchain use case is for the drug or pharmaceutical supply chain. Typical pain points may occur during handovers across stakeholders. Blockchain provides better transparency on supply chain activities and players may have better control over product and service flows. Moreover, primary concerns also come from the provenance of drug supply. Serious fake and counterfeit drugs have prevailed due to poor authentication and traceability from manufacturing and shipping to delivery. The movement of drugs could be recorded on blockchain to provide better real-time monitoring as well as to cease the distribution of fake drugs. This implies that trackable footprints verified by participating players can help secure drug supply chains.

Medical Research and Data Exploitation

Medical records have long been managed with a centralized approach. However, the disconnected health systems that exist across different clinics or health organizations may hinder further usage of EHRs and EMRs for medical researchers [51]. A considerable number of medical records are stored in paper-based documents or in electronic health systems with poor interoperability. Poor efficiency in health care information exchange and rising costs of administrative processing have locked the true value of medical information. In traditional

circumstances, researchers may have difficulties in acquiring patient data and medical records. This phenomenon may result from *where* and *how* questions. To address data sharing and exploitation among parties and research institutes, researchers have proposed a privacy-preserving model [52,53] and an incentive mechanism [54] during the course of data collection, sharing, and collaborative exploitation. With a shared health care ledger system, researchers may reap benefits from the blockchain-changing paradigm. They may access related data by checking smart contract conditions if the use is permitted by patients. Patients could get rewards or credits from the contributions or payments from researchers by granting different levels of permission, which are coded and stored by smart contracts on blockchain, to release specific data. In sum, blockchain may give control of data access to patients, and researchers could pay for access. In this regard, the traditional pain points for collecting patient data could be resolved in order to facilitate research conduct. Data reconciliation during research design and clinical trials may become easier with a shared medical ledger, thus improving health care and medical treatment.

Automation of Financial Transactions and Insurance Procedures

A lack of trust between health care stakeholders may affect the overall performance of financial transactions in the health care industry, for example, impedance in promoting alternative payment models between payers and providers. When the current reimbursement models and claim procedures were examined, we found hindrances on processing efficiency, transparency, and visibility among ecosystem members. For example, in current insurance fields, multiple middlemen and intermediaries exist throughout the procedures of health insurance policies. In addition, shared information could help insurers seek out better providers and provide verification on the fact if providers meet obligations and contractual terms. Smart contracts may replace efforts on drafting complex and value-based paper contracts and may automate the process of execution of terms or agreements. Through the aid of smart contracts, entities may set up logical process flows when preset conditions regarding health care activities are met. The deployment of smart contracts on decentralized immutable ledger systems could also make payment and claims records visible and render postaudit and review. In this sense, the paramount manipulation on data exchanges and payment transfer between insurers and their stakeholders could become easier and less expensive.

Limitations

In this study, we conducted a literature review to investigate the potential impacts of blockchain-based health care innovations. Along with selected pilot cases, we discussed the positions and promises that blockchain may bring to the health care ecosystem. While researchers and practitioners have high hopes, challenges will be faced before the large-scale adoption of blockchain due to limitations from technical health care service operations and regulatory concerns. Confined by the level of blockchain maturity in various health care subsectors, different use cases and clinical trials need more support from empirical work to report on its real performance. We collated extant research efforts and attempted to shed light on a potential paradigm shift in the future health care ecosystem. Such an endeavor may be subject to uncertainties from the changing environment, technology limitations, or emerging innovations.

Conclusions

This study aims to answer questions on the evolution and development of blockchain technology in health care research and on how stakeholders coevolve in this environment. From the perspective of the business ecosystem, we identified research articles about blockchain-enabled health care and we covered prototype designs and leading pilot cases in recent years. The evolutionary trajectory and interactions among major health care stakeholders may potentially formulate the blockchain-based health care ecosystem. Key players have presented their roles and interacted with one another to go through the life cycle of the business ecosystem. We illustrated their potential and the phenomenon of coevolution within the health care ecosystem. It is noted that while the literature in this field has proliferated recently, mostly regarding proof-of-concept studies, framework propositions, and trial pilots, a careful consideration on embracing such technology still needs to address technical limitations, privacy, mindset, and legal concerns. Our perspective and analysis show that large-scale adoption would need long-term support from health care stakeholders. Future research may devote more efforts to building up evaluation models to provide practical implications for practitioners. Whether feasible business models may sustainably survive in such an ecosystem needs attention from scholars. With a better understanding of how stakeholders coevolve within the ecosystem, players may reap their benefits in a more efficient manner to propel a potential blockchain–health care paradigm shift.

Acknowledgments

This research was supported by the Ministry of Science and Technology, Taiwan, under contract numbers MOST-106-2221-E-005-053-MY3 and MOST-109-2221-E-005-043.

Conflicts of Interest

None declared.

References

1. Chang SE, Chen Y. When blockchain meets supply chain: A systematic literature review on current development and potential applications. IEEE Access 2020;8:62478-62494. [doi: 10.1109/access.2020.2983601]
2. Zhou L, Wang L, Sun Y. MIStore: A blockchain-based medical insurance storage system. J Med Syst 2018 Jul 02;42(8):149 [FREE Full text] [doi: 10.1007/s10916-018-0996-4] [Medline: 29968202]
3. Park YR, Lee E, Na W, Park S, Lee Y, Lee J. Is blockchain technology suitable for managing personal health records? Mixed methods study to test feasibility. J Med Internet Res 2019 Feb 08;21(2):e12533 [FREE Full text] [doi: 10.2196/12533] [Medline: 30735142]
4. Casino F, Dasaklis TK, Patsakis C. A systematic literature review of blockchain-based applications: Current status, classification and open issues. Telematics Inform 2019 Mar;36:55-81. [doi: 10.1016/j.tele.2018.11.006]
5. Halamka JD, Lippman A, Ekblaw A. The potential for blockchain to transform electronic health records. Harv Bus Rev 2017 Mar 03;3(3):2-5 [FREE Full text]
6. Swan M. Blockchain: Blueprint for a New Economy. Sebastopol, CA: O'Reilly Media; 2015.
7. Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. Bitcoin.org. 2008. URL: https://bitcoin.org/bitcoin.pdf [accessed 2019-05-25] [WebCite Cache ID 78dvf0G0K]
8. Szabo N. Smart contracts: Formalizing and securing relationships on public networks. First Monday 1997;2(9):1. [doi: 10.5210/fm.v2i9.548]
9. Buterin V. A next-generation smart contract and decentralized application platform. Ethereum. 2013. URL: https://ethereum.org/en/whitepaper/ [accessed 2019-05-25]
10. Mettler M. Blockchain technology in healthcare: The revolution starts here. Presented at: 18th IEEE International Conference on e-Health Networking, Applications and Services (Healthcom); September 14-16, 2016; Munich, Germany. [doi: 10.1109/healthcom.2016.7749510]
11. Mackey TK, Kuo T, Gummadi B, Clauson KA, Church G, Grishin D, et al. 'Fit-for-purpose?' - Challenges and opportunities for applications of blockchain technology in the future of healthcare. BMC Med 2019 Mar 27;17(1):68 [FREE Full text] [doi: 10.1186/s12916-019-1296-7] [Medline: 30914045]
12. Krawiec RJ, Housman D, White M, Filipova M, Quarre F, Barr D, et al. Blockchain: Opportunities for Health Care. New York, NY: Deloitte; 2016 Aug. URL: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/us-blockchain-opportunities-for-health-care.pdf [accessed 2019-05-25]
13. Moore JF. Predators and prey: A new ecology of competition. Harv Bus Rev 1993;71(3):75-86. [Medline: 10126156]
14. Kshetri N. Blockchain's roles in strengthening cybersecurity and protecting privacy. Telecomm Policy 2017 Nov;41(10):1027-1038. [doi: 10.1016/j.telpol.2017.09.003]
15. Moore JF. The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems. New York, NY: HarperBusiness; 1996.
16. Rong K, Hu G, Hou J, Ma R, Shi Y. Business ecosystem extension: Facilitating the technology substitution. Int J Technol Manag 2013;63(3/4):268. [doi: 10.1504/ijtm.2013.056901]
17. Rong K, Shi Y, Yu J. Nurturing business ecosystems to deal with industry uncertainties. Ind Manag Data Syst 2013 Mar 15;113(3):385-402. [doi: 10.1108/02635571311312677]
18. Esmaeilzadeh P, Mirzaei T. The potential of blockchain technology for health information exchange: Experimental study from patients' perspectives. J Med Internet Res 2019 Jun 20;21(6):e14184 [FREE Full text] [doi: 10.2196/14184] [Medline: 31223119]
19. Khatoon A. A blockchain-based smart contract system for healthcare management. Electronics 2020 Jan 03;9(1):94. [doi: 10.3390/electronics9010094]
20. Azaria A, Ekblaw A, Vieira T, Lippman A. MedRec: Using blockchain for medical data access and permission management. In: Proceedings of the 2nd International Conference on Open and Big Data (OBD). New York, NY: IEEE; 2016 Presented at: 2nd International Conference on Open and Big Data (OBD); August 22-24, 2016; Vienna, Austria. [doi: 10.1109/obd.2016.11]
21. O'Donoghue O, Vazirani AA, Brindley D, Meinert E. Design choices and trade-offs in health care blockchain implementations: Systematic review. J Med Internet Res 2019 May 10;21(5):e12426 [FREE Full text] [doi: 10.2196/12426] [Medline: 31094344]
22. Maslove DM, Klein J, Brohman K, Martin P. Using blockchain technology to manage clinical trials data: A proof-of-concept study. JMIR Med Inform 2018 Dec 21;6(4):e11949 [FREE Full text] [doi: 10.2196/11949] [Medline: 30578196]
23. Hau YS, Lee JM, Park J, Chang MC. Attitudes toward blockchain technology in managing medical information: Survey study. J Med Internet Res 2019 Dec 09;21(12):e15870 [FREE Full text] [doi: 10.2196/15870] [Medline: 31815676]
24. Pandey P, Litoriya R. Implementing healthcare services on a large scale: Challenges and remedies based on blockchain technology. Health Policy Technol 2020 Mar;9(1):69-78. [doi: 10.1016/j.hlpt.2020.01.004]
25. Rabah KVO. Challenges & opportunities for blockchain powered healthcare systems: A review. Mara Res J Med Health Sci 2017;1(1):45-52.
26. Hughes L, Dwivedi YK, Misra SK, Rana NP, Raghavan V, Akella V. Blockchain research, practice and policy: Applications, benefits, limitations, emerging research themes and research agenda. Int J Inf Manage 2019 Dec;49:114-129. [doi: 10.1016/j.ijinfomgt.2019.02.005]

27. Fan K, Wang S, Ren Y, Li H, Yang Y. MedBlock: Efficient and secure medical data sharing via blockchain. J Med Syst 2018 Jun 21;42(8):136. [doi: 10.1007/s10916-018-0993-7] [Medline: 29931655]
28. Bouras MA, Lu Q, Zhang F, Wan Y, Zhang T, Ning H. Distributed ledger technology for eHealth identity privacy: State of the art and future perspective. Sensors (Basel) 2020 Jan 15;20(2):483 [FREE Full text] [doi: 10.3390/s20020483] [Medline: 31952172]
29. Thomas G. A typology for the case study in social science following a review of definition, discourse, and structure. Qual Inq 2011 Jun 17;17(6):511-521. [doi: 10.1177/1077800411409884]
30. Yin RK. Case Study Research and Applications: Design and Methods. 6th edition. Thousand Oaks, CA: SAGE Publications; 2017.
31. Gerring J. What is a case study and what is it good for? Am Polit Sci Rev 2004 Jun 21;98(2):341-354. [doi: 10.1017/s0003055404001182]
32. Baxter P, Jack S. Qualitative case study methodology: Study design and implementation for novice researchers. Qual Rep 2008 Dec 1;13(4):544-559 [FREE Full text]
33. Allison I. IBM, Aetna, PNC explore medical data blockchain for 100 million health plans. CoinDesk. 2019 Jan 24. URL: https://www.coindesk.com/ibm-aetna-pnc-explore-medical-data-blockchain-for-100-million-health-plans [accessed 2020-04-01]
34. Iansiti M, Levien R. Strategy as ecology. Harv Bus Rev 2004 Mar;82(3):68-78, 126. [Medline: 15029791]
35. Iansiti M, Levien R. The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability. Boston, MA: Harvard Business School Press; 2004.
36. Chen Y, Ding S, Xu Z, Zheng H, Yang S. Blockchain-based medical records secure storage and medical service framework. J Med Syst 2018 Nov 22;43(1):5. [doi: 10.1007/s10916-018-1121-4] [Medline: 30467604]
37. Ali MS, Vecchio M, Putra GD, Kanhere SS, Antonelli F. A decentralized peer-to-peer remote health monitoring system. Sensors (Basel) 2020 Mar 16;20(6):1656 [FREE Full text] [doi: 10.3390/s20061656] [Medline: 32188135]
38. McGhin T, Choo KR, Liu CZ, He D. Blockchain in healthcare applications: Research challenges and opportunities. J Netw Comput Appl 2019 Jun;135:62-75. [doi: 10.1016/j.jnca.2019.02.027]
39. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing blockchains for efficient health care: Systematic review. J Med Internet Res 2019 Feb 12;21(2):e12439 [FREE Full text] [doi: 10.2196/12439] [Medline: 30747714]
40. Zhang R, George A, Kim J, Johnson V, Ramesh B. Benefits of blockchain initiatives for value-based care: Proposed framework. J Med Internet Res 2019 Sep 27;21(9):e13595 [FREE Full text] [doi: 10.2196/13595] [Medline: 31573899]
41. Shuaib K, Saleous H, Shuaib K, Zaki N. Blockchains for secure digitized medicine. J Pers Med 2019 Jul 13;9(3):35 [FREE Full text] [doi: 10.3390/jpm9030035] [Medline: 31337080]
42. Guo R, Shi H, Zhao Q, Zheng D. Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems. IEEE Access 2018;6:11676-11686 [FREE Full text] [doi: 10.1109/ACCESS.2018.2801266]
43. Kadam S, Meshram A, Suryawanshi S. Blockchain for healthcare: Privacy preserving medical record. Int J Comput Appl 2019 Jul 24;178(36):5-9 [FREE Full text] [doi: 10.5120/ijca2019919092]
44. Al Omar A, Bhuiyan MZA, Basu A, Kiyomoto S, Rahman MS. Privacy-friendly platform for healthcare data in cloud based on blockchain environment. Future Gener Comput Syst 2019 Jun;95:511-521. [doi: 10.1016/j.future.2018.12.044]
45. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare data gateways: Found healthcare intelligence on blockchain with novel privacy risk control. J Med Syst 2016 Oct;40(10):218. [doi: 10.1007/s10916-016-0574-6] [Medline: 27565509]
46. Aetna, Anthem, Health Care Service Corporation, PNC Bank and IBM announce collaboration to establish blockchain-based ecosystem for the healthcare industry. IBM News Room. 2019 Jan 24. URL: https://tinyurl.com/y3tej8x7
47. Rong K, Lin Y, Shi Y, Yu J. Linking business ecosystem lifecycle with platform strategy: A triple view of technology, application and organisation. Int J Technol Manag 2013;62(1):75. [doi: 10.1504/ijtm.2013.053042]
48. Chukwu E, Garg L. A systematic review of blockchain in healthcare: Frameworks, prototypes, and implementations. IEEE Access 2020;8:21196-21214 [FREE Full text] [doi: 10.1109/access.2020.2969881]
49. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences-A scoping review. Int J Med Inform 2020 Feb;134:104040 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104040] [Medline: 31865055]
50. Walker J, Pan E, Johnston D, Adler-Milstein J, Bates DW, Middleton B. The value of health care information exchange and interoperability. Health Aff (Millwood) 2005;Suppl Web Exclusives:W5-W10. [doi: 10.1377/hlthaff.w5.10] [Medline: 15659453]
51. Xia Q, Sifah EB, Asamoah KO, Gao J, Du X, Guizani M. MeDShare: Trust-less medical data sharing among cloud service providers via blockchain. IEEE Access 2017;5:14757-14767 [FREE Full text] [doi: 10.1109/ACCESS.2017.2730843]
52. Chen X, Ji J, Luo C, Liao W, Li P. When machine learning meets blockchain: A decentralized, privacy-preserving and secure design. In: Proceedings of the 2018 IEEE International Conference on Big Data (Big Data). New York, NY: IEEE; 2018 Presented at: 2018 IEEE International Conference on Big Data (Big Data); December 10-13, 2018; Seattle, WA p. 1178-1187. [doi: 10.1109/bigdata.2018.8622598]
53. Kuo T, Gabriel RA, Ohno-Machado L. Fair compute loads enabled by blockchain: Sharing models by alternating client and server roles. J Am Med Inform Assoc 2019 May 01;26(5):392-403. [doi: 10.1093/jamia/ocy180] [Medline: 30892656]

54. Weng J, Weng J, Zhang J, Li M, Zhang Y, Luo W. DeepChain: Auditable and privacy-preserving deep learning with blockchain-based incentive. IEEE Trans Dependable Secure Comput 2019 Nov 08:1. [doi: 10.1109/tdsc.2019.2952332]

Abbreviations

EHR: electronic health record
EMR: electronic medical record
HIE: health information exchange
IoT: internet of things

Edited by G Eysenbach; submitted 19.04.20; peer-reviewed by JT te Gussinklo, TT Kuo; comments to author 12.06.20; revised version received 30.06.20; accepted 19.07.20; published 31.08.20.

Please cite as:
Chang SE, Chen Y
Blockchain in Health Care Innovation: Literature Review and Case Study From a Business Ecosystem Perspective
J Med Internet Res 2020;22(8):e19480
URL: http://www.jmir.org/2020/8/e19480/
doi:10.2196/19480
PMID:32865501

Original Paper

Technological Capabilities to Assess Digital Excellence in Hospitals in High Performing Health Care Systems: International eDelphi Exercise

Marta Krasuska[1], PhD; Robin Williams[2], PhD; Aziz Sheikh[1], MSc, MD; Bryony Dean Franklin[3,4], PhD; Catherine Heeney[2], PhD; Wendy Lane[5], MA; Hajar Mozaffar[6], PhD; Kathy Mason[5], MSc; Sally Eason[5], MSc; Susan Hinder[2], PhD; Rachel Dunscombe[7,8], MSc; Henry W W Potts[9], PhD; Kathrin Cresswell[1], PhD

[1]Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

[2]Institute for the Study of Science, Technology and Innovation, University of Edinburgh, Edinburgh, United Kingdom

[3]UCL School of Pharmacy, London, United Kingdom

[4]National Institute for Health Research Imperial Patient Safety Translational Research Centre, London, United Kingdom

[5]National Health Services Arden and Greater East Midlands Commissioning Support Unit, Warwick, United Kingdom

[6]Business School, University of Edinburgh, Edinburgh, United Kingdom

[7]Imperial College, London, United Kingdom

[8]KLAS Research's Arch Collaborative, London, United Kingdom

[9]Institute of Health Informatics, University College London, London, United Kingdom

Corresponding Author:
Kathrin Cresswell, PhD
Usher Institute
University of Edinburgh
Old Medical School
Teviot Place
Edinburgh, EH8 9AG
United Kingdom
Phone: 44 131 651 7878
Email: Kathrin.Cresswell@ed.ac.uk

Abstract

Background: Hospitals worldwide are developing ambitious digital transformation programs as part of broader efforts to create digitally advanced health care systems. However, there is as yet no consensus on how best to characterize and assess digital excellence in hospitals.

Objective: Our aim was to develop an international agreement on a defined set of technological capabilities to assess digital excellence in hospitals.

Methods: We conducted a two-stage international modified electronic Delphi (eDelphi) consensus-building exercise, which included a qualitative analysis of free-text responses. In total, 31 international health informatics experts participated, representing clinical, academic, public, and vendor organizations.

Results: We identified 35 technological capabilities that indicate digital excellence in hospitals. These are divided into two categories: (a) capabilities within a hospital (n=20) and (b) capabilities enabling communication with other parts of the health and social care system, and with patients and carers (n=15). The analysis of free-text responses pointed to the importance of nontechnological aspects of digitally enabled change, including social and organizational factors. Examples included an institutional culture characterized by a willingness to transform established ways of working and openness to risk-taking. The availability of a range of skills within digitization teams, including technological, project management and business expertise, and availability of resources to support hospital staff, were also highlighted.

Conclusions: We have identified a set of criteria for assessing digital excellence in hospitals. Our findings highlight the need to broaden the focus from technical functionalities to wider digital transformation capabilities.

(J Med Internet Res 2020;22(8):e17022) doi:10.2196/17022

KEYWORDS

digital excellence; digital maturity; Delphi technique; hospitals, eHealth

Introduction

It is now widely recognized that health information technology (HIT) has significant potential to transform health care systems and support continuous quality improvement efforts [1]. With growing recognition of this potential has come a strong international drive towards creating digitally advanced health care organizations. To this end, hospitals worldwide are now implementing ambitious digital transformation programs [2,3].

There are various ways to conceptualize and measure digital excellence in health care [4,5]. These approaches vary in scope from highly specialized models, focusing on a specific technological subsystem [6] to those assessing digital transformation across an entire hospital, and others encompassing the wider integrated health and care ecosystem [7]. The origin of these models is also diverse, including international health care industry organizations such as the Healthcare Information and Management System Society (HIMSS) Analytics [8], national health care providers [9], and academic groups [10]. Common to all existing frameworks is the concept of digital transformation progressing towards advanced levels of digital maturity through a defined set of stages associated with different technological capabilities. Perhaps the best known of these is the HIMSS Analytics Electronic Medical Record Adoption Model (HIMSS EMRAM; Textbox 1). Policymakers and health care organizations commonly use these frameworks for baseline assessments of current levels of digital maturity and as a roadmap for a desired future state of maturity. As such, these frameworks actively shape the direction of digital transformation [11].

Despite substantive worldwide efforts to promote digital excellence, there is no consensus on how to conceptualize it, what capabilities characterize a digitally excellent hospital, and how to best measure progress in a changing environment [12]. New models are beginning to emerge that acknowledge the importance of locally formed priorities and the changing nature of what constitutes digital excellence over time. In this study, we sought to identify and reach consensus on a defined set of internationally relevant technological capabilities for hospitals in order to address current gaps in approaches to conceptualizing and assessing digital excellence.

Textbox 1. HIMSS Analytics Electronic Medical Record Adoption Model (EMRAM).

The HIMSS EMRAM classification evaluates the extent to which electronic medical records (EMRs) have been adopted within a hospital over eight progressive stages (Levels 0-7).

A hospital's digital transformation begins at Level 0, in which no electronic laboratory, pharmacy, or radiology systems are installed. The hospital then moves through Levels 1-7 by progressive adoption of various aspects of EMRs. These include limited ancillary departmental systems (Level 1), and adoption across an increasing number of hospital departments (Levels 1-6), culminating in a virtually paperless environment with complex EMRs implemented in over 90% of the hospital's departments (Level 7).

A hospital can be assessed against the HIMSS classification to establish its current HIMSS Level, which in turn highlights what further technological capabilities the hospital needs to reach the next level of the HIMSS classification. HIMSS Level 7 is often considered a 'gold standard' for the digitization of hospitals and an aspirational endpoint guiding the design of a hospital's digital strategy.

Methods

Study background

This work was conducted as part of a national evaluation of the National Health Service (NHS) Global Digital Exemplar (GDE) Programme in England [13]. The GDE Programme aims to create a cohort of digitally excellent hospitals ("Exemplars"), which are expected to share their experiences and learning with other health care providers and contribute toward creating a national learning health care ecosystem [3]. We followed the reporting recommendations for Delphi studies outlined by Boulkedid and colleagues [14].

Overview of the Delphi Method

The Delphi technique is a structured process that involves presenting a series of surveys to a group of experts to seek their agreement on statements relating to a particular issue [15]. An initial survey informs the development of a second survey, which is returned to the experts, who are asked to reconsider their initial judgment in light of feedback from the first round. Consecutive rounds are carried out until consensus is reached [16]. The key strength of the Delphi method is that it supports consensus development in an area of uncertainty or limited empirical evidence [17]. The method allows drawing on a wide range of experts' knowledge and experiences. The feedback offered to participating experts between rounds has the potential to widen participants' outlooks and stimulate new ideas that can be expressed in subsequent rounds [17]. The anonymity offered by the method (the identity of experts and their contributions are not known to other experts taking part in the Delphi exercise or the public) also has potential to facilitate disclosing opinions that may be underrepresented or not expressed in other forms of consensus-building approaches where participants are aware of each other's identity. The potential risks associated with the Delphi approach include lack of accountability for anonymously presented views, and risks generally associated with consensus-building approaches such as group-think and lack of diversity of views represented in the outcomes [17]. After considering its strengths and weaknesses, we decided that the Delphi method would be an appropriate approach for addressing our aim of developing international agreement on a defined set of technological capabilities to assess digital excellence in hospitals in high performing health care systems.

To ensure a reasonable geographical spread of experts and relatively prompt completion of the Delphi exercise, we used a modified Delphi approach utilizing electronic communication with experts [18]. The modified electronic Delphi (eDelphi) technique has been widely used in health care and medical informatics, for example, to establish a set of readiness criteria for HIT innovations [19], to define key performance indicators to benchmark hospital information systems [20], and to identify ways to improve the delivery of medication alerts within computerized physician order entry (CPOE) systems [21].

The study took place between July 2018 and January 2019. Ethics approval was obtained from an Institutional Review Board at the School of Social and Political Science at The University of Edinburgh, United Kingdom. The Qualtrics platform was used to develop an online survey and collect data. SPSS Version 24 was used to conduct quantitative analyses, and NVivo Version 12 was used to analyze free-text responses.

The eDelphi Process

Identification of Experts

We identified a diverse group of international experts in the field of health informatics from leading clinical, academic, public, and vendor organizations, aiming for maximum variation in terms of geographical location, background (eg, academic, clinical, vendors), and gender. Our eligibility criteria included providing senior-level leadership in the field of health informatics and affiliation with a leading clinical, academic, public, or vendor organization. Experts were identified through the research team's international academic and professional networks.

Development and Piloting of Candidate Capabilities

Our focus was to ensure that the proposed list of candidate technological capabilities forming the basis of the eDelphi exercise drew on ongoing work relating to digital excellence in hospitals. We used NHS England's Digital Maturity Index as a basis for constructing the initial list [9]. The index was developed in 2013 based on HIMSS EMRAM but included additional dimensions of interoperability, technological readiness, and infrastructure. We then piloted this initial list with three clinical academics, which resulted in some changes to the wording to improve clarity.

Round 1 of the eDelphi

Identified experts received an invitation email explaining the rationale for and aim of the study, the reason they were invited, what taking part would involve, and a personalized link to the Round 1 online survey. Experts were asked to follow the link if they wished to participate. We sent up to three follow-up emails at 2-3 week intervals to those who did not complete the survey following the initial invitation.

The opening page of the online survey for Round 1 contained further details of the study and a link to a participant information sheet. We obtained informed consent from each participant before the start of the survey. Participants were given the option to receive a summary of the findings once the study was completed. The main body of the online survey consisted of the list of proposed technological capabilities identified in the piloting stage. Participants were asked to rate how much they agreed that each proposed capability could be used to assess the level of digital excellence in hospitals, using a scale ranging from 1 (strongly agree) to 9 (strongly disagree). Experts were also encouraged to comment on each capability to suggest more appropriate wording, merge, split, or remove the capability, or to add other comments. Finally, we asked experts for suggestions of any additional capabilities they wished to add to the list.

Analysis of Data From Round 1

The purpose of analysis at this stage was to produce material for Round 2 of the eDelphi exercise. First, we revised the list of proposed capabilities based on participants' comments from Round 1, changing the wording and dividing some capabilities into two or more capabilities to improve precision and clarity. We also added capabilities proposed in Round 1 to the revised list. As the majority of candidate capabilities were revised following insights from Round 1, we decided not to remove any capabilities at this stage. We further produced a feedback document that contained a summary of experts' comments and descriptive statistics from Round 1 for each capability.

Round 2 of the eDelphi

Experts who completed Round 1 were invited to take part in Round 2 via an invitation email as before. Again, we sent up to three reminders at 2-3 week intervals to those who did not complete Round 2 following the initial email. An online version of the feedback document from Round 1 was also provided. Experts were given their score for each revised capability and asked if they wished to reconsider given the feedback from Round 1. If they replied "Yes," they were given an option to amend their assessment using the same scale as in Round 1. Experts were asked to rate how much they agreed that each proposed new capability could be used to assess the level of digital excellence in hospitals on the same scale as for other capabilities. Experts were also able to comment on each capability, as above.

Analysis of Data From Round 2 and Definition of Consensus

Analysis following Round 2 aimed to identify any consensus on the capabilities and determine whether an additional round was needed. We defined consensus a priori as 70% agreement among experts that a specific capability should be included [17]. In other words, to be included, at least 70% of experts needed to "agree" or "strongly agree" to the appropriateness of the capability to define digital excellence in hospitals. After calculating the percentage of experts agreeing or strongly agreeing that the capability should be included, we removed all capabilities for which fewer than 70% agreed (see Multimedia Appendix 1), to produce the final list of capabilities.

Qualitative Data Collection and Analysis

To supplement the consensus-building exercise with additional insights, we incorporated several open-ended questions into the surveys, for which experts were able to provide free-text responses. In Round 1, we gave one current definition of digital maturity proposed by the MIT Sloan Management Review and asked experts to comment on this definition in the health care context [22]. We also asked experts to comment on the role of

nontechnological factors (eg, strategy, workforce, culture) in the context of digital excellence in hospitals. Some feedback from Round 1 suggested that the initial list of capabilities focused too narrowly on the internal operations of hospitals. In Round 2, we therefore asked experts to comment on (a) conceptualization of digital excellence in hospitals in the context of the broader health care ecosystem; and (b) digital excellence in the context of a patient-centered health care perspective. We analyzed data from all free-text entries using thematic analysis to identify key themes of digital excellence in health care [23]. The qualitative data were initially analyzed by one researcher (MK). The resulting coding framework and the analyses were then reviewed and expanded by the wider team (KC, RW, AS). Researchers from a variety of backgrounds (eg, social sciences, public health, informatics) were involved in the analysis of the qualitative data, and diverging findings and viewpoints were discussed in detail in order to minimize the risk of bias.

Results

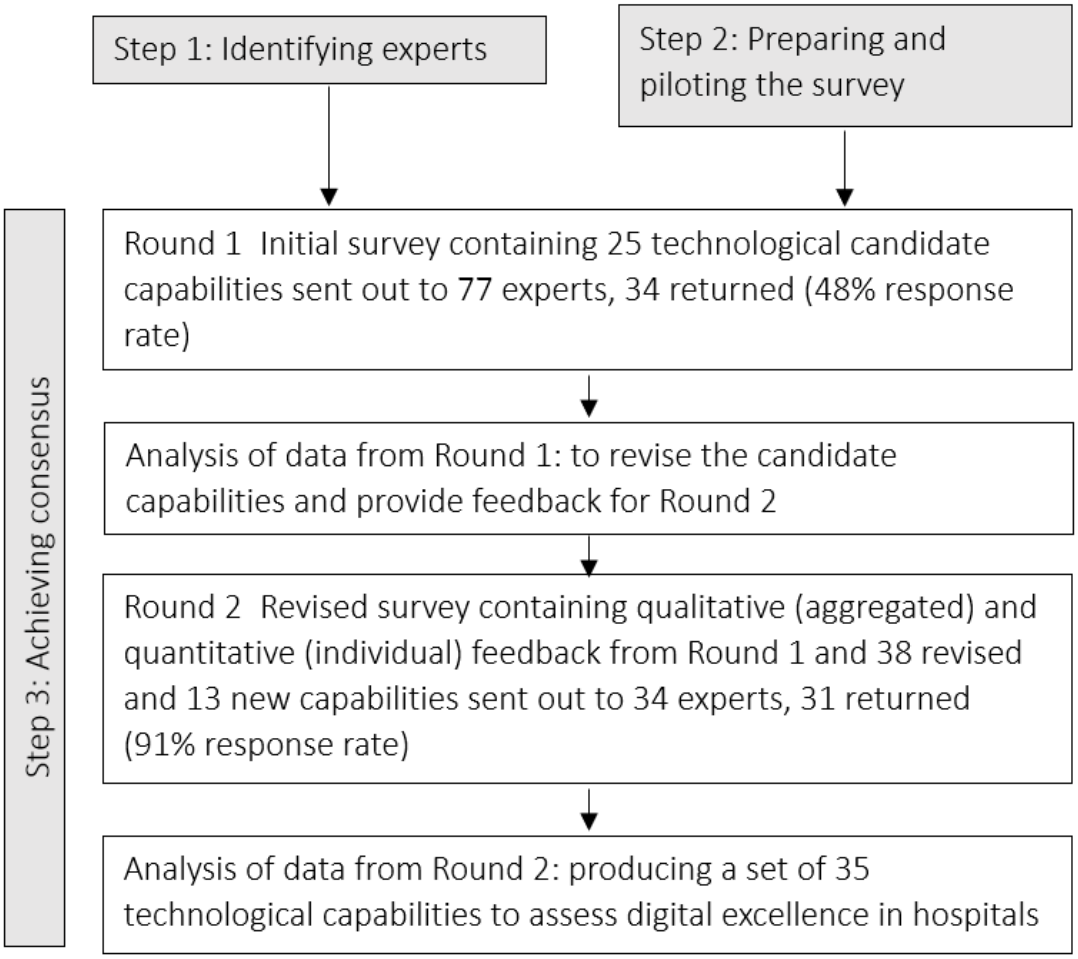
eDelphi Process and Expert Characteristics

In total, 77 experts were invited. Of these, 34 agreed to take part and completed Round 1 (44% response rate), and 31 of the 34 completed Round 2 (91% response rate). Table 1 describes the characteristics of the 31 experts who took part in both rounds; Figure 1 outlines the steps involved.

Table 1. Expert characteristics.

Characteristic	Experts approached to take part (n=77), n (%)	Experts who took part in both rounds (n=31), n (%)
Sector		
Clinical	17 (22)	6 (19)
Academia	35 (45)	17 (55)
Policy	5 (6)	1 (3)
Vendor	20 (26)	7 (23)
Region		
North America	21 (27)	12 (39)
United States	21 (27)	12 (39)
South America	2 (3)	1 (3)
Brazil	2 (3)	1 (3)
Europe	46 (60)	15 (48)
United Kingdom	23 (30)	6 (19)
Spain	6 (8)	4 (13)
Norway	4 (5)	2 (7)
Denmark	3 (4)	1 (3)
Sweden	2 (3)	1 (3)
Slovenia	2 (3)	1 (3)
Belgium	1 (1.5)	0 (0)
Estonia	1 (1.5)	0 (0)
Finland	1 (1.5)	0 (0)
Russia	1 (1.5)	0 (0)
Austria	1 (1.5)	0 (0)
Germany	1 (1.5)	0 (0)
Australasia	6 (8)	3 (10)
Australia	6 (8)	3 (10)
Middle East	2 (3)	0 (0)
Saudi Arabia	1 (1.5)	0 (0)
Israel	1 (1.5)	0 (0)
Gender		
Female	11 (14)	4 (13)
Male	66 (86)	27 (87)

Figure 1. Flow diagram for the eDelphi exercise.



Digital Excellence in Hospitals

Experts identified 35 technological capabilities that were judged to characterize digital excellence in hospitals (Tables 2-4). The technological capabilities fell into two categories: (a) capabilities within a hospital, and (b) communication with other parts of the health care system, and with patients and carers. The need to distinguish between capabilities within hospitals and those relating to the broader context of the health care ecosystem was emphasized in free-text comments, for example:

> *There is an important assessment on where enterprises (eg, hospitals) are, versus where those enterprises sit in an ecosystem and how they interact with those wider ecosystems.* [Vendor]

Technological Capabilities Within a Hospital

The largest category, technological capabilities within a hospital, included 20 items (Table 2), including technologies to promote the appropriate use and administration of medication, capabilities to capture structured and unstructured data, and the ability to integrate new advanced technologies (eg, natural language processing) within existing systems.

The most substantial proportion of capabilities within this category (five of 20) related to medicines management. The highest level of agreement was applied to a capability related to closed-loop electronic medicines management stating that it should be included as a marker of digital excellence (90% agreement), this was closely followed by capabilities relating to the effective capture of clinical data.

Experts proposed four new capabilities in Round 1 (Capabilities 10, 12, 13, and 15; Table 2). These were concerned with advancements in electronic medical records, electronic prescribing and medicines administration systems to improve user experience (for example Capability 10 ‘A single list of all medication for one patient is available’), and integration of new technologies and analytical approaches into existing systems (for example Capability 15 ‘Use of machine learning and adding third party programs through Application Programming Interfaces’).

Table 2. Technological capabilities within hospitals.

Agreed list of capabilities	"Strongly agreed" and "agreed"[a] (%)	Number of experts who agreed (n=31)	Median	IQR[b]
1. Closed-loop electronic medicines management and optimization (electronic prescribing with technology-assisted identification of both patient and medication, eg, bar codes or RFID[c] tags)	90	28	1	1-2
2. Effective mechanisms to collect and record complete, accurate and high-quality patient/clinical data	87	27	2	1-2
3. Structured data (records, assessments, and plans) captured digitally at the point of care	87	27	1	1-2
4. Orders (eg, lab tests) are ordered, and results reported in a coded form (ie, using standard compendiums and international vocabulary standards, including dm+d[d], and acknowledged electronically in the system	84	26	1	1-2
5. Effective mechanisms to review and improve the quality of patient/clinical data	84	26	2	1-2
6. Flexible digital systems guiding clinicians along evidence-based, person-specific, clinical pathways	81	25	2	1-2
7. Unstructured data (eg, notes, free text) captured at the point of care when appropriate	81	25	2	1-2
8. Person reading/acting on the results acknowledges this electronically in the system	81	25	1	1-2
9. Cybersecurity strategy and continuity processes in place and implemented effectively	81	25	1	1-2
10. A single list of all medication for one patient is available[e]	81	25	1	1-2
11. Management intelligence through digital health data	81	25	1.5	1-2
12. Reducing the need for duplicate entry of patient data to near-zero[e]	81	25	2	1-2
13. Third-party tools can be added through Application Programming Interfaces[e]	81	25	2	1-2
14. Advanced clinical decision support (eg, integrated with lab data, diagnosis codes) with alerts that are both sensitive and specific and therefore less likely to result in alert fatigue	77	24	2	1-2
15. Use of machine learning and automation when appropriate (eg, analysis of radiology images)[e]	77	24	2	1-2
16. Clinical intelligence through digital health data	77	24	1	1-2
17. The ability to monitor outcome data for modifying clinical pathways based on digital tools and services	77	24	2	1-2
18. Open Application Programming Interfaces allowing different software components to interact	74	23	1	1-3
19. Supporting end-to-end redesign and improvement of clinical pathways based on digital tools and services	74	23	2	1-3
20. Advanced analytics capability to support the move from reactive to proactive/predictive models of care	74	23	2	1-3

[a]Experts scored each capability using a scale ranging from "1" (strongly agree) to "9" (strongly disagree).

[b]IQR: Interquartile range.

[c]RFID: Radio Frequency Identification.

[d]dm+d: Dictionary of Medicines and Devices.

[e]New capabilities suggested by experts in Round 1 of the eDelphi.

Communication With Other Parts of the Health Care System, Patients, and Carers

This category was related to enabling the exchange of information and communication beyond an individual hospital setting, including communication with other parts of health and social care systems (Table 3), and communication with patients and carers (Table 4). In total, this category comprised 15 capabilities, of which ten related to communication with other parts of health care systems and five to communication with patients and carers. Experts proposed two new capabilities, including the use of a unique patient identifier (Capability 23, Table 3), and the ability to exchange information with other systems based on shared standards (Capability 6, Table 3).

Table 3. Technological capabilities related to communication with other parts of the health and social care system.

Agreed list of capabilities	"Strongly agreed" and "agreed"[a] (%)	Number of experts who agreed, (n=31)	Median	IQR[b]
1. Exchange of prescription information in a structured way within and between organizations and sectors	87	27	1	1-2
2. Local sharing of relevant data across the local health care ecosystem facilitated by interfacing or interoperability of electronic systems	84	26	1	1-2
3. A unique patient identifier used across the health care system[c]	84	26	1	1-2
4. Data analysis at scale and use of insights to deliver targeted care for high-risk and high-use groups of patients (eg, diabetes, chronic obstructive pulmonary disease, asthma) across a population or area	84	26	2	1-2
5. Using digital systems to enable the seamless (through interfaces/integration) flow and use of information/data across organizational boundaries within a local health care ecosystem	81	25	1	1-2
6. Ability to interoperate with other standard-based external systems[c]	81	25	2	1-2
7. Referrals within and between hospitals are always managed electronically	77	24	1	1-2
8. Ability to send communications to primary care and social care through a variety of media	77	24	2	1-2
9. Ability to produce data for audits and other reports based on the routine collection of complete, accurate, and quality data	74	23	2	1-3
10. Discharge to primary care and community is always managed electronically	71	22	1	1-2

[a]Experts rated how much they agree that the capability can be used to assess the level of digital excellence in hospitals on a scale from "1" (strongly agree) to "9" (strongly disagree).

[b]IQR: Interquartile range.

[c]New capabilities suggested by experts in Round 1 of the eDelphi.

Table 4. Technological capabilities related to communication with patients and carers.

Agreed list of capabilities	"Strongly agreed" and "agreed"[a] (%)	Number of experts who agreed, (n=31)	Median	IQR[b]
1. Records, assessments, and plans shared digitally and easily accessible to patients and carers to enter and amend the data securely and confidentially	90	28	1	1-2
2. Records, assessments, and plans shared digitally and easily accessible to patients and carers to view the data securely and confidentially	87	27	1	1-2
3. Ability to receive communications from patients and carers through a variety of media	74	23	2	1-3
4. Ability to send communications to patients and carers through a variety of media	74	23	2	1-3
5. Using mobile technologies to support the delivery of care outside traditional settings and closer to home	71	22	2	1-3

[a]Experts rated how much they agree that the capability can be used to assess the level of digital excellence in hospitals on a scale from "1" (strongly agree) to "9" (strongly disagree).

[b]IQR: Interquartile range.

Broader Aspects of Digitally Enabled Change: Culture, Skills, and Strategy

Free text responses emphasized that technologies should not be viewed in isolation and that social and organizational factors were crucial for digital transformation (Table 5).

Organizational culture, characterized by a willingness to transform established ways of working and an openness to risk-taking, was frequently mentioned as key to promoting digital transformation:

> *It is important to have a culture where individuals are prepared to change their ways of working and take some risks with an understanding of the overall good that will be achieved.* [Policy expert]

Experts also frequently mentioned the need for a diverse set of interdisciplinary skills supporting these transformations. Here, participants called for a range of technological, project management, and business expertise:

> *Digital health is a diverse, interdisciplinary sector, something that is reflected in the skills required in*

> *the field, ranging from higher-level computing, such as software development and software engineering to project management and business-related skills.* [Vendor]

Experts further highlighted the need for sufficient resources for the existing staff base, and their emerging training needs to support digital transformation:

> *A digital agenda cannot be delivered without sufficient staff, who are experienced and well trained within the digital team.* [Clinician]

Table 5. Social and organizational factors contributing to digital maturity.

Factor	Description
Organizational culture	• Willingness to face the new, change the way of thinking, and to take risks • Culture of allowing innovations • Understanding of change management • Culture free of bullying and harassment • Leadership to support digital transformation
Workforce	• Skills within the digital team: software development, software engineering, project management, business-related skills • Skills across the hospital's workforce: the ability to perform one's role using digital tools • Professionalization of health informatics
Strategy	• Putting clinical benefits at the center of clinical strategy • Aligning digital strategy with the overall strategy of the hospital • Support of the digital agenda from the hospital's board

Discussion

Summary of Findings

We have established consensus on a discrete set of internationally relevant technological capabilities to indicate digital excellence in hospitals. Engaging international experts in a transparent process represented by the Delphi technique, allowed us to develop a detailed, multi-axial mapping of digital excellence, which may be used by decision-makers to inform digital transformation strategy and evaluation. The outcomes of this eDelphi process mark a significant departure from existing tools such as HIMSS EMRAM and the NHS Digital Maturity Index [8,9]. First, our results point to a shift away from the description of purely technological functionalities towards digital transformation capabilities and highlight a need to be cognizant of cultural and strategic factors, such as skills and resources, to support the digitally enabled transformation of health care. Second, our findings indicate that the concept of digital excellence is moving beyond the physical boundaries of acute hospitals. Thus, once a certain level of digitization and data sharing is achieved within hospitals, strategic direction needs to shift towards sharing data and integration across local/regional/national ecosystems that encompass primary and social care providers and enable patient self-management.

Strengths and Limitations

This study is the first attempt to achieve international consensus on a defined set of technological capabilities to indicate digital excellence in hospital settings. We recruited a relatively large sample of international experts from a variety of countries and achieved a good overall response rate. There is considerable variation in the number of experts involved in Delphi studies, and no consensus exists as to what constitutes an optimal number of experts [24,25]. However, the available evidence indicates that the Delphi technique produces reliable outcomes for sample sizes of 20 or above, and that increasing the number of experts above that number does not significantly change the outcomes [25]. We do, however, acknowledge that including a larger number of experts in this eDelphi could have provided valuable additional insights. Our participants reached consensus after two rounds, and we decided not to conduct further rounds. Evidence in the literature indicates that Delphi studies consisting of two to three rounds are preferred [17]. Additionally, in the case of busy experts and clinicians (as was the case for this work), response exhaustion occurs after two rounds resulting in limited new insights occurring in consequent rounds [26].

Our identified criteria have the potential to be used internationally, although our sampling reflects a certain subset of predominantly English-speaking economically developed countries and, therefore, high-performing health care systems. Our sample also exhibits a strong gender bias, with 27 of 31 Round 2 participants being men. This bias may mirror a broader gender bias present across the digital health leadership community but is likely to affect our findings and conclusions.

A more general concern is that the eDelphi process itself has some limitations. It may, to some extent, force consensus and reinforce dominant views (although controlled anonymized feedback should minimize normative pressure to align views) [16,27]. The addition of a qualitative component may have helped mitigate against this risk by allowing dissenting voices to be heard and by allowing discussion of the complexity of the context in which attempts to measure excellence are taking place. While we originally intended to examine differences between groups of experts (eg, experts from different regions and commercial and public sector) with regards to ratings of different capabilities, we have not been able to conduct those analyses meaningfully given the sample size and the small overall variance in our data.

Integration of Findings With the Current Literature

Most existing models seeking to define digital excellence in health care settings are hospital-focused and stage-based [28]. Our findings question the appropriateness of such one-directional models, which assume that organizations and people within them progress towards increasingly advanced levels of maturity through a predefined set of consecutive stages associated with certain characteristics. Indeed, numerous accounts of organizations leapfrogging undermine the tacit idea of stage-based models going through a fixed sequence of stages [29]. Stage-based models are popular, perhaps because they promise a simple way to measure progress, but give little scope for health systems and individual organizations to articulate their local priorities [12]. Furthermore, 'one size fits all' assessment criteria enforce a common standard even under circumstances where achieving this may not be appropriate or impose disproportionate costs. Maturity models from outside the health sector can offer insights into possible ways of addressing some of the limitations of the currently dominant, linear, one-directional approaches in health care. For example, the Deloitte Maturity Model [30] developed primarily to meet the needs of the telecommunications industry, proposes to assess digital maturity using 179 digital capabilities grouped into five categories representing the core dimensions for the functioning of an organization (eg, 'customers', 'operations'). An organization can choose which capabilities from which dimension to develop and in what order, based on its local priorities. This flexibility, in turn, allows for articulating local needs and following specific digital journey appropriate for the local context. Our findings support the increasing recognition that particular organizational and cultural environments of health systems are important factors when considering digital excellence [10,31].

The existing literature predominantly places large acute hospitals at the center of discussions of digital excellence [32]. Our study highlights how the entrenched focus on acute hospitals can draw attention away from integration across the health care ecosystem—even though integrated, patient-centered care has become a key component of current health policies internationally [33]. In line with this, HIMSS Analytics recently developed the Continuity of Care Maturity Model (CCMM) [34]. CCMM, like EMRAM, comprises seven stages and includes dimensions such as interoperability, exchange of information, coordination of care, patient involvement, and use of HIT to optimize clinical and financial outcomes. However, this extended HIMSS classification focuses on the individual health care provider rather than considering the entire health care system, and it remains a stage-based approach. It also remains mainly relevant to the hospital-centric United States context.

There is only limited evidence that meeting all criteria in any index of digital excellence leads to improved quality, safety, or efficiency outcomes, although some functionalities such as clinical decision support systems have been shown to improve practitioner performance [35,36]. At the same time renewing digital infrastructure to meet the ever-expanding requirements for each progressive stage of digital maturity indexes such as HIMSS EMRAM is costly. Many hospitals might choose not to pursue the route of advancing across stages of digital maturity due to high costs combined with insufficient evidence of desired return on investment. Thus, although digital excellence indices are commonly viewed as a proxy measure for improvement in efficiency and safety, there is limited evidence that adoption of these models will *per se* deliver such improvements [37].

Implications for Research, Policy, and Practice

The identified technological capabilities have the potential to serve as a practical means to baseline and measure digital progress within acute hospital settings and their wider health care context. They can also promote international comparisons. Future work should focus on developing an assessment tool based on these identified capabilities. This needs to include establishing a scoring mechanism and weighting criteria for the capabilities comprising the tool and demonstrating the tool's reliability and validity, including responsiveness to change and discriminatory properties. This work could be facilitated by using the set of capabilities to assess digital excellence in selected hospitals worldwide. It might also be valuable to investigate whether our set of capabilities could be further divided into more detailed categories to provide a better understanding of dimensions constituting digital excellence in hospitals. Additionally, there is a need for further efforts aiming to develop an agreement on what constitutes digital excellence for health care providers that includes views of additional stakeholders such as politicians, decision makers, and authorities.

Conclusions

We have identified an internationally agreed defined set of technological capabilities that constitute digital excellence in hospitals. Our study also foregrounds the managerial and cultural skills necessary for successful, digitally enabled change. Finally, it highlights the need to address integrating digital capabilities across the wider health and care ecosystem to deliver safe, high-quality, and patient-centered care. Digital implementation strategies and indicators need to be positioned within this wider health system landscape to enable and foster transformational change in health care internationally.

Acknowledgments

We gratefully acknowledge the input of the wider GDE evaluation team and the Steering Group of this evaluation, especially Ann Slee (Associated CCIO Medicines, NHSX). We would also like to thank the experts who completed the study surveys and made this work possible.

This article is based on a program of independent research funded by NHS England. The views expressed are those of the author(s) and not necessarily those of the NHS, NHS England, NHSX or NHS Digital. This work was also supported by the National

Institute for Health Research (NIHR) Imperial Patient Safety Translational Research Centre and Health Data Research UK. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Care.

Authors' Contributions

MK conducted the study and drafted the paper. KC, RW, and AS conceived and designed the study and commented on multiple versions of the draft. BDF, CH, WL, HM, KM, SE, SH, RD, and HWWP commented on multiple versions of the draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Capabilities that did not meet the inclusion criteria.
[DOCX File , 16 KB - jmir_v22i8e17022_app1.docx]

References

1. Wachter RM. Making IT Work: Harnessing the Power of Health Information Technology to Improve Care in England Report of the National Advisory Group on Health Information Technology in England.: Department of Health and Social Care; 2016 Sep 07. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf [accessed 2019-08-15]
2. Blumenthal D. Launching HITECH. N Engl J Med 2010 Feb 04;362(5):382-385. [doi: 10.1056/NEJMp0912825] [Medline: 20042745]
3. Global Digital Exemplars. NHS England. URL: https://www.england.nhs.uk/digitaltechnology/connecteddigitalsystems/exemplars/ [accessed 2019-08-15]
4. Carvalho JV, Rocha A, Abreu A. Maturity Models of Healthcare Information Systems and Technologies: a Literature Review. J Med Syst 2016 Jun;40(6):131. [doi: 10.1007/s10916-016-0486-5] [Medline: 27083575]
5. Gomes J, Romão M. Information System Maturity Models in Healthcare. J Med Syst 2018 Oct 16;42(12):235. [doi: 10.1007/s10916-018-1097-0] [Medline: 30327955]
6. van de Wetering R, Batenburg R. A PACS maturity model: a systematic meta-analytic review on maturation and evolvability of PACS in the hospital enterprise. Int J Med Inform 2009 Feb;78(2):127-140. [doi: 10.1016/j.ijmedinf.2008.06.010] [Medline: 18752988]
7. Grooten L, Borgermans L, Vrijhoef HJ. An Instrument to Measure Maturity of Integrated Care: A First Validation Study. Int J Integr Care 2018 Jan 25;18(1):10. [doi: 10.5334/ijic.3063] [Medline: 29588644]
8. HIMMS Electronic Medical Record Aadoption Model (EMRAM). Healtcare Information and Management Systems Society. URL: https://www.himssanalytics.org/emram [accessed 2019-09-12]
9. Digital Maturity Assessment. NHS England. URL: https://www.england.nhs.uk/digitaltechnology/connecteddigitalsystems/maturity-index/ [accessed 2019-08-16]
10. Carvalho JV, Rocha A, van de Wetering R, Abreu A. A Maturity model for hospital information systems. Journal of Business Research 2019 Jan;94:388-399. [doi: 10.1016/j.jbusres.2017.12.012]
11. Hughes O. Will Smart: England to get first HIMSS 7 hospitals before year's end. Digital Health. 2018 May 17. URL: https://www.digitalhealth.net/2018/05/will-smart-first-himss-7-hospital-before-years-end/ [accessed 2019-09-10]
12. Cresswell K, Sheikh A, Krasuska M, Heeney C, Franklin BD, Lane W, et al. Reconceptualising the digital maturity of health systems. The Lancet Digital Health 2019 Sep;1(5):e200-e201. [doi: 10.1016/s2589-7500(19)30083-4]
13. Global Digital Exemplar (GDE) Programme Evaluation. University of Edinburgh. URL: https://www.ed.ac.uk/usher/digital-exemplars [accessed 2019-06-03]
14. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One 2011;6(6):e20476 [FREE Full text] [doi: 10.1371/journal.pone.0020476] [Medline: 21694759]
15. Dalkey N. The Delphi Method: An experimental study of group opinion. Futures 1969 Sep;1(5):408-426. [doi: 10.1016/s0016-3287(69)80025-x]
16. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000 Oct;32(4):1008-1015. [Medline: 11095242]
17. Powell C. The Delphi technique: myths and realities. J Adv Nurs 2003 Feb;41(4):376-382. [Medline: 12581103]
18. Keeney S, Hasson F, McKenna H. Consulting the oracle: ten lessons from using the Delphi technique in nursing research. J Adv Nurs 2006 Jan;53(2):205-212. [doi: 10.1111/j.1365-2648.2006.03716.x] [Medline: 16422719]
19. Snyder-Halpern R. Indicators of organizational readiness for clinical information technology/systems innovation: a Delphi study. Int J Med Inform 2001 Oct;63(3):179-204. [Medline: 11502432]

20. Hübner-Bloder G, Ammenwerth E. Key performance indicators to benchmark hospital information systems - a delphi study. Methods Inf Med 2009;48(6):508-518. [doi: 10.3414/ME09-01-0044] [Medline: 19893854]
21. Riedmann D, Jung M, Hackl WO, Ammenwerth E. How to improve the delivery of medication alerts within computerized physician order entry systems: an international Delphi study. J Am Med Inform Assoc 2011;18(6):760-766 [FREE Full text] [doi: 10.1136/amiajnl-2010-000006] [Medline: 21697293]
22. Kane G, Palmer D, Phillips A, Kiron D, Buckley N. Achieving Digital Maturity: Adapting Your Company to a Changing World. MIT Sloan Management Review. 2017 Jul 13. URL: https://sloanreview.mit.edu/projects/achieving-digital-maturity/ [accessed 2019-05-03]
23. Ritchie J, Lewis J, McNaughton N, Ormston R, editors. Qualitative Research Practice: A Guide for Social Science Students and Researchers. London: SAGE Publications Ltd; 2013.
24. Williams PL, Webb C. The Delphi technique: a methodological discussion. J Adv Nurs 1994 Jan;19(1):180-186. [Medline: 8138622]
25. Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. BMC Med Res Methodol 2005 Dec 01;5:37 [FREE Full text] [doi: 10.1186/1471-2288-5-37] [Medline: 16321161]
26. McKenna HP. The essential elements of a practitioners' nursing model: a survey of psychiatric nurse managers. J Adv Nurs 1994 May;19(5):870-877. [doi: 10.1111/j.1365-2648.1994.tb01163.x] [Medline: 8056915]
27. Goodman CM. The Delphi technique: a critique. J Adv Nurs 1987 Nov;12(6):729-734. [doi: 10.1111/j.1365-2648.1987.tb01376.x] [Medline: 3320139]
28. Mettler T, Pinto R. Evolutionary paths and influencing factors towards digital maturity: An analysis of the status quo in Swiss hospitals. Technological Forecasting and Social Change 2018 Aug;133:104-117. [doi: 10.1016/j.techfore.2018.03.009]
29. Kharrazi H, Gonzalez CP, Lowe KB, Huerta TR, Ford EW. Forecasting the Maturation of Electronic Health Record Functions Among US Hospitals: Retrospective Analysis and Predictive Model. J Med Internet Res 2018 Aug 07;20(8):e10458 [FREE Full text] [doi: 10.2196/10458] [Medline: 30087090]
30. Digital Maturity Model: Achieving digital maturity to drive growth. Deloitte Development LLC.: Deloitte Development LLC; 2018. URL: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Technology-Media-Telecommunications/deloitte-digital-maturity-model.pdf [accessed 2020-04-20]
31. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3:i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
32. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital. J Am Med Inform Assoc 2017 Nov 01;24(6):1142-1148. [doi: 10.1093/jamia/ocx080] [Medline: 29016973]
33. Flott K, Callahan R, Darzi A, Mayer E. A Patient-Centered Framework for Evaluating Digital Maturity of Health Services: A Systematic Review. J Med Internet Res 2016 Apr 14;18(4):e75 [FREE Full text] [doi: 10.2196/jmir.5047] [Medline: 27080852]
34. HIMSS Continuity of Care Maturity Model (CCMM). Healtcare Information and Management Systems Society. URL: https://www.himssanalytics.org/ccmm [accessed 2019-09-13]
35. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
36. Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005 Mar 9;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
37. Martin G, Clarke J, Liew F, Arora S, King D, Aylin P, et al. Evaluating the impact of organisational digital maturity on clinical outcomes in secondary care in England. NPJ Digit Med 2019;2:41 [FREE Full text] [doi: 10.1038/s41746-019-0118-9] [Medline: 31304387]

Abbreviations

CCMM: Continuity of Care Maturity Model
CPOE: Computerized Physician Order Entry
EMR: electronic medical record
EMRAN: Electronic Medical Record Adoption Model
HIMSS: Healthcare Information and Management System Society
HIT: health information technology
NHS: National Health Service

Edited by G Eysenbach; submitted 20.11.19; peer-reviewed by A Lange, M Alrige, P Pantoja Bustillos, W Gordon; comments to author 10.03.20; revised version received 11.05.20; accepted 14.05.20; published 18.08.20.

Please cite as:
Krasuska M, Williams R, Sheikh A, Franklin BD, Heeney C, Lane W, Mozaffar H, Mason K, Eason S, Hinder S, Dunscombe R, Potts HWW, Cresswell K
Technological Capabilities to Assess Digital Excellence in Hospitals in High Performing Health Care Systems: International eDelphi Exercise
J Med Internet Res 2020;22(8):e17022
URL: https://www.jmir.org/2020/8/e17022
doi:10.2196/17022
PMID:32808938

Original Paper

Managing Illicit Online Pharmacies: Web Analytics and Predictive Models Study

Hui Zhao[1], PhD; Sowmyasri Muthupandi[2], MSc; Soundar Kumara[3], PhD

[1]Smeal College of Business, Pennsylvania State University, University Park, PA, United States

[2]PricewaterhouseCoopers, LLP, New York, NY, United States

[3]Department of Industrial and Manufacturing Engineering, Pennsylvania State University, University Park, PA, United States

Corresponding Author:
Hui Zhao, PhD
Smeal College of Business
419 Business Building
Pennsylvania State University
University Park, PA, 16802
United States
Phone: 1 814 863 1940
Email: huz10@psu.edu

Abstract

Background: Online pharmacies have grown significantly in recent years, from US $29.35 billion in 2014 to an expected US $128 billion in 2023 worldwide. Although legitimate online pharmacies (LOPs) provide a channel of convenience and potentially lower costs for patients, illicit online pharmacies (IOPs) open the doors to unfettered access to prescription drugs, controlled substances (eg, opioids), and potentially counterfeits, posing a dramatic risk to the drug supply chain and the health of the patient. Unfortunately, we know little about IOPs, and even identifying and monitoring IOPs is challenging because of the large number of online pharmacies (at least 30,000-35,000) and the dynamic nature of the online channel (online pharmacies open and shut down easily).

Objective: This study aims to increase our understanding of IOPs through web data traffic analysis and propose a novel framework using referral links to predict and identify IOPs, the first step in fighting IOPs.

Methods: We first collected web traffic and engagement data to study and compare how consumers access and engage with LOPs and IOPs. We then proposed a simple but novel framework for predicting the status of online pharmacies (legitimate or illicit) through the referral links between websites. Under this framework, we developed 2 prediction models, the reference rating prediction method (RRPM) and the reference-based K-nearest neighbor.

Results: We found that direct (typing URL), search, and referral are the 3 major traffic sources, representing more than 95% traffic to both LOPs and IOPs. It is alarming to see that direct represents the second-highest traffic source (34.32%) to IOPs. When tested on a data set with 763 online pharmacies, both RRPM and R2NN performed well, achieving an accuracy above 95% in their predictions of the status for the online pharmacies. R2NN outperformed RRPM in full performance metrics (accuracy, kappa, specificity, and sensitivity). On implementing the 2 models on Google search results for popular drugs (Xanax [alprazolam], OxyContin, and opioids), they produced an error rate of only 7.96% (R2NN) and 6.20% (RRPM).

Conclusions: Our prediction models use what we know (referral links) to tackle the many unknown aspects of IOPs. They have many potential applications for patients, search engines, social media, payment companies, policy makers or government agencies, and drug manufacturers to help fight IOPs. With scarce work in this area, we hope to help address the current opioid crisis from this perspective and inspire future research in the critical area of drug safety.

(J Med Internet Res 2020;22(8):e17239) doi:10.2196/17239

KEYWORDS

online pharmacy; web analytics; classification; illicit online pharmacies; online traffic analysis

Introduction

Online pharmacies (OPs) have grown tremendously in recent years, from US $29.35 billion in 2014 to an expected US $128 billion in 2023 globally, at an annual growth rate of 17.7% [1]. Most consumers pursue OPs for lower prices [2,3], convenience, and access to otherwise unavailable drugs, for example, recalled or on shortage [4,5]. However, there is insufficient awareness of the prevalent illicit online pharmacies (IOPs), which are estimated to represent 67%-75% web-based drug merchants [6]. Although much work has been carried out to restrict prescription for the recent opioid crisis, many IOPs provide access without prescription. IOPs provide unfettered access to prescription drugs and even controlled substances, leading to great concerns about substandard drugs, counterfeits, and supply chain integrity [7,8].

Fighting IOPs is critical in protecting patient safety as well as integrity of the drug supply chain. However, this is very challenging. First, there is low awareness of how to differentiate the legitimacy of OPs among consumers [9], and we still have much to learn about IOPs [6]. IOPs may look very similar to LOPs, and, unlike other consumer products, most consumers have no expertise in differentiating potentially substandard drugs even upon receiving them. Second, even identifying and tracking IOPs, the first step in fighting IOPs, can be challenging because of the sheer scale and the dynamic nature of the problem. According to Legitscript [6], there are 30,000-35,000 online pharmacies, and about 20 new IOPs are created when many *die* on a daily basis. Even if IOPs can be closed down (more difficult than we think as many IOPs have their servers outside of the United States), they can easily pop up using different URLs (eg, 30,000-35,000 known OPs represent only 2000-3500 merchants [6]).

A few checking systems of OP status (legitimate or illicit) do exist but with limitations. Some of them are *not* recommended [10], including the Canadian International Pharmacy Association and Pharmacychecker, which have been criticized for not always classifying the OPs correctly. The 2 sources recommended by the Food and Drug Administration are the National Association Board of Pharmacies (NABP) and Legitscript. However, both sources require consumers to take the initiative to look up the status of the pharmacies. According to a survey of 500 consumers from the United States, conducted by the Alliance for Safe Online Pharmacies, 95% do not know about the certification programs [9], let alone where to check the status of the OPs. Furthermore, there is no exhaustive database because of the aforementioned scale and the dynamic nature of OPs.

This study aims to use web analytics to better understand IOPs and to predict, identify and monitor IOPs using known information. We do this in 2 steps. First, we conducted a traffic analysis based on web-collected data, which assesses the means through which LOPs and IOPs are accessed and how engaged the customers are with them. On the basis of the information from the first step, especially through the analysis of referrals data, in the second step, we proposed a novel framework to predict the status (legitimate or illicit) of OPs based on the referral websites to them. Under this framework, we developed 2 easy-to-understand prediction models, the referral-based K-nearest neighbor (RKNN) and the referral rating prediction method (RRPM), and tested them using a data set with 763 OPs. We then implemented the 2 methods on Google search results for 3 popular drugs: Xanax (alprazolam), OxyContin, and opioids. These methods have many potential applications for consumers when shopping on the web and for other stakeholders to help fight IOPs, as presented in detail in the *Applications and Conclusions* subsection of the *Discussion* section.

Methods

Data Sources

We obtained the ground truth list of LOPs and IOPs from the NABP (Legitscript was not available for the size of our sample). NABP provided a list of approximately 1000 IOPs and 50 LOPs. We filtered out many IOPs that stopped operations at the time of data collection. We then collected usage data (ie, traffic and engagement data) for the remaining OPs from Similarweb and obtained the structure data (ie, referrals and backlink data, detailed later) from SEMrush. As Similarweb does not have data for websites not in its database or whose traffic is too low to monitor, this led us to the final sample sizes for each of the databases in Table 1. The first 5 rows in the table are the usage data, and the last row is the structure data. We collected data from Similarweb through web scraping using R. For SEMrush, we tried to collect the data manually (no crawling allowed for SEMrush). When that was impossible, we purchased the function from SEMrush (it sells different levels of functions through various priced accounts).

Table 1. Data sets and sample size.

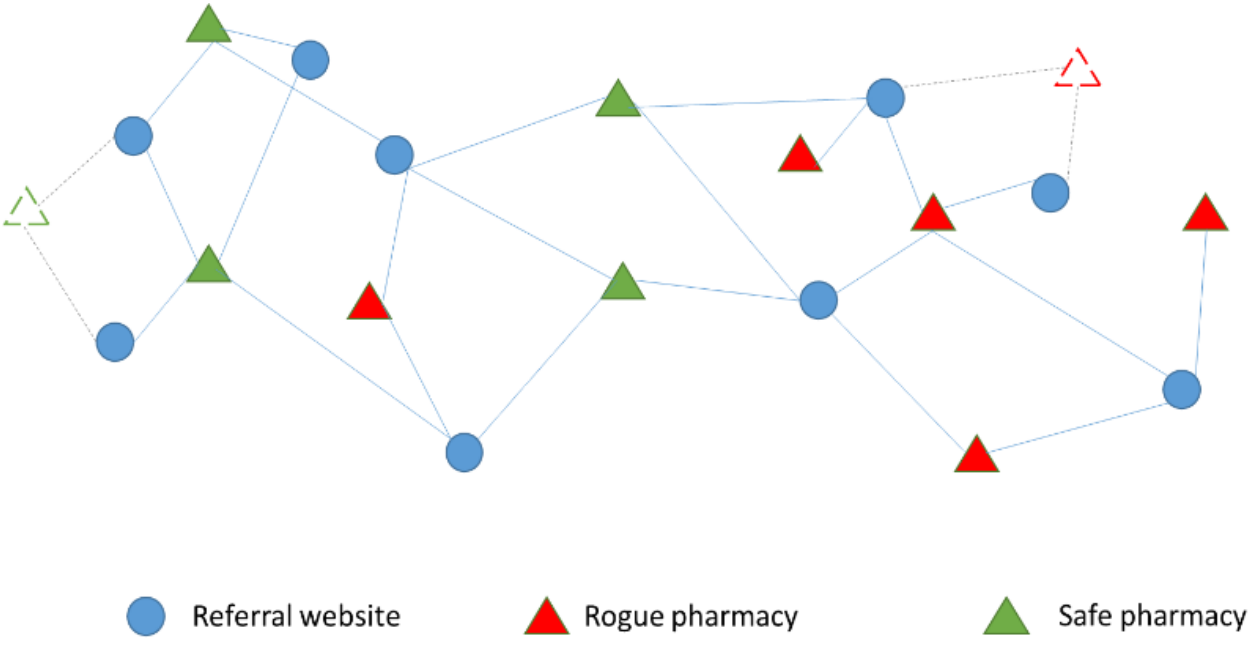
Data set names	Legitimate pharmacies, n	Illicit pharmacies, n	Total samples, n	Data collection period
Traffic sources data	30	127	157	Average over 4 months (October 2015-February 2016)
Engagement data	30	127	157	Average over 4 months (October 2015-February 2016)
Country data	30	139	169	Average over 4 months (October 2015-February 2016)
Social media data	24	41	65	Average over 4 months (October 2015-February 2016)
Search data	30	60	90	Average over 4 months (October 2015-February 2016)
Referral data	50	713	763	September 2016

In Table 1, traffic sources provide the percentage of the sources through which consumers access the OPs, that is, direct, search, referral, social media, display, and email (details later). Engagement data show the extent of users' involvement with the website (eg, the number of pages viewed and time spent on the website). Country data provide the percentage of traffic to OPs from different countries. *Social media data* refer to the proportion of traffic from 26 social media websites, such as Facebook, YouTube, and Google Plus. Search data provide the percentages of traffic resulting from organic or paid searches for OPs. An organic search, also called natural search, provides results by the search engine based on its relevance to the user's query. Paid search results are like advertisements, where the websites pay search engines to promote their web pages for particular keywords. Referral data provide the different referring websites to online pharmacies, their internet protocol addresses, and countries of origin.

OP Status Prediction Model

One of the difficulties in predicting the status of an OP is that the proposed method and the data it uses need to be something that cannot be easily manipulated by IOPs to affect future prediction results. To overcome this challenge, we propose a novel structure-based framework that predicts the status based on the relationship among the referral websites. Basically, we expect that if a pharmacy is mainly reached from referral websites that mostly link/refer to illicit pharmacies, then this pharmacy is more likely to be illicit. Figure 1 depicts an oversimplified demonstration of this idea and the links between referral websites.

Figure 1. Simple Demonstration of Links Between Referral Websites and online pharmacies.



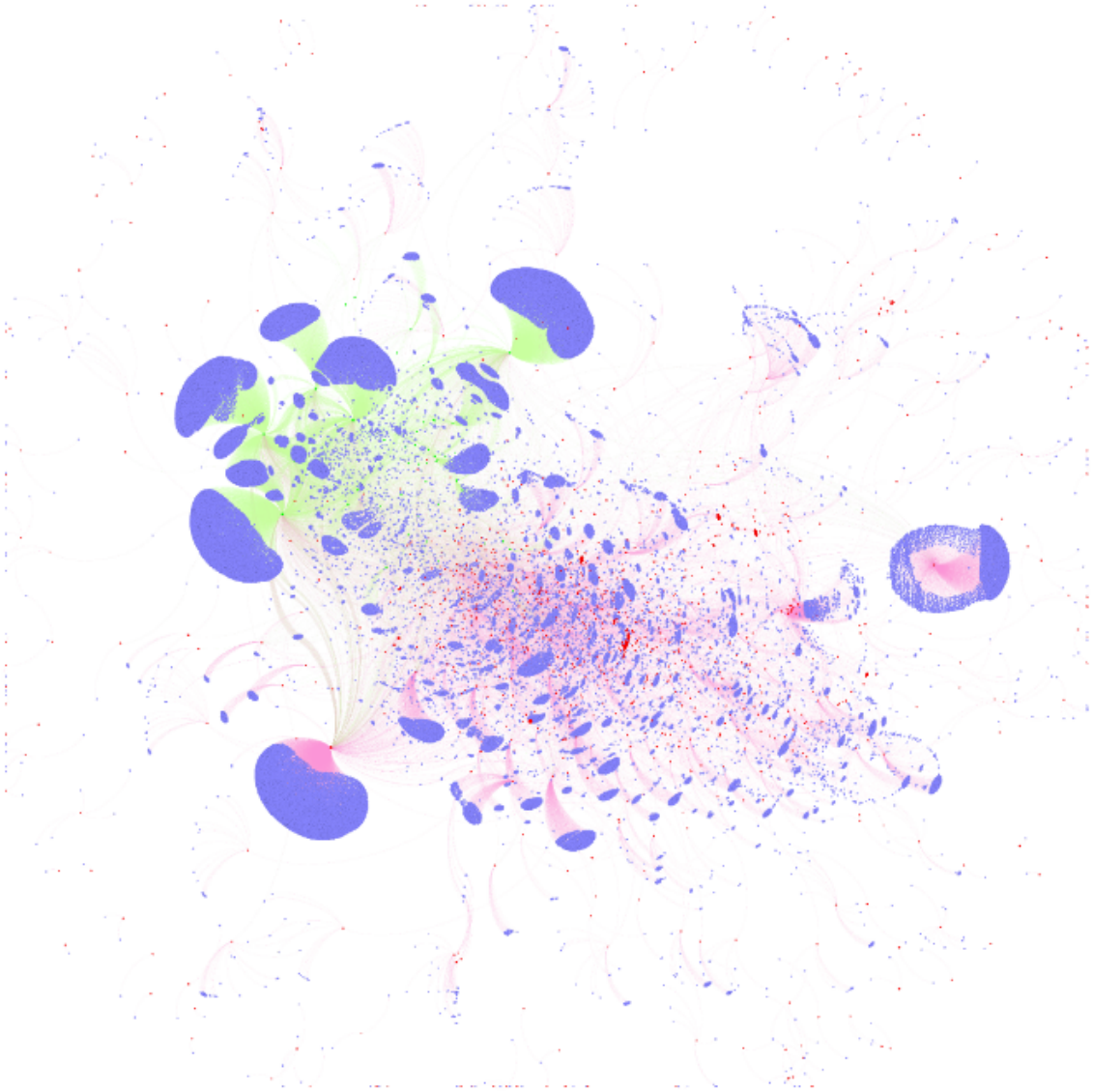
To execute this idea, based on the ground truth list of LOPs and IOPs from NABP, we identified all the websites referring to the OPs in the data set and collected the structure data, that is, referrals and information of the number of backlinks to each OP, where a backlink is a link from a website to another website (eg, the OP here). These data, listed in Table 1 as the referral data, were then used to train the prediction model. Table 2 provides a snapshot of these data (the entries are the number of backlinks from a referral site j to a pharmacy i).

Table 2. Demonstration of our data set for the prediction model.

Pharmacy site, i	Referral site, j					
	1	2	3	...	j	...
1	5	0	3	…	16	…
2	9	3	0	…	0	…
i	0	0	0	…	2	…
...	…	…	…	…	…	…

Figure 2 plots all our referral data. In the figure, the pink nodes are the IOPs, the green nodes are the LOPs, and the blue nodes are their referral websites. This figure shows 2 interesting phenomena: (1) LOPs and IOPs are clearly separated by the referral websites directing to them (although some referral websites refer to both IOPs and LOPs), that is, IOPs tend to be referred to by referring websites referring to other IOPs, and vice versa and (2) *good* referral websites tend to cluster in groups referring to each other's referred pharmacies, whereas *bad* referral websites scatter around (they refer to all kinds of pharmacies far and between). These 2 phenomena, especially the first one, confirms our basic idea of using the *quality* of the referral websites (ie, how much the referring websites refer to LOPs) to predict the status of the OPs.

Figure 2. Relationship among referral websites and LOPs and IOPs based on real data.The pink nodes are the IOPs, the green nodes are the LOPs and the blue nodes are their referral websites. LOPs: legitimate online pharmacies; IOPs: illicit online pharmacies.



Then, we described in detail 2 prediction models that we developed based on this idea, that is, the RRPM and the RKNN.

RRPM

Let  represent the set of pharmacies,  represent the set of referring websites, l_{ij} represent the number of backlinks to pharmacy *i* from referring website *j*, and $y_i = 1$ (*illicit*) or 0 (*licit*) represent the status of pharmacy *i*. We are now ready to present the model.

Step 1: First define the quality of a referral website j (M_j) based on its backlinks to legitimate and illicit pharmacies as

where  *represents the set of safe or legitimate pharmacies and*  *represents the set of rogue or illicit pharmacies.*

Therefore,  represent the number of LOPs, IOPs, and any OPs website *j* refers to, respectively. It is easy to see that M_j is between –1 and 1 with M_j=−1 indicating that website *j* only refers to IOPs and M_j=1 indicating that website *j* only refers to LOPs.

Step 2: For pharmacy i whose status is to be predicted, calculate the reliability score (R_i) as

where  *is the total number of backlinks (referral websites) to pharmacy i*. Note that it is possible that a given pharmacy to be predicted does not have any referral website from the training set. In this case, its R_i will be indeterminate  and R_i is set to 0.

On the basis of our framework, we expect that the higher the R_i is, the more likely it is legitimate. For our prediction model, we set a threshold T for the reliability score above which we predicted the pharmacy to be legitimate. In determining T, we considered a crucial factor, the sensitivity of the model, which measured the proportion of IOPs that were correctly identified as such. Although predicting a pharmacy wrong in either way is risky, for safety reasons, from the consumers' perspective, classifying an illicit pharmacy as legitimate may be more detrimental than classifying a legitimate pharmacy as illicit. Taking this into consideration, we have the following:

Step 3: Set the threshold T from the training set as

such that we will classify pharmacy i as illicit if R_i<T and legitimate otherwise.

Notice that this is a very conservative threshold, meaning that a pharmacy is highly unlikely to be an illicit one when it is predicted legitimate. This could hurt the average accuracy of the model. Therefore, considering the average accuracy, one could select a different threshold. Another way we tried is to test different threshold levels with the training set and choose the one with the highest accuracy. Later, we reported the accuracy for both thresholds.

RKNN

In addition to RRPM, we next adapted one of the established classification methods, K-nearest neighbor (KNN), to our framework based on the referral links to develop another prediction model. KNN is a supervised learning model that classifies the samples in the test set based on their proximity to the samples of different classes in the training set [11,12]. The key to this method is defining proximity (similarity). We now incorporated our idea of the proposed framework into this definition.

Step 1: Compute the Euclidean distance between the pharmacy x (the one whose status is to be predicted) and all the online pharmacies i *with known status* i=*1,2,…,*n, *as*

Note that the smaller the D_i is, the more similar pharmacy x is to pharmacy *i* in terms of the referring websites directing to them.

Step 2: Order the online pharmacies in decreasing order with respect to D_i. Note down the status of the top K pharmacies. According to the traditional KNN, the status/class of x is assigned to the more frequent status among the K pharmacies. Formally, let the number of legitimate pharmacies among the top K be K_s and illicit ones be K_r. We know that K_s+K_r=K. Let

x will be predicted to be legitimate if R_x>0.5. R_x is similar to the reliability score of x, indicating the strength of the prediction, with a higher R_x signifying a stronger prediction.

Similar to KNN, the performance of the RKNN model varies for different values of K. Obviously, a too high or too low value of K may reduce the accuracy of the model. We tested K=1, 2,…9 and reported the performance of the model for each value of K.

Results

Traffic and Use Analysis of the Online Pharmacies

Traffic sources of all websites are classified as direct, search, referral, social media, display, and email. Specifically, *traffic obtained by users'* directly typing in the URL of the website is classified as *direct*; *search* refers to the traffic coming from search engines such as Google, Bing, and Yahoo; traffic from links on other websites are accounted for as *referral*; *social* indicates the traffic from social media such as Facebook and Twitter; *display* indicates the traffic from banner advertising; and *Email* indicates the traffic coming from links in email messages.

Table 3 shows the mean percentage of traffic from each source to the IOPs and LOPs in our traffic data set (the standard deviation is shown in Multimedia Appendix 1). According to Table 3, direct, search, and referral are the 3 major traffic sources, representing more than 95% traffic to both LOPs and IOPs. A high percentage of direct traffic indicates that the specific OP website is a powerful brand and users visiting the website know what they want. Although LOPs are most accessed

through direct traffic (42.48%), it is alarming to see that direct traffic also represents the second highest traffic source for IOPs (34.32%). This indicates that consumers who have previous experiences with IOPs (eg, from search, referrals) may become returning customers without knowing the aforementioned potential danger. Therefore, it is imperative to educate and alert patients, and curb people from using IOPs for the first time.

Table 3. Mean percentages of different traffic sources to online pharmacies.

Traffic source	Legitimate online pharmacy (n=30), %	Illicit online pharmacy (n=127), %
Direct	42.5	34.3
Search	36.3	39.3
Referral	17.7	21.7
Social	1.3	0.9
Email	2.2	0.6
Display	0	2.5

Previous research has indicated the presence of IOP contents on various social media sites [3]. Our data show that the average percentages of traffic through social media are less than 5% for both IOPs and LOPs, possibly because many OPs in our sample do not have traffic from social media. When we focused on the 24 LOPs and 41 IOPs from our sample that have substantial traffic from the 26 social media sites to conduct further analysis, we found that 92% (24/26) of the studied social media websites direct traffic to IOPs and only 50% (13/26) of them direct traffic to LOPs. Although 42% (11/26) of these social media websites direct traffic to both LOPs and IOPs, 50% (13/26) of the sites direct traffic to only IOPs and only 8% (2/26) of the websites direct traffic to only LOPs. Among the various social media (Table 4), we found that Facebook directs the highest traffic to both IOPs (58%) and LOPs (42%), far exceeding the second highest (Reddit), which directs 20% traffic to IOPs and 15% to LOPs.

Table 4. Traffic from social media websites to online pharmacies.

Social media	Proportion of traffic to legitimate pharmacies (n=24), %	Proportion of traffic to illicit pharmacies (n=42), %
Facebook	58	42
Reddit	15	20
YouTube	14	11
Twitter	4	—[a]
LinkedIn	2	—
Askville	—	7
Pinterest	—	4
Others	7	16

[a]Data negligibly small.

Furthermore, our country data (Table 5) show traffic from 52 countries for the 155 online pharmacies for which we were able to collect country data. Traffic from 27 (52%) countries points to only IOPs, and only 3 (6%) countries have traffic only to LOPs. In addition, the United States is the main consumer for online pharmacies, representing the highest proportion of traffic to both LOPs (97%) and IOPs (60%), among all countries.

Table 5. Traffic from different countries to online pharmacies.

Countries	Proportion of traffic to legitimate online pharmacies (n=30), %	Proportion of traffic to illicit online pharmacies (n=139), %
United States	97	71.1
Canada	1	—[a]
India	1	6.7
United Kingdom	—	7.6
Others	1	14.6

[a]Data negligibly small.

Table 6 shows the average engagement metrics across the LOPs and IOPs. This shows that the average monthly views (in millions), the number of pages viewed, and time spent on the sites of LOPs are all higher than those of IOPs, whereas the

bounce rate (the percentage of visitors leaving the website after viewing only one page) is lower for LOPs than that for IOPs. This indicates that when consumers enter the OP websites, they seem to be more engaged with LOPs than IOPs. However, there are large variances in the monthly views of the websites, reflecting the huge differences among the websites in IOP as well as in LOP. For example, among the licit ones, cvs.com and walgreens.com are definitely the giants, whereas many others only attract a small number of views. In addition, although the average time spent on LOP sites (5 min) was significantly higher than that of IOPs (3.3 min) with $P<.001$, the maximum time spent on the sites of IOPs (17.4 min) was much higher than that on LOPs (10.6 min). This indicates that when consumers are interested, they may become very engaged with an IOP, leading to potential transactions. Hence, it is imperative to warn the patients before they enter a potential IOP, using the prediction method proposed by us.

Table 6. Consumers' engagement with online pharmacies.

Types of the online pharmacies	monthly views in millions, mean (SD)	Number of page views, mean (SD)	Bounce rate, mean (SD)	Time on site in minutes, mean (SD)
Legitimate online pharmacy	1.48 (3.05)	7.2 (3.5)	32.2 (16.1)	5.0 (2.7)
Illicit online pharmacy	0.02 (0.05)	4.0 (2.1)	49.4 (17.9)	3.3 (2.2)

OP Status Prediction Models

We now report the performance of the RRPM and RKNN models in their prediction. We consider 4 performance measures: accuracy, kappa, sensitivity, and specificity. Although sensitivity describes the percentage of IOPs correctly identified, specificity describes the percentage of LOPs correctly identified. We can see that *Type I error=1−specificity* and *Type II error=1*−sensitivity. As discussed, we chose the threshold T to pursue a minimum type II error, that is, a maximum sensitivity. In addition, the developed model should have good accuracy and reasonable kappa values, where kappa measures the agreement between observed and predicted classes considering to some extent the possibility of agreement by chance [13].

With 10-fold cross-validation [14], the performance metrics of RKNN (with K=1-9) and RRPM are shown in Table 7. It can be observed that all the RKNN models achieved 100% sensitivity. However, the specificity, accuracy, and kappa first increase and then decrease as K increases with R2NN performing the best, showing excellent metrics. RRPM also performs reasonably well, achieving a sensitivity of 99.2%, with relatively lower values for kappa and specificity. When changing the threshold T for RRPM from the current relatively conservative value to be the reliability score maximizing the model accuracy in the training data set, model accuracy, kappa, and specificity all we improve much. But sensitivity slightly dropped, as expected (Table 7).

Table 7. Performance of the classification models.

Model	Accuracy	Kappa	Specificity	Sensitivity
R1NN[a]	0.984	0.844	0.76	1
R2NN[a]	*0.986* [b]	*0.859*	*0.78*	*1*
R3NN[a]	0.979	0.789	0.68	1
R4NN[a]	0.975	0.729	0.62	1
R5NN[a]	0.975	0.729	0.62	1
R6NN[a]	0.972	0.711	0.58	1
R7NN[a]	0.965	0.600	0.46	1
R8NN[a]	0.954	0.431	0.30	1
R9NN[a]	0.949	0.321	0.22	1
RRPM[c]	0.950	0.434	0.36	0.992
RRPM (alternative threshold)	0.968	0.648	0.78	0.977

[a]RKNN: reference-based K-nearest neighbor, where K=1-9.

[b]Indicates the best performing model.

[c]RRPM: reference rating prediction method.

Implementing RRPM and RKNN on Google Search Results

Our traffic analysis showed that search accounts for the highest traffic to IOPs (39.27%). Our prediction model can be used in a couple of ways for search engines: (1) it can be incorporated on top of search results to filter/flag search results that are likely IOPs and (2) the reliability scores of the OPs can be used to rank the results such that more reliable OPs would appear first. Therefore, we tested our model on Google search results for 3 popular drugs.

Xanax (alprazolam) is a type of benzodiazepine. More than 30 percent of overdoses involving opioids also involve benzodiazepines [15]. Anecdotal evidence indicates that such drugs are typically the target of IOPs. We monitored the top keywords that direct traffic to OPs and identified that keywords with the drugs' names contributed to more traffic than keywords without drug names. Hence, we chose *buy Xanax online* as the keyword and collected the top 100 search results for the keyword search on September 9, 2016. Almost all the search results were pharmacies selling Xanax on the web. As a result of the opioid crisis, along with *buy Xanax online*, we also studied the search results of the keywords *buy opioids online* and *buy OxyContin online* on April 22, 2017. OxyContin carries a boxed warning and contains oxycodone, a Schedule II controlled substance with an abuse potential similar to other Schedule II opioids.

To test our results, we hand collected the status of the OPs obtained through the top 100 search results from the NABP and Legitscript. Table 8 provides the status from both sources. Results demonstrate that neither source has an exhaustive database, although Legitscript (which only allows checking 10 pharmacies daily without a fee) has a bigger database, confirming what was found by Mackey et al [16]—that hand or automated search of opioid-related sites results in websites not covered by the Legitscript database. It is alarming to note that none of the pharmacies from the top 100 search results are legitimate by definition of either NABP or Legitscript. We then used RRPM and RKNN to predict the status of these pharmacies and compared our prediction results with the OP status according to Legitscript and NABP (Table 8).

Table 8. Status of the search results according to Legitscript and National Association Board of Pharmacies.

Keywords searched	IOP[a] by NABP[b]	LOP[c] by NABP	Unknown from NABP	IOP/rogue by Legitscript	LOP/safe by Legitscript	Unknown from Legitscript
Buy Xanax online	11	0	89	48	0	52
Buy Opioids online	6	0	94	34	0	66
Buy OxyContin online	10	0	90	25	0	75

[a]IOP: illicit online pharmacy.

[b]NABP: National Association Board of Pharmacies.

[c]LOP: legitimate online pharmacy.

As our model relies on the referral data, when the referral data for a particular online pharmacy is not available, its status is defined as unknown by our model. Table 9 compares the prediction results from RRPM and R2NN, respectively, with those from the Legitscript and NABP databases (NABP numbers are shown in parentheses) for the pharmacies obtained from the top 100 search results for the 3 keyword searches (hence, 300 overall). For instance, according to Table 9, 104 (27) pharmacies are correctly predicted illicit and 2 (0) are incorrectly predicted as legitimate pharmacies by RRPM when compared with the status defined by Legitscript (NABP). In addition, the status of 7 (0) IOPs according to the Legitscript (NABP) database cannot be identified by RRPM because of the lack of referral data.

Table 9. Comparison of the predicted status of online pharmacies based on reference rating prediction method (RRPM) and reference-based K-nearest neighbor (RKNN) with those obtained from Legitscript and National Association Board of Pharmacies (NABP) databases, with NABP numbers in parentheses.

Prediction results	Status obtained from Legitscript and NABP databases (NABP numbers in parentheses)		
	Illicit	Legitimate	Unknown
Status estimated by RRPM[a]			
Illicit	104 (27)	0 (0)	147 (224)
Legitimate	2 (0)	0 (0)	3 (5)
Unknown	7 (0)	0 (0)	37 (44)
Status estimated by R2NN[b]			
Illicit	106 (27)	0 (0)	145 (225)
Legitimate	0 (0)	0 (0)	5 (5)
Unknown	7 (0)	0 (0)	37 (43)

[a]RRPM: reference rating prediction method.

[b]R2NN: reference-based K-nearest neighbor.

Excluding those that are unknown from the corresponding databases, the tables show that RRPM and R2NN produced an error rate of 7.96% (0%) and 6.20% (0%), respectively, based on the Legitscript (NABP) database. The above results provide evidence that the proposed prediction models can predict online pharmacies with reasonably good accuracy.

Discussion

Comparison With Previous Work

In this study, we conducted a web traffic and engagement analysis of IOPs and LOPs, developed simple prediction models of the status of the OPs based on referral links, and tested the prediction models with data for 763 online pharmacies. Although the previous literature shows evidence of drug selling through IOPs, there has been very limited work on the traffic to these websites. One exception is the study by Mackey et al [5], which estimated traffic to an IOP through fictitious advertisements for selling drugs without prescription that they created on social media. In contrast, we collected true data on the traffic analysis and the prediction models.

Similarly, very limited research is related to identifying and predicting the status of OPs. The study by Fittler et al [10] aimed to identify the indicators of IOPs by evaluating 136 of them based on the longevity of the site, geographical location, display of contact information, medical information exchange, prescription requirement, and pharmacy legitimacy verification. They identified that the prescription requirement or availability of contact information does not correlate with illicit pharmacy status as indicated by Legitscript; however, the long-term continuous operation of the website has a strong correlation with illicit activities. They did not develop a prediction model.

Predicting the status of OPs is related to classifying different websites into certain categories. In general, there are 2 types of approaches: content based and structure based. Hybrid methods also exist. While content-based classification [17] utilizes the website content to classify the website, structure-based classification exploits the patterns in the link structure or the topology of the hyperlinks of the websites. For example, Amitay et al [18] used structural information to classify 8 classes of websites (eg, corporate sites, search engines, and e-store), with the precision of certain classes exceeding 85%. Our prediction model is structure based, and it is easy to see what we try to classify as IOP and LOP is much more subtle.

Research on the prediction/classification of LOPs and IOPs is very scarce. The only other work is the study by Corona et al [19] aimed at building a database of OPs using textual content analysis. Note that content-based prediction could be more easily manipulated than structure-based prediction. For example, if the prediction is based on certain content appearing on the websites, then IOPs could delete or change the content to confuse the model by making it just *like* LOPs. Toward this end, this paper proposes a novel yet simple structure-based idea using relationships among referral websites to predict the status of OPs.

Finally, when searching the literature for general prediction and classification of websites selling counterfeit products (not limited to drugs), only 2 studies were found [20,21]. Both used content analysis in general, achieving an accuracy of 86.4% [20] and 88% [21]. Our approach can potentially be applied to more general products than just drugs.

Limitations

As we propose a new methodology, we face many limitations. First, because of the limited source of the available ground truth of the status of online pharmacies and the data related to traffic analysis, we have a relatively small sample size (for some of the traffic analysis). We expect that a larger sample size when available will improve the accuracy of the results and allow more detailed analysis. When using Google search results, we also face many websites whose true status is unknown; hence, evaluation of our methods using the Google search results presented in our paper is limited. Second, the current website information sources (SEMRush and Similarweb) do not provide reliable (or any) information for small websites lacking sufficient data for traffic overview. Accordingly, the findings

of this research are mostly applicable to larger legitimate and illicit web-based players. Third, our proposed method relies on referral website data. Our current referral database from the data we collected seemingly works well. However, obviously, the bigger the referral link database, the better. Although outdated links do not hurt the performance (they will not be used), updating these links as more ground truth data becomes available would be desired. Finally, we focus on proposing a novel structure-based prediction framework and developing simple models to help resolve an important and practical problem. More advanced models, such as a hybrid of structure based and contexture based, can be developed in the future to further improve performance.

Applications and Conclusions

Previous research shows that illicit online pharmacies are present and widely accessed, posting dramatic risks to the drug supply chain integrity and patient health. However, because of the sheer scale of this problem (>30,000 OPs) and the dynamic nature of online channels, even identifying and monitoring IOPs, the first step to curb IOPs, is a difficult task. In this study, we aimed to fill this gap by conducting a traffic analysis to increase our understanding of IOPs and proposed a new idea to predict the OP status based on referral data and developed 2 specific prediction models (RRPM and RKNN) using this idea. Testing these models on a data set with 763 online pharmacies showed that both models performed well, with an accuracy of 95.0% (RRPM) and 98.6% (RKNN). R2NN outperformed RRPM in more comprehensive metrics (sensitivity, kappa, and specificity). When implementing both models on the Google search results for 3 drugs, we only incurred an error rate of 6.20% for the pharmacies whose true status was known according to the Legitscript database when using the R2NN model and an error rate of 7.96% when using RRPM for the prediction. Although further testing with a larger data set is being pursued (the difficulty is the limited ground truth data), we believe our traffic analysis and the approach to use web analytics of referral websites to predict the status of OPs is among the first in the drug field and proposes a viable and evidently effective way to monitor OPs.

The developed framework/models have numerous exciting application areas. For example, they can be implemented by search engines, social media, web-based markets (eg, Amazon), and payment companies (eg, Visa and Master cards) to filter IOPs or take the status of the online pharmacies into consideration when ranking search results, deciding advertising allocations, making payments, disqualifying vendors, or at least warning consumers of potential IOPs. They can also be used with search engines and social media to develop a warning system to help consumers make informed decisions. The timeliness of this work could help address the current opioid crisis. Policy makers, government agencies, patient advocacy groups, and drug manufacturers may also use such a system to identify, monitor, curb IOPs, and educate consumers.

Given that this is a critical area of concern to patients' health and the integrity of the drug supply chain, we hope this study will inspire additional efficient and effective prediction models or additional applications for the prediction models developed. On a larger scale, we hope to inspire more research in other aspects to fight IOPs. Finally, our literature review also reveals that literature on automatic prediction/identification of websites selling counterfeit products (not limited to drugs) is also very scarce, although selling counterfeit products on the web is a prevalent problem. Our framework and prediction models can be applied to other products, and we hope to inspire research in this general area as well.

Conflicts of Interest

None declared.

Multimedia Appendix 1
More details of the engagement and traffic source data.
[DOCX File , 15 KB - jmir_v22i8e17239_app1.docx]

References

1. ePharmacy Market. Transparency Market Research. 2015. URL: http://www.transparencymarketresearch.com/epharmacies-market.html [accessed 2017-03-03]
2. Quon BS, Firszt R, Eisenberg MJ. A comparison of brand-name drug prices between Canadian-based Internet pharmacies and major US drug chain pharmacies. Ann Intern Med 2005 Sep 20;143(6):397-403. [doi: 10.7326/0003-4819-143-6-200509200-00004] [Medline: 16172438]
3. Liang BA, Mackey TK. Prevalence and Global Health implications of social media in direct-to-consumer drug advertising. J Med Internet Res 2011 Aug 31;13(3):e64 [FREE Full text] [doi: 10.2196/jmir.1775] [Medline: 21880574]
4. Liang BA, Mackey TK. Online availability and safety of drugs in shortage: a descriptive study of internet vendor characteristics. J Med Internet Res 2012 Feb 9;14(1):e27 [FREE Full text] [doi: 10.2196/jmir.1999] [Medline: 22321731]
5. Mackey TK, Aung P, Liang BA. Illicit Internet availability of drugs subject to recall and patient safety consequences. Int J Clin Pharm 2015 Dec;37(6):1076-1085. [doi: 10.1007/s11096-015-0154-8] [Medline: 26148859]
6. The Internet Pharmacy Market in 2016: Trends, Challenges and Opportunities. The Center for Safe Internet Pharmacies (CSIP). 2016. URL: https://safemedsonline.org/resource/the-internet-pharmacy-market-in-2016-trends-challenges-and-opportunities/ [accessed 2020-07-28]

7. Jena AB, Goldman DP. Growing internet use may help explain the rise in prescription drug abuse in the United States. Health Aff (Millwood) 2011 Jun;30(6):1192-1199 [FREE Full text] [doi: 10.1377/hlthaff.2011.0155] [Medline: 21565838]
8. Jena AB, Goldman DP, Foster SE, Califano JA. Prescription medication abuse and illegitimate internet-based pharmacies. Ann Intern Med 2011 Dec 20;155(12):848-850. [doi: 10.7326/0003-4819-155-12-201112200-00008] [Medline: 22184692]
9. Consumers' Lack of Awareness is a Worry for Online Pharmacy Legitimacy. Pharmaceutical Commerce Magazine. 2017. URL: https://pharmaceuticalcommerce.com/latest-news/consumers-lack-awareness-worry-online-pharmacy-legitimacy/ [accessed 2020-04-03]
10. Fittler A, Bősze G, Botz L. Evaluating aspects of online medication safety in long-term follow-up of 136 internet pharmacies: illegal rogue online pharmacies flourish and are long-lived. J Med Internet Res 2013 Sep 10;15(9):e199 [FREE Full text] [doi: 10.2196/jmir.2606] [Medline: 24021777]
11. Fix E, Hodges JL. Discriminatory analysis. Nonparametric discrimination: consistency properties. Int Stat Rev 1989 Dec;57(3):238. [doi: 10.2307/1403797]
12. Lihua Y, Qi D, Yanjun G. An Improved KNN Text Classification Algorithm Based on Clustering. J Comput 2009;4(3):- [FREE Full text]
13. Khun M, Johnson K. Applied Predictive Modeling. New York, USA: Springer Science & Business Media; 2013.
14. Cawley GC, Talbot NL. Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. Pattern Recogn 2003 Nov;36(11):2585-2592. [doi: 10.1016/s0031-3203(03)00136-5]
15. Benzodiazepines and Opioids. National Institute on Drug Abuse (NIDA). 2018 Mar 15. URL: https://www.drugabuse.gov/drug-topics/opioids/benzodiazepines-opioids
16. Mackey T, Kalyanam JJ, Klugman J, Kuzmenko E, Gupta R. Solution to detect, classify, and report illicit online marketing and sales of controlled substances via twitter: using machine learning and web forensics to combat digital opioid access. J Med Internet Res 2018 Apr 27;20(4):e10029 [FREE Full text] [doi: 10.2196/10029] [Medline: 29613851]
17. Pierre J. On the automated classification of websites. Linkoping Electron Artical Comput Inform Sci 2001:6 [FREE Full text]
18. Amitay E, Carmel D, Darlow A, Lempel R, Soffer A. The Connectivity Sonar: Detecting Site Functionality by Structural Patterns. In: Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia. 2003 Presented at: HYPERTEXT'03; June 11-15, 2003; Maryland, USA. [doi: 10.1145/900051.900060]
19. Corona I, Contini M, Ariu D, Giacinto G, Roli F, Lund M, et al. PharmaGuard: Automatic Identification of Illegal Search-Indexed Online Pharmacies. In: 2nd International Conference on Cybernetics. 2015 Presented at: CYBCONF'15; June 24-26, 2015; Gdynia, Poland. [doi: 10.1109/cybconf.2015.7175954]
20. Wadleigh J, Drew J, Moore T. The E-Commerce Market for 'Lemons': Identification and Analysis of Websites Selling Counterfeit Goods. In: Proceedings of the 24th International Conference on World Wide Web. 2015 Presented at: WWW'15; May 18-22, 2015; Florence, Italy. [doi: 10.1145/2736277.2741658]
21. Carpineto C, Romano G. Learning to Detect and Measure Fake Ecommerce Websites in Search-engine Results. In: Proceedings of the International Conference on Web Intelligence. 2017 Presented at: WI'17; September 25-29, 2017; Leipzig, Germany. [doi: 10.1145/3106426.3106441]

Abbreviations

IOP: illicit online pharmacies
KNN: K-nearest neighbor
LOP: legitimate online pharmacies
NABP: National Association Board of Pharmacies
RKNN: reference-based K-nearest neighbor
RRPM: reference rating prediction method

Edited by G Eysenbach; submitted 03.12.19; peer-reviewed by A Fittler, M Spiliopoulou; comments to author 10.02.20; revised version received 14.04.20; accepted 14.05.20; published 25.08.20.

Please cite as:
Zhao H, Muthupandi S, Kumara S
Managing Illicit Online Pharmacies: Web Analytics and Predictive Models Study
J Med Internet Res 2020;22(8):e17239
URL: http://www.jmir.org/2020/8/e17239/
doi:10.2196/17239
PMID:32840485

Original Paper

Views of Implementers and Nonimplementers of Internet-Administered Cognitive Behavioral Therapy for Depression and Anxiety: Survey of Primary Care Decision Makers in Sweden

Anders Brantnell[1,2], PhD; Joanne Woodford[1], PhD; Enrico Baraldi[2], PhD; Theo van Achterberg[3], PhD; Louise von Essen[1], PhD

[1]Clinical Psychology in Healthcare, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[2]Division of Industrial Engineering and Management, Department of Civil and Industrial Engineering, Uppsala University, Uppsala, Sweden

[3]KU Leuven, Department of Public Health and Primary Care, Academic Centre for Nursing and Midwifery, Leuven, Belgium

Corresponding Author:
Anders Brantnell, PhD
Clinical Psychology in Healthcare
Department of Women's and Children's Health
Uppsala University
Dag Hammarskjölds Väg 14B
Uppsala
Sweden
Phone: 46 18 471 70 55
Email: anders.brantnell@angstrom.uu.se

Abstract

Background: Internet-administered cognitive behavioral therapy (ICBT) has been demonstrated to be an effective intervention for adults with depression and/or anxiety and is recommended in national guidelines for provision within Swedish primary care. However, the number and type of organizations that have implemented ICBT within primary care in Sweden is currently unclear. Further, there is a lack of knowledge concerning barriers and facilitators to ICBT implementation.

Objective: The two primary objectives were to identify and describe primary care organizations providing ICBT in Sweden and compare decision makers' (ie, directors of primary care organizations) views on barriers and facilitators to implementation of ICBT among ICBT implementers (ie, organizations that offered ICBT) and nonimplementers (ie, organizations that did not offer ICBT).

Methods: An online survey based on a checklist for identifying barriers and facilitators to implementation was developed and made accessible to decision makers from all primary care organizations in Sweden. The survey consisted of background questions (eg, provision of ICBT and number of persons working with ICBT) and barriers and facilitators relating to the following categories: users, therapists, ICBT programs, organizations, and wider society.

Results: The participation rate was 35.75% (404/1130). The majority (250/404, 61.8%) of participants were health care center directors and had backgrounds in nursing. Altogether, 89.8% (363/404) of the participating organizations provided CBT. A minority (83/404, 20.5%) of organizations offered ICBT. Most professionals delivering ICBT were psychologists (67/83, 80%) and social workers (31/83, 37%). The majority (61/83, 73%) of organizations had 1 to 2 persons delivering ICBT interventions. The number of patients treated with ICBT during the last 12 months was 1 to 10 in 65% (54/83) of the organizations, ranging between 1 and 400 treated patients across the whole sample. There were 9 significant (P<.05) differences out of 37 possible between implementers and nonimplementers. For example, more implementers (48/51, 94%) than nonimplementers (107/139, 76.9%) perceived few technical problems (P<.001), and more implementers (53/77, 68%) than nonimplementers (103/215, 47.9%) considered that their organization has resources to offer ICBT programs (P<.001).

Conclusions: Despite research demonstrating the effectiveness of ICBT for depression and anxiety and national guidelines recommending its use, ICBT is implemented in few primary care organizations in Sweden. Several interesting differences between implementers and nonimplementers were identified, which may help inform interventions focusing on facilitating the implementation of ICBT.

(J Med Internet Res 2020;22(8):e18033) doi:10.2196/18033

KEYWORDS

mental health; internet-administered CBT; self-management; implementation; barriers and facilitators; decision-making; eHealth; primary care

Introduction

Background

According to the World Health Organization, over 300 million people suffer from depression worldwide [1]. Further, over 260 million people suffer from anxiety disorders (eg, generalized anxiety disorder, panic disorder, and social anxiety disorder) [1]. Both depression and anxiety result in individual suffering and are associated with increased costs for both individuals and wider society [2-5]. One evidence-based intervention for depression and anxiety disorders (hereafter anxiety) is cognitive behavioral therapy (CBT), traditionally delivered face to face by a therapist [6-9]. However, despite the evidence base for CBT, a mental health treatment gap remains globally [10,11]. Some reasons for this treatment gap concern the fact that therapy is resource-intensive (eg, needing facilities to meet patients, requiring patients to travel to clinics, and needing trained, competent, and expensive therapists) [12]. One potential solution to overcome this treatment gap is the provision of psychological interventions via information technology and new media, referred to as e-mental health [13,14]. Internet-administered CBT (ICBT) [15-18] is a form of e-mental health that has been demonstrated to be as effective as face-to-face CBT [19] and may represent a cost-effective solution, which may increase the availability of evidence-based psychological interventions [20,21].

Sweden was among the first countries to conduct research on the efficacy of ICBT [16,22-25], leading to the introduction of national guidelines regarding the treatment of depression and anxiety [26]. These national guidelines are directed to decision makers at the level of primary care and offer health care providers with evidence-based intervention recommendations, including the provision of CBT and ICBT [26]. Specifically, national guidelines recommend ICBT to be offered to adults with mild to moderate symptoms of depression and anxiety at the primary care level [26]. The guidelines were recently revised to state health care providers should be able to choose the mode of CBT delivery by themselves, with recommended modes of delivery being individual, group, or internet-administered [27].

However, guideline introduction and uptake is difficult, and guidelines are not implemented into routine care automatically [28-32]. While barriers and facilitators to implementation in health care more generally have been identified and related to patients, informal caregivers, practitioners, health care organizations, and wider society [33], few studies have focused on barriers and facilitators to ICBT implementation more specifically [34-36]. Indeed, studies on ICBT implementation into clinical practice focus mostly on effectiveness and clinical feasibility [20,37-41]. Despite the current literature on ICBT implementation being in its infancy, recent studies have begun to identify barriers and facilitators. For example, evidence supporting ICBT programs is reported to be a key facilitator to implementation [34,35], whereas lack of resources [34] and lack of integration with the mental health care system [36] are reported to be important barriers to ICBT implementation. Implementation of e-mental health has received some interest, and studies have identified barriers such as lack of therapists' knowledge on e-mental health [42] and lack of evidence-based programs [43]. Further, a recent systematic review of barriers and facilitators to implementation included 47 studies and identified factors such as patient and professional acceptance of e-mental health and fit with existing technologies [44]. In addition, while implementation of e-health interventions more generally is well covered in existing research [45-51], there have been recent calls to place more focus on implementation of e-health [52] to enable health care organizations to benefit from the promises of e-health solutions [53].

Studies examining the implementation of ICBT in clinical practice generally focus on the perspectives of patients and practitioners [20,37,39,41]. More recent studies have also included health care managers [34,36]. Although these three groups are important for successful implementation, these studies do not consider the formal decision to implement, which is usually made by primary care and specialized care directors. Consequently, research considering important stakeholders (ie, decision makers directly affecting the implementation of ICBT and their corresponding health care organizations) is lacking. More precisely, at present there is a lack of knowledge about the opinions of key decision makers regarding barriers and facilitators to implementation and which organizations have implemented ICBT in Sweden. Further, while some research has focused on decision makers operating at systems and national level (ie, policy makers and academic researchers) [43], there is a lack of research on decision makers closer to the implementation context and the health care setting such as primary care organization directors.

Aims and Objectives

This study has two main objectives: (1) identify and describe the primary care organizations providing ICBT in Sweden and (2) compare decision makers' (ie, directors of primary care organizations) views on barriers and facilitators for the implementation of ICBT in organizations that have implemented ICBT (ie, are offering ICBT [implementers]) and have not implemented ICBT (ie, are not offering ICBT [nonimplementers]). In implementation research, it is common to distinguish between diffusion (ie, passive spread of innovations), dissemination (ie, active efforts to convince an organization to adopt and innovation), implementation (ie, active efforts to offer an innovation and integrate it within the organization), and sustainability (ie, making an innovation part of routine care) [54]. The Swedish guidelines recommending ICBT are relatively new and thus it is unlikely ICBT would be part of routine care in organizations; the assumption is that those who offer ICBT are in the process of integrating ICBT. To this end, the term implementation is used to describe organizations that offer ICBT.

Methods

Study Design

An online self-report survey was conducted between February and May 2016 with decision makers in primary care organizations in Sweden. The Checklist for Reporting Results of Internet e-Surveys [55] was followed.

Setting

Sweden has 290 municipalities situated across 21 administrative units called regions, which are responsible for health care provision. Each region has several primary care organizations forming the basis of the Swedish health care system. According to the Swedish law on health care (Chapter 2, 6§), primary care is part of the open care system and should provide basic medical treatment, ongoing care, preventive measures, and rehabilitation in cases that do not require medical and technical resources or other specialized competence accessible at hospitals [56]. Primary care is thus the first point of care and from here patients can be referred to specialized care. A typical primary care organization employs medical doctors, nurses, physiotherapists, and psychologists and can thus treat many patients. The number of health care professionals employed at primary care organizations is related to the number of listed patients, which ranges from approximately 3000 to 30,000 per organization. Consequently, some primary care organizations have relatively few listed patients whereas others have very many.

Both private and public primary care organizations receive funding from regions and operate under the same conditions in terms of personnel competence, financial resources, opening hours, patient access, and adherence to national guidelines regarding care provision. For primary care organizations willing to implement ICBT, there are three main options available: (1) buying a license from a company for proprietary ICBT programs and delivering the support by themselves, (2) hiring a company to deliver ICBT including support, or (3) connecting to the Platform for Support and Care. The Platform for Support and Care is owned and run by a company (Inera) owned by the Swedish Association of Local Authorities and Regions (SALAR). The Platform for Support and Care offers ICBT programs developed by companies and/or research groups who in turn receive financial compensation when their programs are used. All programs delivered via the Platform for Support and Care undergo a careful procedure to ensure effectiveness and safety. When connecting to this platform, regions can hire companies to support ICBT or use their own therapists to support ICBT. Connecting to the Platform for Support and Care and gaining access to ICBT programs results in financial costs for the regions, such as costs for connecting to the platform and purchasing treatment programs either with or without therapist support. As such, access to ICBT programs and therapist support may differ depending on the financial resources of the health care organization. Further, access to ICBT comes at a financial cost, often per treatment contact with a therapist. For instance, in Stockholm each contact costs around €10 (US $11.22) per session.

Recruitment and Study Procedures

Decision makers were heads of Swedish primary care organizations identified via SALAR, which supports development and provision of health care, and the Inspectorate for Treatment and Care, responsible for monitoring Swedish health care. We compiled a list of 1156 primary care organizations and mailing addresses and sent invitations to the decision makers via regular mail (see Multimedia Appendix 1 for an English translation of the invitation). Invitations were sent to the organizations because we did not have a complete list of decision makers' email addresses. Invitations included a link to the online survey administered using the web-based tool, SurveyMonkey, full study information, and participation number and password allowing potential participants to reach the survey. Informed consent was provided online via SurveyMonkey, after which potential participants could access the survey. No incentives were offered for survey completion.

To maximize response rates, reminders were used [57]. Decision makers not completing the survey within two weeks received up to two telephone reminders and an email (except those whose email addresses were not available). The reminder email consisted of study information, the online link to the survey, participation number and password, and a reminder that the survey should be answered within two weeks. The reminder email was also accompanied by the letter previously sent through regular mail. After two weeks, nonrespondent decision makers were reminded once more following the same procedure.

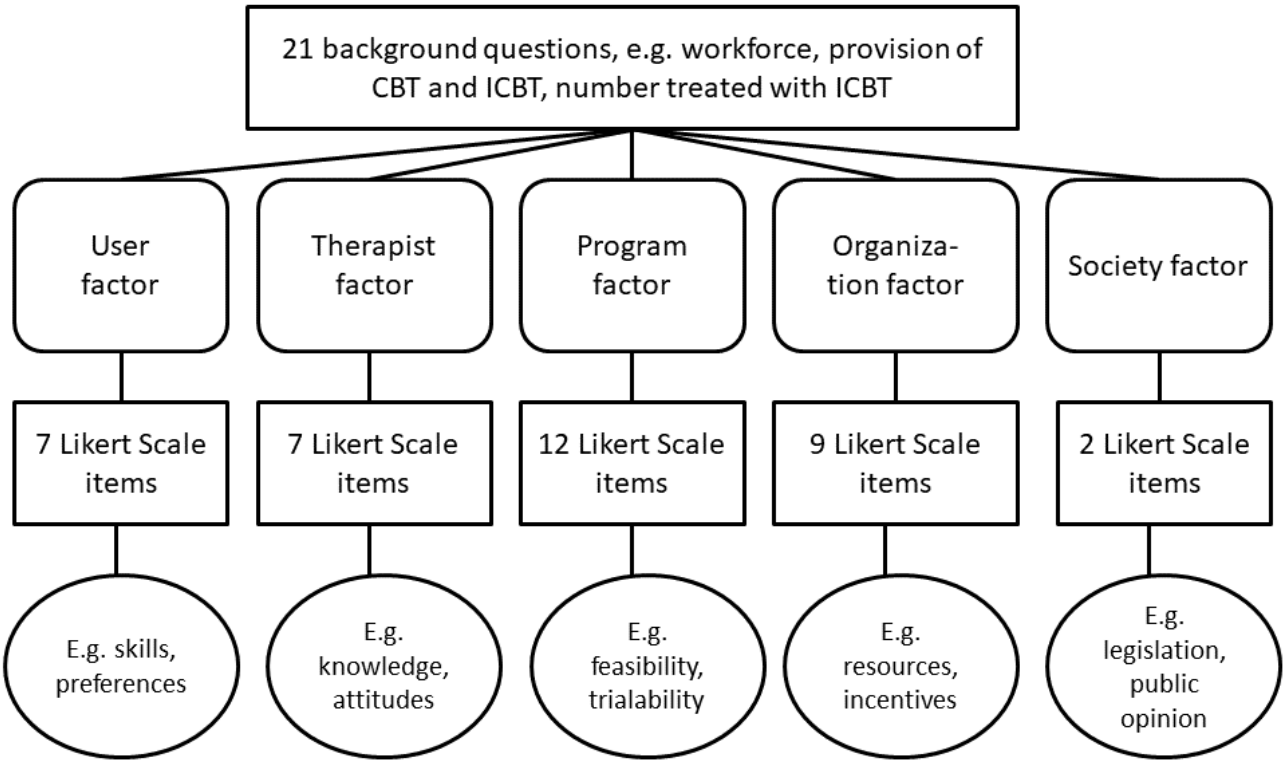
To avoid duplicate responses from decision makers, the SurveyMonkey function allowing only one answer per computer was selected. Respondents who did not complete the survey at one occasion were able to return to the survey at a later time. The survey could be submitted even if all items were not answered. Approval was granted by the ethical review board that reviews applications from Uppsala University, Sweden (application number 2015/461).

Measures

The study team developed an online survey consisting of 21 background questions and 37 items about 5 factors: (1) user, (2) therapist, (3) program, (4) organization, and (5) society, with a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree) and options for "do not know" and "do not wish to answer." The survey also included two open-ended questions about barriers and facilitators to implementation and space for additional comments. This qualitative data will be reported elsewhere. The survey was informed by the checklist of barriers and facilitators for improvement of health care practice by Flottorp et al [33] and modified based on knowledge of barriers and facilitators specific to ICBT implementation originating from studies on effectiveness and clinical feasibility available when the study was designed [20,21,38,58-60]. The survey was piloted by a panel of 8 persons with knowledge of ICBT who were not part of the study population, including former primary care directors and persons responsible for information technology and health care in certain Swedish regions. The SALAR database was used to recruit the panel. Panel members were asked to indicate whether the questions were easy to understand and to provide suggestions for improvements.

Overall, panel members were positive toward survey content but had some suggestions for improvements (see Multimedia Appendix 2 for a detailed description of the development and a summary of the changes). Figure 1 displays the final overall survey structure, and Multimedia Appendix 3 provides an English translation of the survey.

Figure 1. Survey structure.



Statistical Analysis

All calculations were performed in R version 3.5.2 (R Foundation for Statistical Computing).

Responders, Survey Reminders, and Reasons for Not Answering

Descriptive statistics were used to describe numbers and proportions concerning survey completeness, reminders, and reasons for not answering.

Characteristics of Study Sample and Organizational Characteristics

Study sample and organizational characteristics are described using descriptive statistics.

Representativeness of the Study Sample

As noncoverage (ie, the survey fails to include some parts of the population [61]) is common in survey research, the coverage of the obtained sample was examined. Data from three known background variables for the entire population were used (ie, organizational form, city size, and health care region) and described through descriptive statistics. A chi-square test for equality of proportions was used to examine the representativeness of the sample in terms of the entire population.

Barriers and Facilitators to Implementation of Internet-Administered Cognitive Behavioral Therapy

The potential difference regarding the number of agree answers between implementers and nonimplementers was examined using the chi-square test for equality of proportions. The chi-square tests included both the number of nonagree (Likert scale options 1, 2, 3, and 4) and agree (Likert scale options 5, 6, and 7) answers. It was assumed that if a decision maker answered that the organization did not offer CBT (Q6), they did not offer ICBT (Q12), and the respondent was therefore transferred directly from Q6 to Q22. To this end, nonimplementation was imputed for Q12 for the decision makers who answered that they did not offer CBT (Q6).

Results

Responders

A total of 1156 survey invitations were sent, of which 97.75% (1130/1156) were shown to be eligible. Noneligible answers that were excluded were duplicate answers (n=13),

bankruptcy/closed down (n=10), and not a primary care organization (n=3). A total of 426 decision makers answered at least the first survey question, providing an indication of interest to participate rate of 37.69% (426/1130). Participation [43] was defined as answering questions 1 to 22 (ie, background questions). The participation rate was 35.75% (404/1130). Completeness [55] was defined as completion of the last item (Q58). The completeness rate was 31.94% (361/1130).

Reminders

A total of 5.4% (22/404) of participants (based on participation rate) responded without a reminder, 47.5% (192/404) responded after being reminded via phone and email, 46.5% (188/404) responded after being reminded via email, and 0.4% (2/404) could not be reminded.

Reasons for Not Responding

Of the decision makers who did not answer the survey, 17.6% (128/726) provided reasons for not responding. The most frequent reason was no time (n=67), followed by not relevant for us (n=17) and decision maker employment ending (n=14).

Characteristics of the Study Sample and Organizational Characteristics

The majority (250/404, 61.8%) of decision makers were health care center directors with a background in nursing. A total of 89.8% (363/404) of organizations provided CBT, and 30.1% (122/404) had tried ICBT. Only 20.5% (83/404) reported that they currently offered ICBT and were accordingly categorized as ICBT implementers. Among the professionals delivering ICBT, 80% (67/83) were psychologists (including intern psychologists and psychotherapists) and 37% (31/83) were social workers. The majority (61/83, 73%) of organizations had 1 to 2 persons delivering ICBT. ICBT was offered to 1 to 10 patients during the last 12 months in 65% (54/83) of the organizations. Only 2% (2/83) of organizations had offered ICBT to over 100 persons during the last 12 months. Access to ICBT seemed to depend upon referral by a CBT therapist (36/83, 43%) or general practitioner (25/83, 30%) within the organization (for details on all collected characteristics see Multimedia Appendix 4).

Representativeness of the Sample

To evaluate the coverage of the sample, data on three known demographic variables for the entire population of primary care organizations in Sweden were used: organizational type (public/private), localization in cities of different sizes (city size), and health care regions. Table 1 displays the distribution of these variables in the study sample and entire population. Data indicates the study sample includes different organizational forms, city sizes, and health care regions. There was a significant difference between distribution of respondents and nonrespondent in terms organizational form (P<.001) and health care region (P<.001). There was no significant difference between distribution of respondents and nonrespondents regarding city size (P=.24). Consequently, the sample was representative in terms of city size but not concerning organizational form and health care region.

Table 1. Distribution of respondents in study sample and population.

Demographic	Sample (n=404), n (%)	Population, n/N (%)
Organizational form		
Private	131 (32.4)	131/465 (28.1)
Public	273 (67.5)	273/665 (41.0)
City size		
Small (<30,000)	148 (36.6)	148/382 (38.7)
Medium (30,000-100,000)	148 (36.6)	148/418 (35.4)
Large (>100,000)	108 (26.7)	108/330 (32.7)
Health care region[a]		
North	61 (15.0)	61/123 (49.5)
Uppsala-Örebro	85 (21.0)	85/231 (36.7)
Stockholm	64 (15.8)	64/208 (30.7)
West-East	57 (14.1)	57/130 (43.8)
South	68 (16.8)	68/222 (30.6)
West	69 (17.0)	69/216 (31.9)

[a]In Sweden, 21 regions are organized into 6 health care regions that facilitate cooperation and strategic work between the 21 regions.

Barriers and Facilitators to Implementation of Internet-Administered Cognitive Behavioral Therapy

Likert Scale Options and Item Nonresponse

Tables 2 to 6 display the numbers and percentages of decision makers, both implementers and nonimplementers, who agreed (Likert scale options 5, 6, and 7) with items Q22 through Q58. The survey options "do not know" and "do not wish to answer" are treated as item nonresponse and also shown in Tables 2 to 6 [61]. Item nonresponse varied from 22% to 66%.

Table 2. Implementers and nonimplementers agreeing to user-related items.

Variable	Item agreement and nonresponse			
	No ICBT[a] agree n/sample size (%)	ICBT agree n/sample size (%)	*P* value	Item nonresponse[b] (n=403)
Q22. Adults with depression and/or anxiety have the computer skills needed to use ICBT programs.	177/236 (75.0)	42/62 (67.7)	.32	105 (26.0)
Q23. Adults with depression and/or anxiety are capable of working on their own with ICBT programs.	151/217 (69.5)	39/59 (66.1)	.72	127 (31.5)
Q24. Adults with depression and/or anxiety have interest in ICBT programs.	71/156 (45.5)	29/60 (48.3)	.86	188 (46.6)
Q25. Adherence increases when treatment is delivered online to adults with depression and/or anxiety.	46/113 (40.7)	16/36 (44.4)	.84	254 (63.0)
Q26. The barrier to seek help for adults with depression and/or anxiety is decreased when care is online.	123/183 (67.2)	19/37 (51.3)	.10	183 (45.4)
Q27. Adults with depression and/or anxiety prefer to give confidential information to a computer.	51/144 (35.4)	7/35 (2.0)	.12	224 (55.5)
Q28. Adults with depression and/or anxiety in rural areas can be reached with ICBT programs.	208/243 (85.5)	53/64 (82.8)	.72	96 (23.8)

[a]ICBT: internet-administered cognitive behavioral therapy.

[b]Item nonresponse (do not know, do not wish to answer) of the entire sample (n=403).

Table 3. Implementers and nonimplementers agreeing to therapist-related items.

Variable	Item agreement and nonresponse			
	No ICBT[a] agree n/sample size (%)	ICBT agree n/sample size (%)	*P* value	Item nonresponse[b] (n=396)
Q29. Therapists treating adults with depression and/or anxiety are positive toward the ICBT programs.	108/166 (65.0)	57/71 (80.2)	.03	159 (40.1)
Q30. Therapists treating adults with depression and/or anxiety have knowledge of the ICBT programs.	125/192 (65.1)	68/76 (89.4)	.001	128 (32.3)
Q31. Therapists treating adults with depression and/or anxiety only need a little training in ICBT programs.	94/140 (67.1)	31/52 (59.6)	.42	204 (51.5)
Q32. Therapists treating adults with depression and/or anxiety have the computer skills needed for ICBT.	195/227 (85.9)	66/72 (91.6)	.28	97 (24.4)
Q33. Therapists treating adults with depression and/or anxiety have confidence in the guidelines recommending ICBT programs.	103/148 (69.5)	55/62 (88.7)	.001	186 (46.9)
Q34. Therapists treating adults with depression and/or anxiety can motivate patients to participate.	159/199 (79.8)	54/66 (81.8)	.87	131 (33.0)
Q35. More therapists support the introduction of ICBT programs than oppose it.	89/119 (74.7)	40/49 (81.6)	.45	228 (57.5)

[a]ICBT: internet-administered cognitive behavioral therapy.

[b]Item nonresponse (do not know, do not wish to answer) of the entire sample (n=396).

Table 4. Implementers and nonimplementers agreeing to program-related items.

Variable	Item agreement and nonresponse			
	No ICBT[a] agree n/sample size (%)	ICBT agree n/sample size (%)	*P* value	Item nonresponse[b] (n=389)
Q36. ICBT programs for adults with depression and/or anxiety should come with a help desk for therapists.	164/198 (82.8)	41/45 (91.1)	.25	146 (37.5)
Q37. ICBT programs are well suited for adults with depression and/or anxiety.	142/184 (77.1)	52/62 (83.8)	.37	143 (36.7)
Q38. ICBT programs for adults with depression and/or anxiety offer alternative learning formats.	77/101 (76.2)	15/32 (46.8)	.001	256 (65.8)
Q39. ICBT programs for adults with depression and/or anxiety are not plagued with big technical problems.	107/139 (76.9)	48/51 (94.1)	.001	199 (51.1)
Q40. It should be possible to trial the ICBT programs.	222/247 (89.8)	45/55 (81.8)	.15	87 (22.3)
Q41. It is possible to measure the effect on depression and/or anxiety when providing ICBT programs.	122/137 (89.0)	36/42 (85.7)	.75	210 (53.9)
Q42. ICBT programs for adults with depression and/or anxiety are easy to use.	78/95 (82.1)	46/52 (88.4)	.44	242 (62.2)
Q43. ICBT programs for adults with depression and/or anxiety can be integrated with the care structure.	161/208 (77.4)	63/73 (86.3)	.15	108 (27.7)
Q44. ICBT programs for adults with depression and/or anxiety can replace face-to-face CBT.	58/202 (28.7)	18/62 (29.0)	.99	125 (32.1)
Q45. It is easy to get access to ICBT programs for adults with depression and/or anxiety.	59/111 (53.1)	39/57 (68.4)	.08	221 (56.8)
Q46. ICBT programs for adults with depression and/or anxiety are well grounded on research evidence.	95/117 (81.1)	39/49 (79.5)	.98	223 (57.3)
Q47. GPs[c] referring adults with depression and/or anxiety to ICBT are positive toward ICBT programs.	88/130 (67.6)	30/46 (65.2)	.90	213 (54.7)

[a]ICBT: internet-administered cognitive behavioral therapy.

[b]Item nonresponse (do not know, do not wish to answer) of the entire sample (n=389).

[c]GPs: general practitioners.

Table 5. Implementers and nonimplementers agreeing to organization-related items.

Variable	Item agreement and nonresponse			
	No ICBT[a] agree n/sample size (%)	ICBT agree n/sample size (%)	*P* value	Item nonresponse[b] (n=381)
Q48. Our organization has resources to offer ICBT programs to adults with depression and/or anxiety.	103/215 (47.9)	53/77 (68.8)	.001	89 (23.3)
Q49. ICBT programs can decrease care costs in treatment of adults with depression and/or anxiety.	157/196 (80.1)	42/57 (73.6)	.39	128 (33.5)
Q50. Existing information management systems allows administration of patients enrolled in ICBT.	57/105 (54.2)	34/54 (62.9)	.38	222 (58.2)
Q51. Existing quality assurance and patient safety systems are compatible with the requirements to offer ICBT to adults with depression and/or anxiety.	56/85 (65.8)	35/44 (79.5)	.16	252 (66.1)
Q52. Existing continuing education systems of therapists are compatible with the training to introduce ICBT to adults with depression and/or anxiety.	58/91 (63.7)	31/44 (70.4)	.56	246 (64.5)
Q53. Internal regulations allow introduction of ICBT for adults with depression and/or anxiety.	123/152 (80.9)	64/71 (90.1)	.12	158 (41.4)
Q54. Contracts with service providers allow introduction of ICBT for adults with depression and/or anxiety.	59/89 (66.2)	41/47 (87.2)	.02	245 (64.3)
Q55. The concept of online treatment to adults with depression and/or anxiety is well established at our organization.	35/226 (15.4)	36/72 (50.0)	.001	83 (21.7)
Q56. The patient referral process allows introduction of ICBT for adults with depression and/or anxiety.	87/164 (53.0)	58/66 (87.8)	.001	151 (39.6)

[a]ICBT: internet-administered cognitive behavioral therapy.

[b]Item nonresponse (do not know, do not wish to answer) of the entire sample (n=381).

Table 6. Implementers and nonimplementers agreeing to society-related items.

Variable	Item agreement and nonresponse			
	No ICBT[a] agree n/sample size (%)	ICBT agree n/sample size (%)	*P* value	Item nonresponse[b] (n=361)
Q57. Legislation does not hinder the introduction of ICBT for adults with depression and/or anxiety.	128/143 (89.5)	48/53 (90.5)	.99	165 (45.7)
Q58. Public opinion supports the introduction of online treatments for adults with depression and/or anxiety.	91/123 (73.9)	26/37 (70.2)	.81	201 (55.6)

[a]ICBT: internet-administered cognitive behavioral therapy.

[b]Item nonresponse (do not know, do not wish to answer) of the entire sample (n=361).

Users

There were no significant differences ($P<.05$) between implementers and nonimplementers in terms of agreeing to user-related items (Table 2).

Therapists

Three therapist-related items were significant ($P<.05$; Table 3). More implementers (57/71, 80%) than nonimplementers (108/166, 65.0%) believed that therapists treating adults with depression and/or anxiety are positive toward the ICBT programs ($P=.03$; Q29). More implementers (68/76, 89%) than nonimplementers (125/192, 65.1%) believed therapists treating adults with depression and/or anxiety have knowledge of the ICBT programs ($P<.001$; Q30). More implementers (55/62, 88%) than nonimplementers (103/148, 69.5%) also believed therapists treating adults with depression and/or anxiety have confidence in the guidelines recommending ICBT programs ($P<.001$; Q33).

Program

Two program-related items were significant ($P<.05$; Table 4). More nonimplementers (77/101, 76.2%) than implementers (15/32, 46%) believed that ICBT programs for adults with depression and/or anxiety offer alternative learning formats ($P<.001$; Q38). More implementers (48/51, 94%) than nonimplementers (107/139, 76.9%) in turn believed that ICBT programs for adults with depression and/or anxiety are not plagued with big technical problems ($P<.001$; Q39).

Organization

Four organization-related items were significant ($P<.05$; Table 5). More implementers (53/77, 68%) than nonimplementers (103/215, 47.9%) considered that they have resources to offer

ICBT programs (*P*<.001; Q48). More implementers (41/47, 87%) than nonimplementers (59/89, 66%) believed that contracts with service providers allow introduction of ICBT (*P*=.02; Q54). More implementers (36/72, 50%) than nonimplementers (35/226, 15.4%) considered that the concept of online treatment is well established at their organization (*P*<.001; Q55). More implementers (58/66, 87%) than nonimplementers (87/164, 53.0%) believed the patient referral process allows for the introduction of ICBT (*P*<.001; Q56).

Society

There were no significant differences (*P*<.05) between implementers and nonimplementers in terms of society-related items (Table 6).

Discussion

Principal Findings

The participation rate was 35.75% (404/1130). The majority (250/404, 61.8%) of participants were health care center directors with a background in nursing. A total of 89.9% (363/404) of the participating organizations provided CBT. A minority (83/404, 20.5%) provided ICBT and thus were implementers. In general, psychologists (67/83, 80%) and social workers (31/83, 37%) delivered ICBT to patients, and the majority (61/83, 73%) of the organizations had 1 to 2 persons delivering ICBT. There were 9 significant (*P*<.05) differences out of 37 possible between implementers and nonimplementers. For example, more implementers (68/76, 89%) than nonimplementers (125/192, 65.1%) considered that therapists treating adults with depression and/or anxiety have knowledge of the ICBT programs (*P*<.001), and more implementers (58/66, 87%) than nonimplementers (87/164, 53.0%) believed the patient referral process allows the introduction of ICBT (*P*<.001).

Although e-mental health initiatives are encouraged, and the Swedish national guidelines include provision of ICBT in Swedish primary care among their recommendations, few organizations in our sample provided ICBT. Moreover, in a majority of cases only 1 to 10 patients had been treated with ICBT during the last 12 months, and the majority of organizations only had 1 to 2 CBT therapists working with ICBT. The number of ICBT implementers is not as high as one would expect based on the strong recommendations and claimed benefits of ICBT, such as increased patient access and lower costs [21,22]. Previous research indicates the implementation rate of guidelines across various treatment areas has been between 50% and 70% [62-65]. However, implementation research does not provide detailed guidance concerning whether a 20% implementation rate could be considered low or high. Further, the implementation of a complex health care intervention, such as ICBT, is more demanding than implementing clinical guidelines in general. Indeed, our findings are in line with a recent study examining ICBT implementation among psychologists in the Netherlands where the implementation rate was around 16% [66].

One reason for the low implementation rate may pertain to not having a workforce who can support the provision of ICBT. Indeed, the majority of implementing organizations had only 1 to 2 therapists working with ICBT and CBT. Moreover, all organizations providing ICBT also provided CBT, and in cases ICBT and CBT rely on the same workforce, this workforce could be inadequate to support ICBT implementation. Further, CBT therapists may not represent the best solution, given they are a highly trained and expensive workforce, with a demand exceeding supply [67]. A possible solution could be to train mental health workers to support ICBT, as done by the Improving Access to Psychological Therapies program in England, whereby a workforce of psychological well-being practitioners was established to deliver low-intensity CBT [67]. Future research could explore the workforce aspect in Sweden, and current findings could be helpful for policymakers trying to promote the implementation of ICBT.

Our findings show that 94% (48/51) of implementers consider that ICBT is not plagued with technical problems, whereas only 76.9% (107/139) of nonimplementers share this opinion (Q39). Existing research on implementation is clear regarding technical problems, showing innovations low in complexity to be more easily adopted [68-70]. Another interesting finding is that 50% (36/72) of implementers believed that the concept of online treatment regarding depression and/or anxiety is well established at their organization compared with 15.4% (35/226) of nonimplementers who believed this be the case (Q55). The survey did not define the word concept, and thus this could have been interpreted differently by decision makers. However, as relatively many implementers compared with nonimplementers perceived the concept to be well established there seems to be a clear pattern in how this item is perceived by the decision makers. One possible reason for this finding could be that implementation of ICBT may increase understanding of online treatments and thus strengthen the concept.

Another interesting finding is that the majority of implementers (53/77, 68%) consider that they have resources to offer ICBT programs compared with a minority among nonimplementers (103/215, 47.9%; Q48), which indicates lack of resources may be an important barrier to implementation of ICBT in Swedish primary care. This factor is identified as central in existing ICBT implementation research [34] and in research concerning implementation of eHealth [45,71]. Findings also show that 80% (57/71) of implementers consider therapists treating adults with depression and/or anxiety are positive toward the ICBT programs compared with 65.0% (108/166) of nonimplementers (Q29). This finding implies therapist acceptance could be an important facilitator to implementation, in line with existing ICBT implementation research [21,60] and research concerning e-mental health [42,44]. Findings also demonstrate that 87% (41/47) of implementers believe contracts with service providers allow for the introduction of ICBT compared with 66% (59/89) of nonimplementers (Q54). Indeed, contracts with service providers are included in the framework regarding barriers and facilitators to implementation by Flottorp et al [33]. As such, our findings may imply lack of contracts with service providers could be a sizeable barrier to implementation of ICBT in Sweden.

While 28 barriers and facilitators to implementation did not show significant differences between implementers and

nonimplementers, they still provide important findings. The majority (219/298, 73.4%) of decision makers considered users to possess the computer skills needed to use ICBT programs. Further, the majority (261/299, 87.2%) considered therapists to have the computer skills required to offer ICBT. Lack of computer skills among patients and therapists was found to be a barrier to implementation in wider research concerning implementation of e-mental health [43]. Further, therapists' computer skills is reported to be a facilitator to implementation of eHealth [45]. Findings also show that the majority (134/166, 80.7%) of decision makers considered ICBT programs to be evidence-based. This factor is a key facilitator in ICBT implementation [34,35] and an important barrier, if absent, in e-mental health [43] and eHealth implementation [47,48]. Further, the majority (199/253, 78.6%) of decision makers considered that ICBT can decrease costs related to providing care, which is also suggested in existing studies concerning eHealth programs [46,47].

Limitations

First, this study aimed to cover all primary care organizations in Sweden. However, recruitment of decision makers was difficult and thus the study only includes a limited sample of primary care organizations. However, participation rate was 35.75% (404/1130), which is in line with a recent study examining the implementation of ICBT from therapists' point of view [66] and the sample was representative of the population on city size. Second, we cannot exclude the possibility that people who were interested in ICBT were more inclined to respond and thus our sample may include a larger proportion of implementers than nonimplementers compared with the entire population. Third, we focused on decision makers in Swedish primary care, who are often not experts in ICBT. However, decision makers decide whether or not to implement and thus their perceptions of barriers and facilitators are key for understanding the chances of successful implementation. Future studies may also include the views of Swedish primary care therapists representing implementing organizations and could compare the views of therapists with those of decision makers. Fourth, this study presents results of a survey focusing on a set of factors and whether respondents agreed on their importance on a quantitative scale but not to what extent each factor represented a barrier or facilitator to the implementation process. Future studies combining both survey and semistructured interview data may provide a more detailed exploration of barriers and facilitators to implementation and aid interpretation of survey findings.

Fifth, the survey did not explore whether implementing organizations used specific implementation strategies associated with improved changes to existing practice [29]—for example, opinion leaders or audit and feedback to facilitate implementation [72]. Nor did the survey explore to what extent decision makers considered ICBT implementation a success (eg, in terms of increased ICBT use, costs, and resources). Future research into barriers and facilitators of ICBT implementation may examine ICBT implementation strategies used and the perceived success of these strategies. Sixth, our results are influenced by multiple testing: with alpha .05 and 37 tests, 2 of the 9 differences are likely to be produced by chance alone. Seventh, nonresponse per item was relatively high (varied between 22% and 66%), which could impact the validity of the results. However, we observe that nonresponse is relatively evenly distributed between items and thus a reasonable assumption is that nonresponse does not impact validity.

Despite the aforementioned limitations, to our knowledge, this is the first study to provide data on the implementation of ICBT in one of the countries that pioneered ICBT research and treatment, elucidating the views of decision makers on the implementation of ICBT in primary care.

Conclusions

Despite existing scientific evidence supporting implementation of ICBT in primary care and guidelines recommending implementation, most primary care organizations in Sweden still only offer traditional CBT. This study provides an overview of the characteristics of implementers in one of the countries that pioneered research in ICBT, identifying interesting differences in terms of perceived barriers and facilitators between implementers and nonimplementers that may inform future implementation interventions to improve the routine uptake of ICBT in Swedish primary care but also in other countries introducing ICBT.

Acknowledgments

The authors thank all decision makers in primary care who participated in the study. They also thank SALAR for their advice concerning ensuring the survey was more suitable for decision makers in Swedish primary care. The authors also thank several colleagues for their input and support. This work was supported by U-CARE, which is a strategic research environment funded by the Swedish Research Council (dnr 2009-1093). The funder did not contribute to design and conducting of the study and did not participate in analysis of the results.

Conflicts of Interest

None declared.

Multimedia Appendix 1
English translation of invitation letter used to recruit respondents.
[DOCX File , 79 KB - jmir_v22i8e18033_app1.docx]

Multimedia Appendix 2
Survey development.
[DOCX File , 29 KB - jmir_v22i8e18033_app2.docx]

Multimedia Appendix 3
English translation of the online questionnaire as it was presented to the respondents.
[DOCX File , 28 KB - jmir_v22i8e18033_app3.docx]

Multimedia Appendix 4
Characteristics of the study sample and organizational characteristics.
[DOCX File , 26 KB - jmir_v22i8e18033_app4.docx]

References

1. Depression and other common mental disorders: global health estimates. Geneva: World Health Organization; 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf?sequence=1 [accessed 2020-06-17]
2. Lewis C, Pearce J, Bisson JI. Efficacy, cost-effectiveness and acceptability of self-help interventions for anxiety disorders: systematic review. Br J Psychiatry 2012 Jan;200(1):15-21 [FREE Full text] [doi: 10.1192/bjp.bp.110.084756] [Medline: 22215865]
3. Stuhldreher N, Leibing E, Leichsenring F, Beutel ME, Herpertz S, Hoyer J, et al. The costs of social anxiety disorder: the role of symptom severity and comorbidities. J Affect Disord 2014 Aug;165:87-94. [doi: 10.1016/j.jad.2014.04.039] [Medline: 24882183]
4. Newman MG, Llera SJ, Erickson TM, Przeworski A, Castonguay LG. Worry and generalized anxiety disorder: a review and theoretical synthesis of evidence on nature, etiology, mechanisms, and treatment. Annu Rev Clin Psychol 2013;9:275-297 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050212-185544] [Medline: 23537486]
5. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJL, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 2013 Nov;10(11):e1001547 [FREE Full text] [doi: 10.1371/journal.pmed.1001547] [Medline: 24223526]
6. Cuijpers P, Cristea IA, Karyotaki E, Reijnders M, Huibers MJH. How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. World Psychiatry 2016 Oct;15(3):245-258 [FREE Full text] [doi: 10.1002/wps.20346] [Medline: 27717254]
7. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. Can J Psychiatry 2013 Jul;58(7):376-385. [Medline: 23870719]
8. Olatunji BO, Cisler JM, Deacon BJ. Efficacy of cognitive behavioral therapy for anxiety disorders: a review of meta-analytic findings. Psychiatr Clin North Am 2010 Sep;33(3):557-577. [doi: 10.1016/j.psc.2010.04.002] [Medline: 20599133]
9. Hofmann SG, Asnaani A, Vonk IJJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012 Oct 1;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
10. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al. Undertreatment of people with major depressive disorder in 21 countries. Br J Psychiatry 2017 Dec;210(2):119-124 [FREE Full text] [doi: 10.1192/bjp.bp.116.188078] [Medline: 27908899]
11. Alonso J, Liu Z, Evans-Lacko S, Sadikova E, Sampson N, Chatterji S, WHO World Mental Health Survey Collaborators. Treatment gap for anxiety disorders is global: results of the World Mental Health Surveys in 21 countries. Depress Anxiety 2018 Mar;35(3):195-208 [FREE Full text] [doi: 10.1002/da.22711] [Medline: 29356216]
12. Kazdin AE, Blase SL. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
13. Riper H, Andersson G, Christensen H, Cuijpers P, Lange A, Eysenbach G. Theme issue on e-mental health: a growing field in internet research. J Med Internet Res 2010;12(5):e74 [FREE Full text] [doi: 10.2196/jmir.1713] [Medline: 21169177]
14. Musiat P, Tarrier N. Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. Psychol Med 2014 Nov;44(15):3137-3150. [doi: 10.1017/S0033291714000245] [Medline: 25065947]
15. Andersson G. Internet interventions: past, present and future. Internet Interv 2018 Jun;12:181-188 [FREE Full text] [doi: 10.1016/j.invent.2018.03.008] [Medline: 30135782]
16. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]

17. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-869. [doi: 10.1586/ern.12.63] [Medline: 22853793]
18. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
19. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
20. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
21. Andersson G. The promise and pitfalls of the internet for cognitive behavioral therapy. BMC Med 2010;8:82 [FREE Full text] [doi: 10.1186/1741-7015-8-82] [Medline: 21138574]
22. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
23. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
24. Sjöström J, von Essen L, Grönqvist H. The origin and impact of ideals in ehealth research: experiences from the U-CARE research environment. JMIR Res Protoc 2014 May 23;3(2):e28 [FREE Full text] [doi: 10.2196/resprot.3202] [Medline: 24860071]
25. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
26. Socialstyrelsen. Nationella Riktlinjer för Vård vid Depression och Ångestsyndrom. Stockholm: Socialstyrelsen; 2010. URL: http://www.sfbup.se/wp-content/uploads/2016/11/2010-3-4.pdf [accessed 2020-06-22]
27. Socialstyrelsen. Nationella Riktlinjer för Vård vid Depression och Ångestsyndrom. Stockholm: Socialstyrelsen; 2017. URL: https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/nationella-riktlinjer/2017-12-4.pdf [accessed 2020-06-22]
28. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. Health Technol Assess 2004 Feb;8(6):1-72 [FREE Full text] [Medline: 14960256]
29. Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. Implement Sci 2012 May 31;7:50 [FREE Full text] [doi: 10.1186/1748-5908-7-50] [Medline: 22651257]
30. Wang Z, Norris SL, Bero L. The advantages and limitations of guideline adaptation frameworks. Implement Sci 2018 May 29;13(1):72 [FREE Full text] [doi: 10.1186/s13012-018-0763-4] [Medline: 29843737]
31. Liang L, Bernhardsson S, Vernooij RWM, Armstrong MJ, Bussières A, Brouwers MC, Members of the Guidelines International Network Implementation Working Group. Use of theory to plan or evaluate guideline implementation among physicians: a scoping review. Implement Sci 2017 Feb 27;12(1):26 [FREE Full text] [doi: 10.1186/s13012-017-0557-0] [Medline: 28241771]
32. DeJonckheere M, Robinson CH, Evans L, Lowery J, Youles B, Tremblay A, et al. Designing for clinical change: creating an intervention to implement new statin guidelines in a primary care clinic. JMIR Hum Factors 2018 Apr 24;5(2):e19 [FREE Full text] [doi: 10.2196/humanfactors.9030] [Medline: 29691206]
33. Flottorp SA, Oxman AD, Krause J, Musila NR, Wensing M, Godycki-Cwirko M, et al. A checklist for identifying determinants of practice: a systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice. Implement Sci 2013 Mar 23;8:35 [FREE Full text] [doi: 10.1186/1748-5908-8-35] [Medline: 23522377]
34. Hadjistavropoulos HD, Nugent MM, Dirkse D, Pugh N. Implementation of internet-delivered cognitive behavior therapy within community mental health clinics: a process evaluation using the consolidated framework for implementation research. BMC Psychiatry 2017 Sep 12;17(1):331 [FREE Full text] [doi: 10.1186/s12888-017-1496-7] [Medline: 28899365]
35. Titov N, Dear B, Nielssen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: a descriptive analysis of successful clinics in five countries. Internet Interv 2018 Sep;13:108-115 [FREE Full text] [doi: 10.1016/j.invent.2018.07.006] [Medline: 30206525]
36. Folker AP, Mathiasen K, Lauridsen SM, Stenderup E, Dozeman E, Folker MP. Implementing internet-delivered cognitive behavior therapy for common mental health disorders: a comparative case study of implementation challenges perceived by therapists and managers in five European internet services. Internet Interv 2018 Mar;11:60-70 [FREE Full text] [doi: 10.1016/j.invent.2018.02.001] [Medline: 30135761]
37. Nordgreen T, Gjestad R, Andersson G, Carlbring P, Havik OE. The implementation of guided Internet-based cognitive behaviour therapy for panic disorder in a routine-care setting: effectiveness and implementation efforts. Cogn Behav Ther 2018 Jan;47(1):62-75. [doi: 10.1080/16506073.2017.1348389] [Medline: 28714775]

38. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
39. Hedman E, Ljótsson B, Rück C, Bergström J, Andersson G, Kaldo V, et al. Effectiveness of internet-based cognitive behaviour therapy for panic disorder in routine psychiatric care. Acta Psychiatr Scand 2013 Dec;128(6):457-467. [doi: 10.1111/acps.12079] [Medline: 23406572]
40. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
41. Hedman E, Ljótsson B, Kaldo V, Hesser H, Kraepelien M, Andersson E, et al. Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. J Affect Disord 2014 Feb;155:49-58. [doi: 10.1016/j.jad.2013.10.023] [Medline: 24238951]
42. Feijt MA, de Kort YA, Bongers IM, IJsselsteijn WA. Perceived drivers and barriers to the adoption of emental health by psychologists: the construction of the levels of adoption of emental health model. J Med Internet Res 2018 Dec 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]
43. Wozney L, Newton AS, Gehring ND, Bennett K, Huguet A, Hartling L, et al. Implementation of eMental Health care: viewpoints from key informants from organizations and agencies with eHealth mandates. BMC Med Inform Decis Mak 2017 Jun 02;17(1):78 [FREE Full text] [doi: 10.1186/s12911-017-0474-9] [Medline: 28577543]
44. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving implementation of emental health for mood disorders in routine practice: systematic review of barriers and facilitating factors. JMIR Ment Health 2018 Mar 16;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
45. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
46. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
47. Mileski M, Kruse CS, Catalani J, Haderer T. Adopting telemedicine for the self-management of hypertension: systematic review. JMIR Med Inform 2017 Oct 24;5(4):e41 [FREE Full text] [doi: 10.2196/medinform.6603] [Medline: 29066424]
48. Ariens LF, Schussler-Raymakers FM, Frima C, Flinterman A, Hamminga E, Arents BW, et al. Barriers and facilitators to ehealth use in daily practice: perspectives of patients and professionals in dermatology. J Med Internet Res 2017 Sep 05;19(9):e300 [FREE Full text] [doi: 10.2196/jmir.7512] [Medline: 28874336]
49. Lennon MR, Bouamrane M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a national digital health innovation program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]
50. Swinkels ICS, Huygens MWJ, Schoenmakers TM, Oude Nijeweme-D'Hollosy W, van Velsen L, Vermeulen J, et al. Lessons learned from a living lab on the broad adoption of ehealth in primary health care. J Med Internet Res 2018 Mar 29;20(3):e83 [FREE Full text] [doi: 10.2196/jmir.9110] [Medline: 29599108]
51. van der Meij E, Huirne JA, Ten Cate AD, Stockmann HB, Scholten PC, Davids PH, et al. A perioperative ehealth program to enhance postoperative recovery after abdominal surgery: process evaluation of a randomized controlled trial. J Med Internet Res 2018 Jan 02;20(1):e1 [FREE Full text] [doi: 10.2196/jmir.8338] [Medline: 29295808]
52. Buis L. Implementation: the next giant hurdle to clinical transformation with digital health. J Med Internet Res 2019 Nov 20;21(11):e16259 [FREE Full text] [doi: 10.2196/16259] [Medline: 31746763]
53. Kukafka R. Digital health consumers on the road to the future. J Med Internet Res 2019 Nov 21;21(11):e16359 [FREE Full text] [doi: 10.2196/16359] [Medline: 31750835]
54. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
55. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
56. Socialdepartementet. Hälso-och Sjukvårdslag SFS 2017:30. Stockholm: Socialdepartementet; 2017. URL: https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/halso--och-sjukvardslag_sfs-2017-30 [accessed 2020-06-22]
57. Shih T, Xitao Fan. Comparing response rates from web and mail surveys: a meta-analysis. Field Methods 2008 Mar 03;20(3):249-271. [doi: 10.1177/1525822X08317085]
58. Andrews G, Titov N. Hit and miss: innovation and the dissemination of evidence based psychological treatments. Behav Res Ther 2009 Nov;47(11):974-979. [doi: 10.1016/j.brat.2009.07.007] [Medline: 19625014]
59. Gega L, Marks I, Mataix-Cols D. Computer-aided CBT self-help for anxiety and depressive disorders: experience of a London clinic and future directions. J Clin Psychol 2004 Feb;60(2):147-157. [doi: 10.1002/jclp.10241] [Medline: 14724922]

60. Titov N. Status of computerized cognitive behavioural therapy for adults. Aust N Z J Psychiatry 2007 Feb;41(2):95-114. [doi: 10.1080/00048670601109873] [Medline: 17464688]
61. Brick JM, Kalton G. Handling missing data in survey research. Stat Methods Med Res 1996 Sep;5(3):215-238. [doi: 10.1177/096228029600500302] [Medline: 8931194]
62. Grol R. Successes and failures in the implementation of evidence-based guidelines for clinical practice. Med Care 2001 Aug;39(8 Suppl 2):II46-1154. [doi: 10.1097/00005650-200108002-00003] [Medline: 11583121]
63. Gustafsson I, Hvelplund A, Hansen KW, Galatius S, Madsen M, Jensen JS, et al. Underuse of an invasive strategy for patients with diabetes with acute coronary syndrome: a nationwide study. Open Heart 2015;2(1):e000165 [FREE Full text] [doi: 10.1136/openhrt-2014-000165] [Medline: 25685362]
64. Hulscher ME, Grol RP, van der Meer JW. Antibiotic prescribing in hospitals: a social and behavioural scientific approach. Lancet Infect Dis 2010 Mar;10(3):167-175. [doi: 10.1016/S1473-3099(10)70027-X] [Medline: 20185095]
65. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. N Engl J Med 2003 Jun 26;348(26):2635-2645. [doi: 10.1056/NEJMsa022615] [Medline: 12826639]
66. IJzerman RVH, van der Vaart R, Evers AWM. Internet-based cognitive behavioral therapy among psychologists in a medical setting: a survey on implementation. J Med Internet Res 2019 Aug 09;21(8):e13432 [FREE Full text] [doi: 10.2196/13432] [Medline: 31400101]
67. Richards DA, Ekers D, McMillan D, Taylor RS, Byford S, Warren FC, et al. Cost and Outcome of Behavioural Activation versus Cognitive Behavioural Therapy for Depression (COBRA): a randomised, controlled, non-inferiority trial. Lancet 2016 Aug 27;388(10047):871-880 [FREE Full text] [doi: 10.1016/S0140-6736(16)31140-0] [Medline: 27461440]
68. Grilli R, Lomas J. Evaluating the message: the relationship between compliance rate and the subject of a practice guideline. Med Care 1994 Mar;32(3):202-213. [doi: 10.1097/00005650-199403000-00002] [Medline: 8145598]
69. Rogers E. Diffusion of Innovations. London: Free Press; 2003.
70. Denis J, Hébert Y, Langley A, Lozeau D, Trottier L. Explaining diffusion patterns for complex health care innovations. Health Care Manage Rev 2002;27(3):60-73. [doi: 10.1097/00004010-200207000-00007] [Medline: 12146784]
71. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of ehealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
72. Kok G, Gottlieb NH, Peters GY, Mullen PD, Parcel GS, Ruiter RAC, et al. A taxonomy of behaviour change methods: an Intervention Mapping approach. Health Psychol Rev 2015 Oct 15:1-16. [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]

Abbreviations

CBT: cognitive behavioral therapy
ICBT: internet-administered cognitive behavioral therapy
SALAR: Swedish Association of Local Authorities and Regions

Edited by G Eysenbach; submitted 29.01.20; peer-reviewed by H Hadjistavropoulos, D Austin, K Mathiasen; comments to author 18.03.20; revised version received 11.05.20; accepted 13.05.20; published 12.08.20.

Please cite as:
Brantnell A, Woodford J, Baraldi E, van Achterberg T, von Essen L
Views of Implementers and Nonimplementers of Internet-Administered Cognitive Behavioral Therapy for Depression and Anxiety: Survey of Primary Care Decision Makers in Sweden
J Med Internet Res 2020;22(8):e18033
URL: https://www.jmir.org/2020/8/e18033
doi:10.2196/18033
PMID:32784186

Original Paper

Clinical Genomic Sequencing Reports in Electronic Health Record Systems Based on International Standards: Implementation Study

Borim Ryu[1*], MSc; Soo-Yong Shin[2,3*], PhD; Rong-Min Baek[4], MD, PhD; Jeong-Whun Kim[5], MD, PhD; Eunyoung Heo[1], MSc; Inchul Kang[6], BSc; Joshua SungWoo Yang[7], PhD; Sooyoung Yoo[1], PhD

[1]Office of eHealth and Business, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

[2]Department of Digital Health, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul, Republic of Korea

[3]Big Data Research Center, Samsung Medical Center, Seoul, Republic of Korea

[4]Department of Plastic and Reconstructive Surgery, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

[5]Department of Otorhinolaryngology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

[6]Research and Development Center, ezCaretech, Seoul, Republic of Korea

[7]Big Data Center, Macrogen, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Sooyoung Yoo, PhD
Office of eHealth and Business
Seoul National University Bundang Hospital
82 173rd Street
Gumi-ro, Bundang-gu
Seongnam, 136036
Republic of Korea
Phone: 82 1090537094
Email: yoosoo0@snubh.org

Abstract

Background: To implement standardized machine-processable clinical sequencing reports in an electronic health record (EHR) system, the International Organization for Standardization Technical Specification (ISO/TS) 20428 international standard was proposed for a structured template. However, there are no standard implementation guidelines for data items from the proposed standard at the clinical site and no guidelines or references for implementing gene sequencing data results for clinical use. This is a significant challenge for implementation and application of these standards at individual sites.

Objective: This study examines the field utilization of genetic test reports by designing the Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR) for genomic data elements based on the ISO/TS 20428 standard published as the standard for genomic test reports. The goal of this pilot is to facilitate the reporting and viewing of genomic data for clinical applications. FHIR Genomics resources predominantly focus on transmitting or representing sequencing data, which is of less clinical value.

Methods: In this study, we describe the practical implementation of ISO/TS 20428 using HL7 FHIR Genomics implementation guidance to efficiently deliver the required genomic sequencing results to clinicians through an EHR system.

Results: We successfully administered a structured genomic sequencing report in a tertiary hospital in Korea based on international standards. In total, 90 FHIR resources were used. Among 41 resources for the required fields, 26 were reused and 15 were extended. For the optional fields, 28 were reused and 21 were extended.

Conclusions: To share and apply genomic sequencing data in both clinical practice and translational research, it is essential to identify the applicability of the standard-based information system in a practical setting. This prototyping work shows that reporting data from clinical genomics sequencing can be effectively implemented into an EHR system using the existing ISO/TS 20428 standard and FHIR resources.

(J Med Internet Res 2020;22(8):e15040) doi:10.2196/15040

KEYWORDS

standardization; genomics; electronic health record; information system; data exchange

Introduction

In order to realize the potential of personalized precision medicine through genetic testing, the results of genetic testing must be integrated with other clinical information to improve patient care. Because the electronic health records (EHRs) of today seem ill suited for managing genomic data, other solutions are required, including universal data standards and applications that use application programming interfaces (APIs), along with a commitment on the part of sequencing laboratories to consistently provide structured genomic data for clinical use [1]. In addition, a clinical interface is needed for genetic test reporting that provides genetic evidence as support for clinical decision making. Many standards for genetic test reporting have been presented so far, but they often lack guidance for specific data items and realistic implementation guidelines. Tarczy-Hornoch et al reported that the parallel efforts across clinical sequencing exploratory research sites in the creation of systems for report generation and integration of reports into the EHR system, as well as the lack of standardized approaches to interfacing with variant databases to create active clinical decision support, create opportunities for cross-site and vendor collaborations [2-4]. This has prevented clinical implementation of these standards. Nonetheless, these standards offer a potential foundation for the creation of clinically relevant standards. ISO/TC 215 is the International Organization for Standardization's (ISO) Technical Committee (TC) on health informatics. TC 215 works on the standardization of health information and communications technology (ICT) to ensure compatibility and interoperability between independent systems. ISO also published two genomics standards [5,6] to exchange genomic sequence variation in XML format and to incorporate a structured clinical sequencing report to EHRs. In addition, many other genomics standards for whole genome sequencing, gene fusion reports, and omics data are now under development [7-11]. ISO 25720:2009 is a standard that defines how gene sequence variation information is presented in an XML-based data exchange format [5,12]. This standard enables the transmission and reception of genetic variation information around the world. The scope of ISO 25720:2009 is the data exchange format, but notably it does not include a database schema. From a biological point of view, single nucleotide polymorphisms are the main target of this standard. ISO/TS (Technical Specification) 22220:2011 defines data elements and structures in the medical sector to ensure accurate personal identification through calculation [13]. This standard also defines statistical and other identification data elements to facilitate the improved identification of patients at medical institutions. Finally, ISO/TS 20428 is a standard for defining structured techniques so that genetic screening reports can be effectively communicated and utilized by the clinical workforce [6]. The standard is called *Health informatics— Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records*, which proposes standard formatting for the results of next generation sequencing (NGS) genetic testing of patient samples in electronic medical records. These standards are being employed by various domestic companies and hospitals. For example, a recent study by Park et al showed the implementation of ISO/TS 20428 standard and developed a clinical research information system to integrate patient NGS data with clinical data [14].

Despite these standards, most hospitals in Korea produce genetic test reports in text or PDF format and provide them to medical staff. Furthermore, detailed examination results are managed in different formats by different agencies. To overcome these discrepancies, standards for utilizing genetic information at medical sites should be developed, and an analysis pipeline should be created where data can be freely processed in the form that meets the needs of the site. Of note, it is difficult to apply the current standards to actual sites because the previously proposed ISO/TS 20428 standard contains only definitions and concepts of data items. Significantly, it does not provide the practical guidance needed for actual development and implementation, nor does it provide guidelines for integrating this data with clinical records. As such, it is beneficial to investigate existing data from hospital systems that have implemented these standards. This study examined the field utilization of genetic test report data by designing the Fast Healthcare Interoperability Resources (FHIR) Genomics for Health Level 7 (HL7) for data elements based on the ISO/TS 20428 standard for genomic test reports. This study aims to verify the actual implementation of the ISO/TS 20428 standard in clinical settings and to report an approach for the implementation of a structured clinical sequencing report in EHRs according to international standards. We demonstrated the feasibility of implementing the ISO/TS 20428 standard by studying its application at various sites and integrating the clinical sequencing reports into the EHR system of a tertiary university hospital according to the HL7 FHIR standard.

Methods

International Standards for Clinical Sequencing Reports in EHRs

ISO/TS 20428:2017, entitled *Health informatics – Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records*, defines the data elements and their metadata required for a structured clinical genomic sequencing report [6]. ISO/TS 20428 focuses on genomic data generated using NGS technologies from only human samples. The data elements are divided into data fields required for clinical applications and optional data fields for clinical research, such as clinical trials and translational research. The metadata explains how and where particular appropriate terminological systems, which describe the genomes and/or diseases, can be applied. Thus, this standard attempts to integrate clinical and genomic information for a structured clinical genomic sequencing report template into an EHR system.

HL7 FHIR is an emerging standard that is rapidly gaining attention and being adopted by the health information technology industry and health care organizations [15]. This standard implements various types of information generated in the clinical environment in the form of a *resource* for exchanging clinical data and, thus, it ensures interoperability. Resources are an instance-level representation of a certain type of health care entity. Tailoring the FHIR specification to a specific use case is an important aspect of implementing the

FHIR standard. This process, also called profiling, enables FHIR to be adjusted to the needs of users and defines how exchanging organizations use the FHIR specification. FHIR profiles, as conformance resources, are an important aspect of the FHIR standard. Among the diverse FHIR profiles, FHIR Genomics has also been developed. FHIR Genomics comprises the MolecularSequence resource and several profiles built on top of existing FHIR resources, including a DiagnosticReport-genetics profile on DiagnosticReport, a ServiceRequest-genetics profile on ServiceRequest, an Observation-genetics profile on Observation, and a human leukocyte antigen (HLA)-genotyping-results profile on DiagnosticReport [16]. The MolecularSequence resource is designed for NGS data. The observed sequences of patients should be represented by recording reference sequence IDs and strings as well as detected variants. Extensions of the MolecularSequence resource address complex cases, and they can associate these with repositories for retrieving a patient's full sequence data. Other modifications include a set of genetics profiles for other FHIR resources. In addition, the Observation-genetics profile adds new references such that the Observation can report the genetics test results to be integrated into the EHR. The Observation-genetics profile is used to interpret variants from the sequence resource. Clinical usage may need more specific representation of variants at the locus or structural variants in the entire genome. This DiagnosticReport-genetics is designed based on the DiagnosticReport. The new profile is then used to describe the genetics test report. The resulting element in the DiagnosticReport refers to the Observation resource that can lead to a bundle of genetic observations. The elements of code, effective[x], issued, performer, request, and specimen are used to describe the details of the genetic test. Extensions of AssessedCondition and FamilyMemberHistory are also added. Overall, this profile extends the DiagnosticReport resource to enable the reporting of structured genetic test results. In addition, it denotes the condition context for genetic testing, which may influence the reported variants and interpretations for large genomic testing panels. Further, there are new genetics-extension profiles for extending DiagnosticReport, ServiceRequest, and FamilyMemberHistory for reporting genetics results. We have provided the suffix "-genetics" to all the aforementioned FHIR genetics profiles (eg, "DiagnosticReport-genetics profile"). New profiles built upon the DiagnosticReport have been created for reporting HLA genotyping results. FHIR Genomics focuses on clinical genetic data reporting and covers many aspects of genetic reporting, which includes bacterial and viral specimens, representations of simple discrete variants and structural variants, and full or partial DNA sequencing. In this study, existing FHIR Genomics resources were used, and additional extensions were further defined to expose gene variant data for presentation to the end user.

ISO/TS 20428 Standard Implementation Using FHIR

FHIR is a standard that focuses on providing a secure interoperability environment for health service providers to exchange data using the Representational State Transfer (REST) API [15,17,18]. We built a repository with the FHIR server that applied the STU3 (Standard Trial Use 3) version of FHIR to provide *create*, *read*, and *delete* operations for resources (Observation, Sequence) containing genome analysis result information. The build environment consisted of C# (.NET 4.5), Windows Server 2012, Oracle Database, etc. The Seoul National University Bundang Hospital (SNUBH) FHIR server supported the receiving and processing of the following sequence resource and genetics profiles operations: create, history, read, search, and update. To apply the ISO/TS 20428 standard to clinical sequencing reports within the FHIR infrastructure and to investigate its clinical applicability, profile tooling is an essential process. Therefore, we conducted a series of five collaborative workshops to compare the ISO/TS 20428 reporting standard and FHIR Genomics resources to better understand the comprehensive data requirements to define a mapping and create additional extensions. A total of 10 researchers participated in these workshops: one clinician, one nurse, five bioinformatics specialists, and three FHIR developers. The clinician—an otolaryngologist—reviewed the data items required for a clinician interface as the project supervisor. The nurse had 10 years of experience working in the medical informatics field and helped to identify the location, type, and structure of the data stored in EHRs and provided guidance. The bioinformatics specialists in this workshop had a variety of backgrounds: licensed HL7 standard professional, ISO/TS 20428 standard author, experts in medical informatics with biological backgrounds, and a PhD in genomics and systems biology. These experts were familiar with the content of ISO/TS 20428 and HL7 FHIR standards. They examined how real genomic data can be expressed and transmitted for the data elements presented in each standard. The three FHIR developers worked to implement the new standards in compliance with the security requirements for HL7 FHIR, facilitated further development for scalability, and enabled interagency data transmission.

The aim of these workshops was to develop a method for implementing the ISO/TS 20428 standard using HL7 FHIR. The opinions of all participants were collected for internal discussions; subsequently, a mapping table between the ISO/TS 20428 standard and the FHIR Genomics resources for implementing the ISO/TS 20428 standard using HL7 FHIR was constructed. All members of the workshop reviewed the ISO/TS 20428 standard and FHIR resources; they chose suitable FHIR profiles to implement the standard in the form of ISO/TS 20428 elements and a specification spreadsheet was derived (see Multimedia Appendix 1). Both the required and optional fields in ISO/TS 20428 were considered. In the case of inadequate or undefined resources in FHIR Genomics, new extension definitions were suggested, including datatypes, when there was a need for supplementation. We used the Forge program tool (Firely), the official HL7 FHIR profile editor, for FHIR profiling [19]. It allows the modelers to easily create, edit, and validate the FHIR profiles, extensions, and implementation guides. The users can fetch and publish FHIR resources and profiles. Forge enables users to create their own FHIR profiles, based on the FHIR base resources [19]. When using these extensions, the extension registry at the HL7 website [20] or Simplifier.net has to be examined to identify the already-defined extensions that may be suitable for appropriate needs. For example, a patient profile can be extended with the place of

birth with an already-existing extension from the HL7 extension registry. Based on the FHIR server infrastructure, SNUBH EHRs used the REST API of the FHIR server to obtain the genome analysis results of a specific patient and provided a service that shows the genome analysis results to a clinician through a web interface.

Regarding data security and privacy, we applied Transport Layer Security (TLS) and OAuth 2.0, which is an industry-standard protocol for the authorization process [21]. The TLS was used for encryption of transmission data and authentication for trusted servers. OAuth 2.0 was used to allow only authorized users to have limited access to FHIR resources. Access tokens were used for security credentials for applications to make API requests on the behalf of a user. These tokens represent the authorization of a specific application to access specific parts of a user's data [21].

Results

Mapping Table for Resource and Extension Definitions of ISO/TS 20428 to FHIR

A total of 32 data elements were included in the required fields of ISO/TS 20428 for patient information, type of sample, variant information, and recommended treatments. In the nomenclature of optional data, 29 items were defined to deliver more detailed information for biomedical and translational research. Five existing FHIR resources—ProcedureRequest, DiagnosticReport, Observation, Medication, and Patient—were used to represent the required fields of ISO/TS 20428. ProcedureRequest-genetics is a record of a request for a service such as diagnostic investigations, treatments, or operations to be performed; its name is changed to ServiceRequest-genetics. The DiagnosticReport-genetics profile contains the findings and interpretation of diagnostic tests performed on patients or groups of patients and on devices, locations, and/or specimens derived from these. The report includes the clinical context, such as request and provider information, and a combination of analysis results, images, textual and coded interpretations, and formatted representations of diagnostic reports. The required fields in ISO/TS20428 include clinical sequencing order information; information about the subject of care; and information about the person legally authorized to make the order, the performing laboratory, biomaterials, and genetic variations. Therefore, existing FHIR genetics elements sufficiently cover the genetic reporting standard as well. The other five targeted FHIR resources—Condition, DiagnosticReport, MolecularSequence, Observation, and Device—were leveraged and mapped by the ISO/TS 20428 standard's optional field elements.

The ISO/TS 20428 standard and detailed data element definitions were reviewed. Opinion gathering and decision making were conducted through workshops in such a way that the individual data elements of the ISO/TS 20428 standard were mapped to the FHIR Genomics resource. The data items of the ISO/TS 20428 standard and the FHIR Genomics resource were made into mapping tables and an analysis workflow based on this table was created.

Table 1 shows part of the ISO/TS 20428 standard and FHIR resource mapping table. It indicates how the data elements in the ISO/TS 20428 genetic analysis reporting standard are mapped to the existing FHIR Genomics resource. Multimedia Appendix 1 shows the full mapping table.

Table 2 shows the number of FHIR Genomics resources in ISO/TS 20428. The Observation resource includes the most data elements, whereas the Medication resource contains only one subelement for the ISO/TS 20428 standard's required fields. The existing FHIR resources must be defined to cover more content for the ISO/TS 20428 standard's optional field that focuses on clinical trials and translational biomedical research. To manipulate the detailed sequencing information on quality control metrics, base calling, sequencing platform, and analysis platform, additional FHIR extensions should be defined in this study.

Table 1. Specification example of the International Organization for Standardization Technical Specification (ISO/TS) 20428 standard and the Fast Healthcare Interoperability Resources (FHIR) Genomics resources.

ISO/TS 20428 standard		Example data		Maps to FHIR Genomics resources (proposed)	
Data elements	Metadata	Value	Representation	Target FHIR resource	FHIR.Xpath
Information on subject of care					
Identifiers	ISO/TS 22220:2011	12345678	12345678	DiagnosticReport	DiagnosticReport.subject(patient)
Name	ISO/TS 22220:2011	Gildong Hong	Gildong Hong	DiagnosticReport	DiagnosticReport.subject(patient)
Birth date	ISO 8601	1947-04-29	April 29, 1947	DiagnosticReport	DiagnosticReport.subject(patient)
Sex	ISO/TS 22220:2011	1	Male	DiagnosticReport	DiagnosticReport.subject(patient)
Ethnicity	Health Level 7 (HL7) version 3 code system race	2040-4	Korean	DiagnosticReport	DiagnosticReport.subject(patient).extension(ethnicity)
Genetic variation					
Gene symbols and names	Human Genome Organisation (HUGO) Gene Nomenclature Committee (HGNC)	HGNC:1097, B-Raf proto-oncogene (BRAF)	BRAF	Observation	Observation.extension(observation-geneticsGene)
Sequence variation information					
Notation	Human Genome Variation Society (HGVS)	c.1799T > A_p.V600E	c.1799T > A_p.V600E, Kinase domain (exon 15)	Observation	Observation.extension(observation-geneticsDNASequenceVariantName)
Effects of variants	Text	Substitution (missense)	Substitution (missense)	Observation	Observation.extension(observation-geneticsDNASequenceVariantType)
Sequence variant ID	Database unique ID	COSM476	Catalogue of Somatic Mutations in Cancer (COSMIC) website	Observation	Observation.extension(observation-geneticsDNAVariantId)

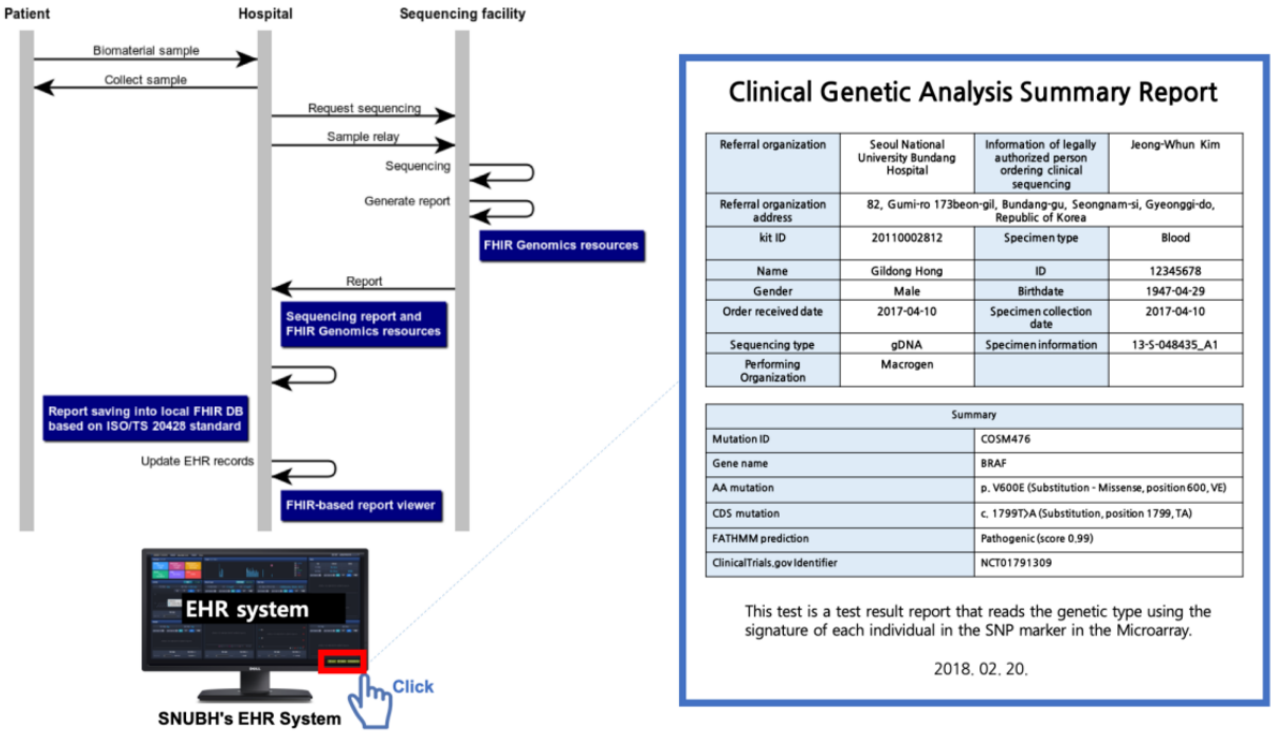
Table 2. Number of data elements per Fast Healthcare Interoperability Resources (FHIR) Genomics resource mapped by the International Organization for Standardization Technical Specification (ISO/TS) 20428 standard.

Standard field and FHIR resources	Number of elements in each resource	Number of extension elements
ISO/TS 20428 required field		
ProcedureRequest	3	0
DiagnosticReport	12	5
Observation	8	8
Medication	1	0
Patient	2	2
Total	26	15
ISO/TS 20428 optional field		
Condition	1	0
DiagnosticReport	2	2
MolecularSequence	16	10
Observation	6	6
Device	3	3
Total	28	21

Implementation of ISO/TS 20428 Standard and FHIR Genomics Resources

A bundled FHIR resource was created for clinical sequencing. It was delivered to the hospital from the sequencing facility and reconstructed in the clinician interface. The clinical sequencing report interface is populated with information about the variant results of the test, a summary, and information about the test (see Figure 1). Finally, we tested the feasibility of the ISO/TS 20428 standard on FHIR by building a clinician interface with sequencing from the sequencing facility. On this interface, patient data and FHIR Genomics resources from the FHIR server were combined. The clinical genetic analysis summary report consists of basic metadata for the sequencing order and a summary of the sequencing results. Data could be accessed, presented, and delivered with other types of patient information from the FHIR server. The clinical genetic analysis summary report was populated with information about the interpretation summary and the test. Multimedia Appendix 2 describes sample testing data for diagnostic requests.

Figure 1. Implementation of the clinical next generation sequencing (NGS) report viewer. DB: database; EHR: electronic health record: FHIR: Fast Healthcare Interoperability Resources; ISO/TS: International Organization for Standardization Technical Specification; SNP: single nucleotide polymorphism; SNUBH: Seoul National University Bundang Hospital.



Discussion

Principal Findings

To implement standardized machine-processable clinical sequencing reports in an EHR system, the ISO/TS 20428 international standard was proposed for a structured template.

Genomic sequencing laboratory reports are still provided in PDF or text formats in many clinical practices. Several difficulties, therefore, arise in clinical applications of these results due to unstructured image-based or narrative text–based genomic sequencing reports [22,23]. The ISO/TS 20428 standard was established to solve the variability in NGS reporting. It also defines the composition of a structured clinical sequencing report, which consists of the required and optional data fields and their metadata for a structured clinical sequencing report.

HL7 FHIR introduced web programming techniques to provide all standards and related resources in JSON (JavaScript Object Notation) or XML, making it easy for developers to implement. In particular, Sync for Genes, one of the pilot programs addressing significant technical challenges in this area, aims to standardize the sharing of genetic information by building FHIR Genomics. To this end, a pilot program involving laboratories, providers, governments, developers, and patients is developing ways to convert each participant's internal format to the FHIR Genomics format. Sync for Genes aims to use HL7's FHIR as the basis for data exchange to enable sharing of genetic information and, ultimately, to incorporate clinically tested genomes into patient care [24]. FHIR Genomics resources predominantly focus on transmitting or representing sequencing data, which is of less clinical value.

Genetic test reporting standards are beginning to develop in many standard development organizations, and there are no standard implementation guidelines for data items from the proposed standard at the clinical site and no guidelines or references for implementing gene sequencing data results for clinical use. This is a significant challenge for implementation and application of these standards at individual sites. The goal of this pilot was to facilitate the reporting and viewing of genomic data for clinical applications.

In this study, we developed a prototype interface based on ISO and HL7 FHIR standards and verified its field applicability. Further, we demonstrated that genomic data reporting can be successfully integrated with an EHR system based on these two standards. We described the design and delivery of a clinical genome sequencing report, including a summary suitable for interpretation by clinicians.

Lessons learned from this study are as follows. First, to manipulate the detailed sequencing information on quality control metrics, base calling, the sequencing platform, and the analysis platform, additional FHIR extensions were required to be defined. For this prototype, FHIR Genomics required a new source, *extension*, which was introduced in this study. For example, sequence variation information such as "c.1799T > A_p.V600E, Kinase domain (exon 15)" is related to the Observation resource domain, but there is no specific entity that is appropriate to contain the associated expression. Thus, we defined the Observation extension as *Observation.extension(observation-geneticsDNASequenceVariantName)*. In addition, pathological tier information, which is written as "Tier 1 (Pathogenic, Identified)," was divided into three extension entities: (1) *Observation.extension(observation-classificationVariants).Pathogeny* to express "Pathogenic," "Likely pathogenic," "Unknown significance," "Likely benign," and "Benign"; (2) *Observation.extension(observation-classificationVariants).Tier* to express "Tier 1"; and (3) *Observation.extension(observation-classificationVariants).ClinicalRelavance* to express "Identified," "Likely identified," "Uncertain," and "Not identified." In addition, the existing FHIR resources must be defined in order to cover more content for the ISO/TS optional field that focuses on clinical trials and translational biomedical research. Second, we found several challenges related to the lack of terminology standardization in the genomics domain. The HL7 FHIR standard could be easily applied to EHR systems to incorporate the genome sequencing reports for clinical practice. However, semantic interoperability should be further improved by standardizing a vocabulary to describe genetic mutations. Third, as the level of detail in reporting clinical genomic sequencing varies between hospitals, the ISO/TS 20428 standard needs to be expanded to meet the requirements of various clinicians. The development and evaluation of clinical decision support systems that utilize the standardized sequencing reports should also be considered in future research. Lastly, during this implementation, it was also found that the system integration workflow between the EHR system and the sequencing facility should be designed to fit clinical process in a secure manner. The OAuth 2.0 authentication technique and dynamic informed consent–based data sharing by patients should be considered for protecting patient data as well. Further study of usability and screen interface design is needed for individuals using mobile apps in the genome sequencing workflow and with physicians in verifying and using the genomic data in the EHR system.

As a limitation of this study, since the standard interface was implemented and tested only for one EHR system in Korea, the standard implementation may be restricted depending on the EHR system level. In addition, as previously mentioned, because the usability test was not conducted on patients and medical professionals, there was a limitation that the system performance and usability of the integrated system using the FHIR standard could not be verified. As the FHIR standards are evolving rapidly, a review of the latest standards may be needed again.

With the pipeline definition for technical frameworks and the mapping resource presented in this study, we hope that the ISO and HL7 standards will be applied to real genomic data at various institutions.

Conclusions

Sequencing laboratories still provide their reports primarily in unstructured formats, making it challenging to leverage or populate EHR data. To share and apply genomic data, especially for clinical practice, it is crucial to identify the applicability of the standard-based information system infrastructure in a practical setting. With existing FHIR resources, this prototyping study showed that reporting data from clinical genomics sequencing can be effectively stored and derived for EHRs for use by clinicians using the ISO standard and FHIR resources. We implemented a standardized clinical genomic sequencing report using the ISO/TS 20428 standard and FHIR. This systematic infrastructure is expected to enable the transfer and observation of clinical genomic sequencing reports in actual clinical applications.

Acknowledgments

This work was supported by the Bio & Medical Technology Development Program of the National Research Foundation of Korea (NRF), funded by the Korean government, Ministry of Science and ICT (MSIT) (No. NRF2017M3A9B6062319).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The mapping table of the International Organization for Standardization Technical Specification (ISO/TS) 20428 standard and Fast Healthcare Interoperability Resources (FHIR) Genomics resources.
[DOCX File , 29 KB - jmir_v22i8e15040_app1.docx]

Multimedia Appendix 2
BundleTransactionRequest.json file.
[DOCX File , 92 KB - jmir_v22i8e15040_app2.docx]

References

1. Conway J, Warner J, Rubinstein W, Miller R. Next-generation sequencing and the clinical oncology workflow: Data challenges, proposed solutions, and a call to action. JCO Precis Oncol 2019 Dec(3):1-10. [doi: 10.1200/po.19.00232]
2. Tarczy-Hornoch P, Amendola L, Aronson SJ, Garraway L, Gray S, Grundmeier RW, et al. A survey of informatics approaches to whole-exome and whole-genome clinical reporting in the electronic health record. Genet Med 2013 Oct;15(10):824-832 [FREE Full text] [doi: 10.1038/gim.2013.120] [Medline: 24071794]
3. Overby CL, Kohane I, Kannry JL, Williams MS, Starren J, Bottinger E, et al. Opportunities for genomic clinical decision support interventions. Genet Med 2013 Oct;15(10):817-823 [FREE Full text] [doi: 10.1038/gim.2013.128] [Medline: 24051479]
4. Tarczy-Hornoch P, Shannon P, Baskin P, Espeseth M, Pagon RA. GeneClinics: A hybrid text/data electronic publishing model using XML applied to clinical genetic testing. J Am Med Inform Assoc 2000;7(3):267-276 [FREE Full text] [doi: 10.1136/jamia.2000.0070267] [Medline: 10833163]
5. Nakaya J, Kimura M, Hiroi K, Ido K, Yang W, Tanaka H. Genomic Sequence Variation Markup Language (GSVML). Int J Med Inform 2010 Feb;79(2):130-142. [doi: 10.1016/j.ijmedinf.2009.11.003] [Medline: 19969503]
6. ISO/TS 20428:2017. Health informatics — Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records. International Organization for Standardization (ISO). 2017 May. URL: https://www.iso.org/standard/67981.html [accessed 2019-02-20]
7. Cappuccio FP, Taggart FM, Kandala N, Currie A, Peile E, Stranges S, et al. Meta-analysis of short sleep duration and obesity in children and adults. Sleep 2008 May;31(5):619-626 [FREE Full text] [doi: 10.1093/sleep/31.5.619] [Medline: 18517032]
8. ISO/PRF 21393. Genomics informatics — Omics Markup Language (OML). International Organization for Standardization (ISO). URL: https://www.iso.org/standard/70855.html [accessed 2019-09-20]
9. ISO/CD TS 23357. Genomics Informatics - Clinical genomics data sharing specification for next generation sequencing. International Organization for Standardization (ISO). URL: https://www.iso.org/standard/75310.html [accessed 2019-09-20]
10. ISO/CD TR 21394.2. Genomics informatics — Whole Genomics Sequence Markup Language (WGML). International Organization for Standardization (ISO). URL: https://www.iso.org/standard/75956.html [accessed 2019-09-20]
11. ISO/CD TS 22693. Genomics Informatics — Structured clinical gene fusion report in electronic health records. International Organization for Standardization (ISO). URL: https://www.iso.org/standard/73694.html [accessed 2019-09-20]
12. SO 25720:2009. Health informatics — Genomic Sequence Variation Markup Language (GSVML). International Organization for Standardization (ISO). 2009 Aug. URL: https://www.iso.org/standard/43182.html [accessed 2019-03-26]
13. ISO/TS 22220:2011. Health informatics — Identification of subjects of health care. International Organization for Standardization (ISO). 2011 Dec. URL: https://www.iso.org/standard/59755.html [accessed 2019-09-20]
14. Park P, Shin S, Park SY, Yun J, Shin C, Jung J, et al. Next-generation sequencing-based cancer panel data conversion using international standards to implement a clinical next-generation sequencing research system: Single-institution study. JMIR Med Inform 2020 Apr 24;8(4):e14710 [FREE Full text] [doi: 10.2196/14710] [Medline: 32329738]
15. Lamprinakos G, Mousas A, Kapsalis A, Kaklamani D, Venieris I, Boufis A, et al. Using FHIR to develop a healthcare mobile application. In: Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH). 2015 Jan 22 Presented at: 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH); November 3-5, 2014; Athens, Greece p. 132-135. [doi: 10.1109/MOBIHEALTH.2014.7015927]
16. HL7 FHIR Release 4: Genomics implementation guidance. HL7 International. 2019. URL: https://www.hl7.org/fhir/genomics.html [accessed 2019-03-18]
17. HL7 Fast Healthcare Interoperability Resources Specification (FHIR®), DSTU Release 1. HL7 International. URL: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=343 [accessed 2019-03-06]
18. Crump JK, Del Fiol G, Williams MS, Freimuth RR. Prototype of a standards-based EHR and genetic test reporting tool coupled with HL7-compliant infobuttons. AMIA Jt Summits Transl Sci Proc 2018;2017:330-339 [FREE Full text] [Medline: 29888091]
19. Forge. Firely. URL: https://fire.ly/forge/ [accessed 2019-03-06]
20. HL7 FHIR Release 4: FHIR core-defined extension registry. HL7 International. 2019. URL: http://hl7.org/fhir/extensibility-registry.html [accessed 2020-07-14]
21. OAuth 2.0. URL: https://oauth.net/2/ [accessed 2019-10-18]
22. Warner JL, Jain SK, Levy MA. Integrating cancer genomic data into electronic health records. Genome Med 2016 Oct 26;8(1):113 [FREE Full text] [doi: 10.1186/s13073-016-0371-3] [Medline: 27784327]

23. Tornheim J, Starks A, Rodwell T, Gardy J, Walker T, Cirillo D, et al. Building the framework for standardized clinical laboratory reporting of next-generation sequencing data for resistance-associated mutations in Mycobacterium tuberculosis complex. Clin Infect Dis 2019 Oct 15;69(9):1631-1633 [FREE Full text] [doi: 10.1093/cid/ciz219] [Medline: 30883637]
24. Sync for Science (S4S). URL: http://syncfor.science/ [accessed 2019-09-20]

Abbreviations

API: application programming interface
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level 7
HLA: human leukocyte antigen
ICT: information and communications technology
ISO: International Organization for Standardization
JSON: JavaScript Object Notation
MSIT: Ministry of Science and ICT
NGS: next generation sequencing
NRF: National Research Foundation of Korea
REST: Representational State Transfer
SNUBH: Seoul National University Bundang Hospital
STU3: Standard Trial Use 3
TC: Technical Committee
TLS: Transport Layer Security
TS: Technical Specification

Edited by G Eysenbach; submitted 14.06.19; peer-reviewed by S Thiebes, X Jin; comments to author 03.10.19; revised version received 30.10.19; accepted 03.06.20; published 10.08.20.

Please cite as:
Ryu B, Shin SY, Baek RM, Kim JW, Heo E, Kang I, Yang JS, Yoo S
Clinical Genomic Sequencing Reports in Electronic Health Record Systems Based on International Standards: Implementation Study
J Med Internet Res 2020;22(8):e15040
URL: https://www.jmir.org/2020/8/e15040
doi:10.2196/15040
PMID:32773368

Original Paper

ACTION-EHR: Patient-Centric Blockchain-Based Electronic Health Record Data Management for Cancer Care

Alevtina Dubovitskaya[1,2], PhD; Furqan Baig[3], MSc; Zhigang Xu[4], PhD, DABR; Rohit Shukla[3], MSc; Pratik Sushil Zambani[3], MSc; Arun Swaminathan[3], MSc; Md Majid Jahangir[3], MSc; Khadija Chowdhry[3], MSc; Rahul Lachhani[3], MSc; Nitesh Idnani[3], MSc; Michael Schumacher[5], PhD; Karl Aberer[6], PhD; Scott D Stoller[3], PhD; Samuel Ryu[4], MD; Fusheng Wang[3,7], PhD

[1]School of Information Technology, Lucerne University of Applied Sciences and Arts, Rotkreuz, Switzerland

[2]Swisscom, Zurich, Switzerland

[3]Department of Computer Science, Stony Brook University, Stony Brook, NY, United States

[4]Department of Radiation Oncology, Stony Brook Medicine, Stony Brook, NY, United States

[5]Applied Intelligent Systems Lab, University of Applied Sciences of Western Switzerland - Valais, Sierre, Switzerland

[6]Polytechnic University of Lausanne, Lausanne, Switzerland

[7]Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY, United States

Corresponding Author:
Alevtina Dubovitskaya, PhD
School of Information Technology
Lucerne University of Applied Sciences and Arts
Suurstoffi 1
Rotkreuz, 6343
Switzerland
Phone: 41 757 68 11
Email: alevtina.dubovitskaya@hslu.ch

Abstract

Background: With increased specialization of health care services and high levels of patient mobility, accessing health care services across multiple hospitals or clinics has become very common for diagnosis and treatment, particularly for patients with chronic diseases such as cancer. With informed knowledge of a patient's history, physicians can make prompt clinical decisions for smarter, safer, and more efficient care. However, due to the privacy and high sensitivity of electronic health records (EHR), most EHR data sharing still happens through fax or mail due to the lack of systematic infrastructure support for secure, trustable health data sharing, which can also cause major delays in patient care.

Objective: Our goal was to develop a system that will facilitate secure, trustable management, sharing, and aggregation of EHR data. Our patient-centric system allows patients to manage their own health records across multiple hospitals. The system will ensure patient privacy protection and guarantee security with respect to the requirements for health care data management, including the access control policy specified by the patient.

Methods: We propose a permissioned blockchain-based system for EHR data sharing and integration. Each hospital will provide a blockchain node integrated with its own EHR system to form the blockchain network. A web-based interface will be used for patients and doctors to initiate EHR sharing transactions. We take a hybrid data management approach, where only management metadata will be stored on the chain. Actual EHR data, on the other hand, will be encrypted and stored off-chain in Health Insurance Portability and Accountability Act–compliant cloud-based storage. The system uses public key infrastructure–based asymmetric encryption and digital signatures to secure shared EHR data.

Results: In collaboration with Stony Brook University Hospital, we developed ACTION-EHR, a system for patient-centric, blockchain-based EHR data sharing and management for patient care, in particular radiation treatment for cancer. The prototype was built on Hyperledger Fabric, an open-source, permissioned blockchain framework. Data sharing transactions were implemented using chaincode and exposed as representational state transfer application programming interfaces used for the web portal for patients and users. The HL7 Fast Healthcare Interoperability Resources standard was adopted to represent shared EHR data, making it easy to interface with hospital EHR systems and integrate a patient's EHR data. We tested the system in a distributed environment at Stony Brook University using deidentified patient data.

Conclusions: We studied and developed the critical technology components to enable patient-centric, blockchain-based EHR sharing to support cancer care. The prototype demonstrated the feasibility of our approach as well as some of the major challenges. The next step will be a pilot study with health care providers in both the United States and Switzerland. Our work provides an exemplar testbed to build next-generation EHR sharing infrastructures.

(J Med Internet Res 2020;22(8):e13598) doi:10.2196/13598

KEYWORDS

electronic health records; data sharing; blockchain; Hyperledger Fabric; privacy; security

Introduction

Timely sharing of electronic health records (EHR) across providers is essential for prompt medical care. For instance, transition and coordination of care for cancer patients are very common phenomena. A patient's history of health, tests, diagnoses, and treatments provides necessary knowledge for physicians to make clinical decisions. Access to EHR history is also preferred by individual patients to support personal and family engagement with user-centric control of data sharing and access [1]. Historical EHR data can also empower predictive modeling to drive personalized medicine and improve health care quality through machine learning [2].

EHR data are highly private and sensitive. According to the Health Insurance Portability and Accountability Act (HIPAA) [3], a patient has the right over his health information and can set rules and limits on who can access and receive health information. In current practice, if a patient needs to transfer his clinical data from one hospital to another, he is typically required to sign paper-based consent that specifies what type of data will be shared and the information about the recipient. EHR data sharing is mostly still a tedious manual process through fax or mail and often takes days or even months for the records to become available. This is mainly due to a lack of systematic infrastructure support for secure and trustable EHR data sharing, which may also incur major delays for patient care.

Ecosystems for health information exchange (HIE) aim to ensure that patient data from EHR are securely, efficiently, and accurately shared nationwide. However, HIEs have limited adoption, and many regional networks are still isolated [4]. Furthermore, the current system lacks standard architecture, resulting in a failure to ensure proper security and access control for patients once data are shared.

HIEs are generally designed as a single, fully trusted entity that is solely responsible for managing and storing EHR data from multiple participating hospitals. While a centralized system may be easier to manage, it suffers from a single point of failure and may prove to be a performance bottleneck for real-world deployment. In addition, a centralized authority with access to sensitive health information has proven to have more security and privacy concerns from end users. The experience with GoogleHealth wallet [5], for example, has shown that patients are concerned about their privacy and aware of the potential risk that their sensitive data might be misused. Alternatively, all the data can be stored and managed in the encrypted form (eg, using homomorphic encryption) for increased security and privacy. However, it requires large amounts of memory and extensive computations [6] that can be prohibitive for a hospital environment. Partial data encryption can improve the efficiency of such methods [7]; however, in medical settings, there might be a need to access and analyze all historical data and images (that lack any encryption) for better health care decisions.

Blockchain technologies have recently emerged with tremendous momentum based on the success of the Bitcoin cryptocurrency [8]. Blockchain uses a distributed ledger to provide a shared, immutable, and transparent history of all the actions that have happened to all the participants of the network. It enables a new generation of transactional applications that establish trust, accountability, and transparency. Blockchain, in particular permissioned blockchain technology, makes it possible for a user to have complete control of data and privacy without a central point of control; thus, it is highly cost-effective and efficient for building applications for sharing EHR data [9-15].

In this paper, we propose ACTION-EHR, a patient-centric, secure, trustable EHR data sharing framework with permissioned blockchain framework that can not only accelerate the data sharing process but also enable patients to take action on their own EHR data for sharing with full access control. The system will not only allow individual patients to stay at the center of their care but also enable medical practitioners and researchers to have fast, secure data access to enhance cancer treatment with significantly reduced cost and improved efficiency.

Methods

Background on Blockchain Technology

Blockchain is a peer-to-peer distributed ledger technology that provides a shared, immutable, and transparent append-only register of all the transactions that happen in the network [8]. Data in the form of transactions, digitally signed and broadcasted by the participants, are grouped into blocks in chronological order and timestamped. A hash function is applied to the content of the block and forms a unique block identifier, which is stored in the subsequent block, thus forming a "chain." Due to the properties of the hash function (the result is deterministic and cannot be reversed), one can easily verify if the block was modified by hashing the block content and comparing it with its identifier. The hash of the previous block, as a part of the block content, allows one to ensure the block belongs to its "location." The blockchain is replicated and maintained by every participant. With this decentralized approach, there is no need to set up a single trusted centralized entity for managing the registry. The participants (in particular, in the permissioned settings) will notice a malicious attempt to tamper with the

information stored in the registry; hence, the immutability of the ledger is guaranteed.

Blockchain technology relies on public key infrastructure (PKI) and employs cryptographic primitives such as digital signatures and asymmetric encryption. In case of an asymmetric encryption scheme [16], two keys are generated: One of the keys is publicly known (public key), and the other (private, or secret, key) is kept private by its owner. To send a secret message, the sender encrypts the message with the recipient's public key and sends it. The recipient can decrypt it using his private key. A digital signature is a construct that authenticates both the origin and contents of a message in a provable manner. A user signs the message with his private key, and other users can check the signature with the public key of the signer [17]. PKI is a framework for managing the creation, distribution, identification, authentication, and revocation of public keys [18]. Adding a new block to the existing ledger is defined by the consensus protocol employed in the implementation of the blockchain technology. A consensus protocol is defined as a protocol employed to disseminate requests among the nodes, such that each node executes the same sequence of requests on its instance of the service [19]. Based on the membership mechanism (ie, how the identity of the participant and their right to participate in the consensus are defined within a network, such as proof of work or endorsement policy), one could distinguish between permissionless and permissioned blockchain systems [20]. The role of the proof of work is to define who will be adding the next block (something that is defined at the policy level in case of permissioned blockchains where identities of the participants are known). Permissionless and permissioned systems also usually employ different consensus protocols (eg, Nakamoto consensus, Practical Byzantine Fault Tolerance [PBFT]). Nakamoto consensus realizes a replicated state machine abstraction, where nodes in a permissionless network reach agreement about a set of committed transactions as well as their ordering [21]. The protocol relies on chaining blocks of transactions. Nodes express their acceptance of the block by working on creating the next block in the chain, using the hash of the accepted block as the previous hash. In case of PBFT or other protocols used in permissioned blockchains, the designated set of nodes verifies the validity of transactions based on the credentials of the transaction’s origin. The architecture design of permissioned blockchain technology provides privacy and security guarantees that are impossible to achieve in the permissionless settings.

We restrict our work by following the settings of the permissioned blockchain technology for highly sensitive EHR [22]. A permissioned blockchain network is operated by known entities such as stakeholders of a given industry. Thus, this design choice enables improved control of the participating network and registration of users and minimizes the computational power compared to the expensive proof of work process used in a permissionless blockchain. As a result, better transaction throughput can be achieved while providing improved trustworthiness and preserving the privacy and audit trails of data sharing.

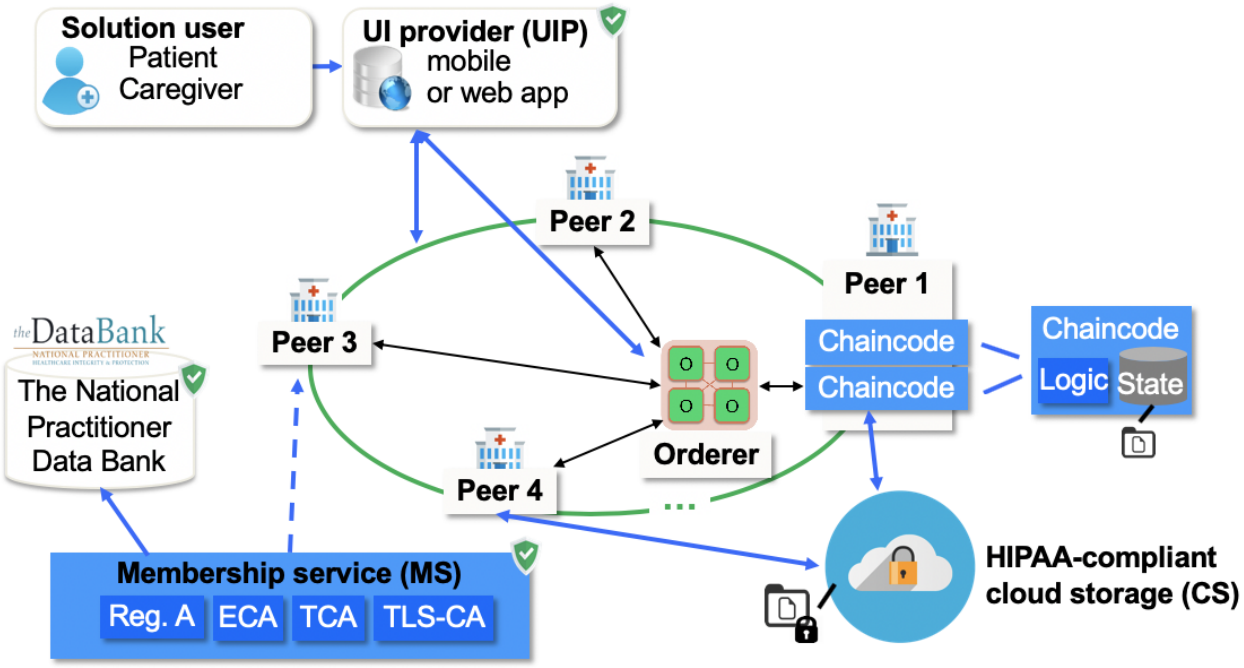
Hyperledger Fabric [23,24] — an implementation of a permissioned blockchain — is an open-source blockchain initiative hosted by the Linux Foundation. Hyperledger Fabric contains a security infrastructure for authentication and authorization (membership service [MS], employing a certificate authority [CA], which is an entity that can generate certificates for key pairs for signing and encryption for the peer nodes and solution users [SUs]). The goal of an MS is to support enrollment and transaction authorization of peers and users through public-key certificates. This is one of the main differences from the permissionless blockchain framework. Hyperledger Fabric also provides the support of anonymous credentials with multiple CAs and the use of threshold signatures. In addition to the MS, the other main architectural components are peers and an ordering-service node, or orderer. Orderer is a node (or a cluster of nodes) running the communication service that implements delivery guarantee, such as atomic or total order broadcast. This is done by transaction verification and ordering. During the verification phase, the digital signature of the transaction issuer is verified, as well as the so-called endorsement policy. The endorsement policy is defined for a chaincode and is used to instruct a peer on how to decide whether a transaction is valid. An example of such a policy can be defined as a requirement that all the peers in the network have to validate (and therefore sign) the transaction. Then, the orderer, during the verification, must ensure that the transaction is indeed signed by all the peers and that the signatures are valid.

In Hyperledger Fabric, smart contracts are implemented by the chaincode. The chaincode is defined by its logic and associated world state (state). The chaincode logic is a set of rules that define how the transactions will be executed and how the state will change. The logic can be written using general-purpose programming language. The state is a database that stores the information in a form of key-value pairs, where the value is an arbitrary byte array. The state also contains the block number to which it corresponds. The ledger manages the blockchain by including an efficient cryptographic hash of the state when appending a block. This enables efficient synchronization if a node was temporary offline, minimizing the amount of stored data at the node.

System Model of ACTION-EHR

ACTION-EHR (patient-centric, blockchain-based EHR management) is a permissioned blockchain-based system for EHR data sharing and integration. Each hospital will provide a blockchain node integrated with its own EHR system to form the blockchain network, and a web application will be used by patients and doctors to initiate EHR sharing transactions. To achieve scalability for the EHR data, ACTION-EHR takes a hybrid data management approach, where metadata on data sharing will be stored on the chain and shared EHR data will be encrypted and stored off-chain in HIPAA-compliant cloud-based storage. A patient (or his health care proxy) will be able to initiate a record-sharing request and define sharing permissions, thus having full control of the shared data. PKI-based [25] asymmetric encryption (that distinguishes between encryption and decryption keys) and digital signatures are employed to secure shared EHR data. The system model of ACTION-EHR framework is shown in Figure 1, and a prototype was implemented using Hyperledger Fabric v1.4.

Figure 1. The system model for ACTION-EHR, a distributed patient-centric blockchain-based electronic health record data sharing system based on permissioned blockchain technology and implemented using Hyperledger Fabric v1.4. ECA: enrollment certificate authority; HIPAA: Health Insurance Portability and Accountability Act; TCA: transaction certificate authority; TLS-CA: transport layer security certificate authority; UI: user interface.



ACTION-EHR consists of the following components: peer node, SU, client-server web application, MS, orderer, and HIPAA-compliant cloud storage (CS).

The peer node (1-n) is a peer in the EHR blockchain network representing a health care institution. As each hospital has its own EHR system, the peer node will have access to the EHR system for pulling data for sharing. The peer node has a web server, EHR integration layer, chaincode defining the sharing operations, and a database (ie, CouchDB [26]) for on-chain data management (Figure 2). The peers also agree on the validity of the transactions and maintain the current state of the blockchain ledger by adding new blocks of transactions and updating on-chain data accordingly. Metadata are stored on-chain and consist of EHR metadata (eg, data source, data category) and permission metadata for each EHR record to be shared.

The SU is an end user of the system. Currently, there are 2 SU roles available: patient and caregiver. We also assume that at every hospital, there is a trusted user with a role of administrator for registering new users. The user management will be automatized once the system is integrated into the clinical data flow, and the electronic identity management system in a hospital (ie, ID cards) can be used. The identity of the administrator, as well as the identities of the users, is maintained by a CA.

A client-server web application is used for SUs, who will access the system through a web client (user interface provider). A web server is deployed on each peer node, which interacts with the chaincode. A hospital administrator (Admin) enrolls the users and retrieves the data from the local EHR database. To ensure that the software is trustworthy, the source code can be digitally signed and made available as open source for verification.

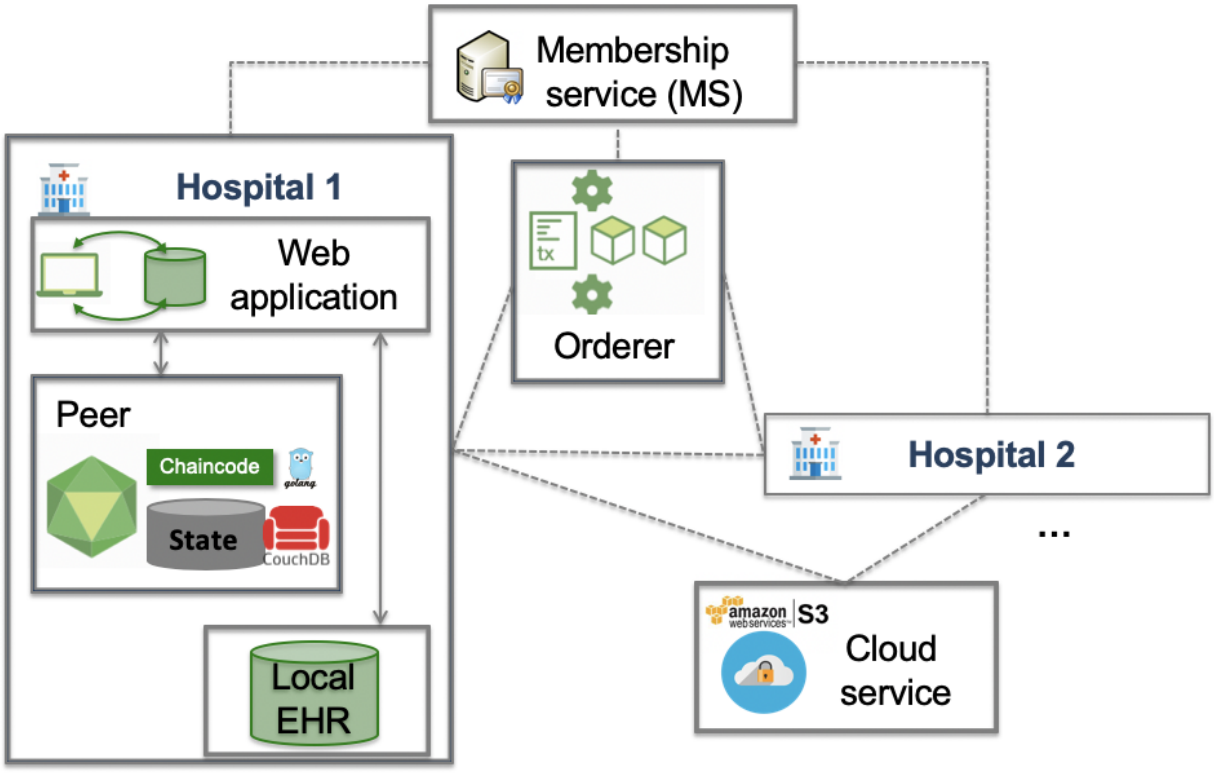
The MS is an entity that manages the network identities of all member organizations and users but does not have access to the EHR data or metadata stored on the blockchain. Before registering a peer, the MS uses a trusted database (such as the National Practitioner Data Bank) to verify the peer. To relax the assumptions and provide stronger security and distributed trust, collective authority servers could substitute for a single MS [27]. We assume that the MS is trusted and hosts a standard CA that can generate certificates for key pairs for signing and encryption for the peer nodes and SUs.

The orderer is a service that provides the verification and ordering of the transactions.

The HIPAA-compliant CS is a server where highly sensitive health care data in an encrypted form are stored according to the access-control policy specified by the patient. The CS is used to support exchange of large files such as medical images and can also be employed when constructing the full history of the patient data.

Based on this design, Action-EHR provides the following two user scenarios for the patient and caregiver, respectively.

Figure 2. Communication between the components of the ACTION-EHR system (single organization). EHR: electronic health record.



A patient can log in into the web platform using a web application. The patient can then share his health care data (using automatic fetching via a "record sharing service" responsible for verification of the patient's existence, the data existence, and pulling the data) from the local database, emulating the hospital database management system with registered caregivers from hospitals that form the blockchain network. To share the data with a caregiver, the patient will need to specify the caregiver, the category of the data to be shared, and for which period of time this caregiver will be able to access the data. The transaction is generated automatically based on the information provided by the patient via the web application and is broadcasted in the network. Simultaneously, prior to being uploaded to the cloud, the data are encrypted. The corresponding transactions that define the metadata of the uploaded data are then added to the ledger.

A caregiver can log into the system and query the ledger to view the permissions specified by the patient, download from the CS with respect to the permissions, and decrypt the data. The patient can revoke the permission given earlier to the doctor by updating the ledger with the corresponding transaction. The patient can also retrieve all historical transactions from the blockchain in chronological order. This can be also used for auditing purposes. Permissions can also be indirectly used to delete the data corresponding to the patient. If the patient wants to delete his data from the CS, he can modify the permissions on the blockchain accordingly. Implementation of the data-deletion process is the next step in our future work.

Implementation Considerations

Cloud Storage

One major challenge for sharing EHR data over blockchain is scalability, as EHR data such as images can be large. Due to the distributed replicated nature of a blockchain network, storing and replicating EHR data on the network for sharing are infeasible, as the large data volume will significantly slow performance. Instead, we propose a hybrid data management approach: All metadata (such as transactions, metadata of EHR, access control) are stored on the chain, but shared sensitive EHR data are stored and managed in a HIPAA-compliant cloud. We adopted Amazon Web Services (AWS), which provides HIPAA compliance through the "AWS Business Associate Addendum" [28]. The shared EHR will be encrypted and stored in AWS storage, which provides high scalability, high availability, and low latency.

Blockchain Nodes

The components of the Hyperledger Fabric are provided in the form of virtual containers — a standard unit of software that packages code and all its dependencies. However, in a real-work scenario, each peer will be physically located on the hospital premises; thus, we have to be able to run each peer on a separate machine. For metadata management on-chain, we take a key-value approach, where the "key" is a pseudonym of the patient (that can be generated as a randomly selected combination of letters and numbers or using a hash function), and the value represents the metadata represented in a JavaScript Object Notation document stored in a chaincode state database CouchDB [26], a document-oriented NoSQL database provided by Hyperledger.

Web Portal

To interact with the chaincode and manage the users (patients and caregivers), a web portal and set of methods that allow communication between the user interface and the server (chaincode) are required. For testing purpose, we created a simulated EHR database with example patients, EHR data, and caregivers for each node. The web portal of the applications makes asynchronous calls to the representational state transfer application programming interfaces implemented in Javascript. The technologies used to implement the web portal are HTML, cascading style sheets, and Javascript, as well as open-source Bootstrap libraries.

Cryptographic Operations

Following best practices for applied cryptography, we ensured that all the SUs possess 2 different key pairs for signing and encryption. The keys are generated during the registration phase. With his secret key for signing (SK_{SU}), the SU signs every transaction when exploiting the functionality of the chaincode. Users can verify the authenticity of the transactions and permissions by verifying the digital signature. A hybrid cryptosystem (ie, a cryptosystem that combines the convenience of a public-key cryptosystem with the efficiency of a symmetric-key cryptosystem [29]) is used to encrypt the patient's data. Patient data are being encrypted with the symmetric key before being uploaded to the CS. Then, the symmetric key is encrypted with the public key of a patient for storage and with the public key of the doctor, with whom the patient wishes to share the data. To decrypt the patient's data, the doctor first uses his corresponding private key to decrypt the symmetric key and then uses it for data decryption. Different approaches can be chosen depending on the available mechanisms to manage the symmetric key. One solution is to use only one patient-specific symmetric key to encrypt the data of this patient and ensure strong protection of this key. Although this theoretically gives the doctors with whom the patient shared the data the opportunity to decrypt all the patient's data, permissions stored on the ledger strictly manage access to the patient's data, only allowing the doctors to download the data according to the patient's access control policy. An advantage is that if a patient wants to share the same data with multiple doctors, he only needs to upload the encrypted data once, and only the key will be shared multiple times. Yet, if such a patient-specific key is compromised, a set of actions have to be immediately taken by the patient to prevent violation of his privacy and restore data availability. Using a newly generated symmetric key for each data-sharing operation will minimize this privacy threat in case of a compromised key yet will require establishment of comprehensive key management and duplication of the patient's data when shared with different doctors. While both approaches are viable, for the prototype, we used the latter and assumed that the keys are encrypted and managed off-chain and can be stored using existing conventional approaches (eg, smart cards, security tokens, or cloud-based hardware security modules [30]). The public-key encryption ensures confidentiality of the symmetric keys, the symmetric encryption ensures confidentiality of the patient's data, and the digital signature ensures integrity, non-repudiation (ie, provides proof of the origin), and authenticity.

The properties of the blockchain technology, architecture design, implementation approaches, and cryptographic interfaces guarantee the protection of the sensitive data that flow in the system. These include the following privacy and security properties: data integrity, confidentiality, authenticity, and availability according to the access-control policy, as well as unlinkability between system metadata and the corresponding patient's identity for any unauthorized user (ie, only the users authorized by the patient are permitted to link the patient's identity and his record stored on the blockchain).

Results

In this section, we present our solution prototype that demonstrates the feasibility of the approach. We describe the data model and the data sharing transaction that are in-line with the system model and required functionalities defined in the section, System Model of ACTION-EHR.

Overview of the Prototype

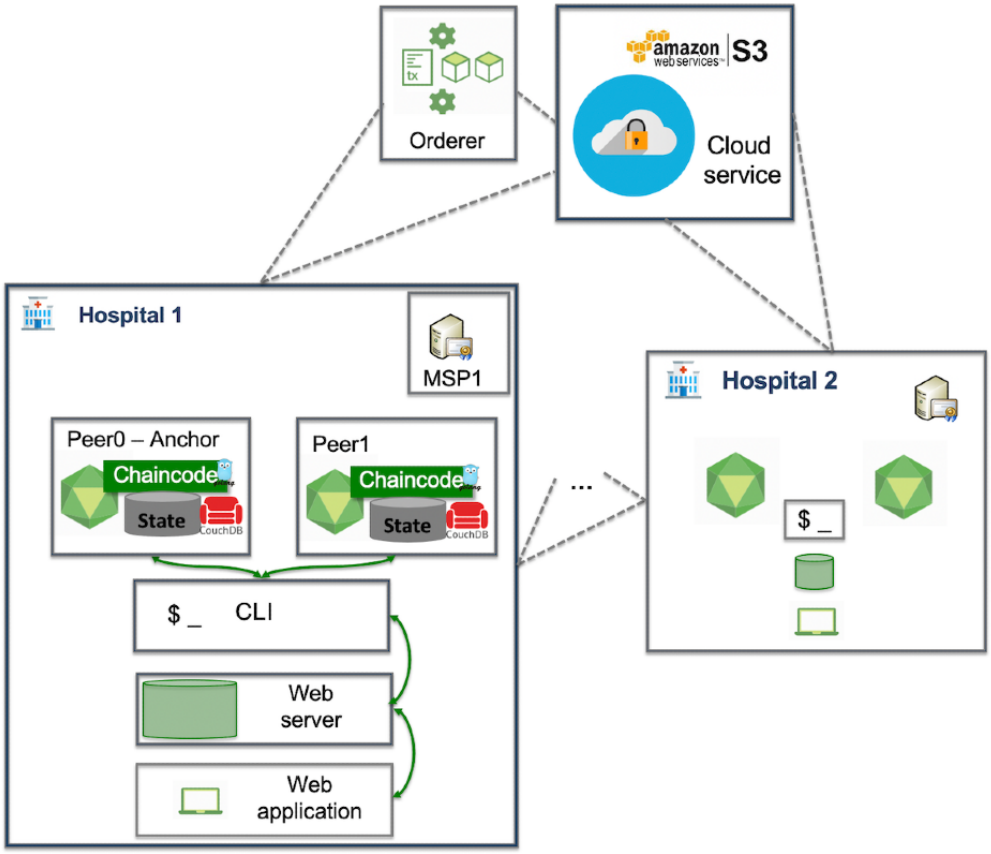
To present ACTION-EHR, we use the example of sharing EHR data between oncology information systems for radiation oncology. The EHR data include radiation, medical, and surgical information to assist radiation oncologists and medical physicists to manage different types of medical data, develop oncology-specific care plans, and monitor radiation doses for patients. We also describe one-organization and multiple-organization settings, both of which have been implemented.

Figure 2 shows the one-organization settings: The EHR blockchain network is formed by a cloud server, MS, orderer, and peer nodes (health care institutions; Figure 2). Every hospital participating in the network needs to deploy a server running a Hyperledger Fabric v1.4 peer, interfacing an EHR database (for testing, a simulated EHR database using MySQL was used per peer), an instance of CouchDB for on-chain metadata management, and a web application to interact with the chaincode and EHR.

However, while such approach is easier to set up and maintain, it may not be the best fit in practice: The MS is then required to manage the identities of all the users from different hospitals. Moreover, a certain level of centralization is unavoidable, as separate designated entities are required to host the MS and the orderer.

In order to address this issue and provide better levels of decentralization and trust, we employed the concept of organizations from Hyperledger Fabric to create a network of hospitals (as shown in Figure 3), where each hospital will be represented as an independent organization and will have its own MS, orderer, and set of peers and will host a web application. In such settings, user management is distributed, yet a patient is able to choose a hospital and doctor to add corresponding permissions. The network is dynamic: A hospital can join the network according to the policy set up in the network and after exchanging the public part of the cryptomaterials.

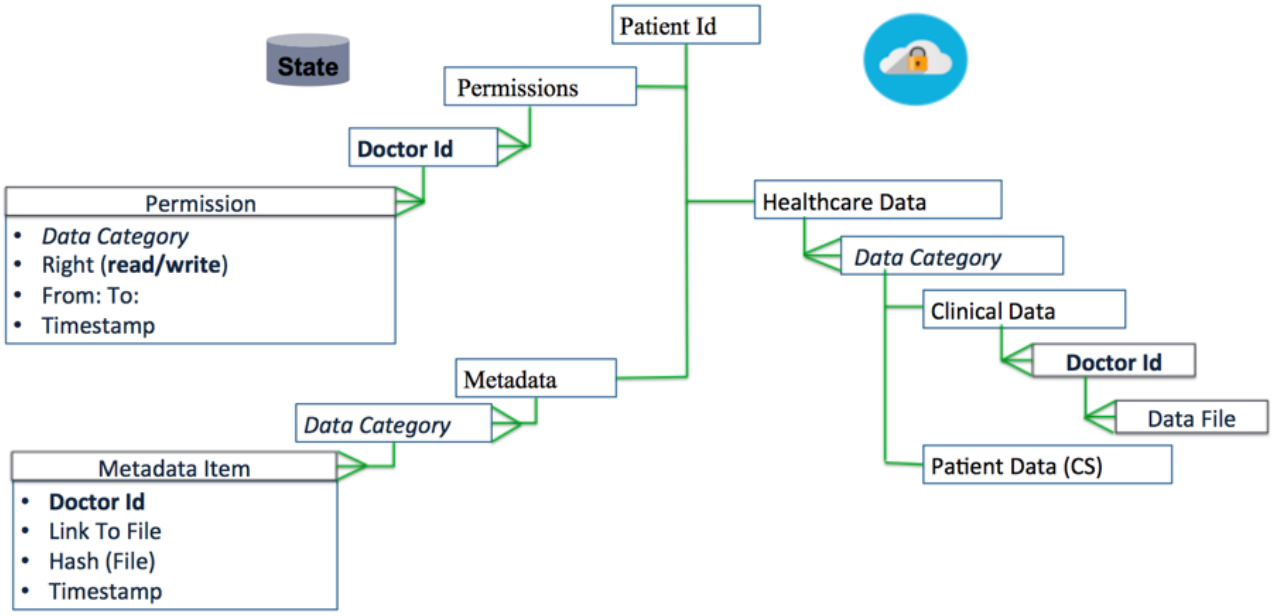
Figure 3. Communication between the components of the ACTION-EHR system (multiple organizations). CLI: command line interface; MSP: membership service provider.



Data Model of Action-EHR

Figure 4 presents the data model of ACTION-EHR, representing the data structure of EHR data and metadata that are stored on the chaincode and cloud server, respectively. Records on the blockchain are stored in a "key-value" form; "key" is a pseudonym of the patient, and a corresponding patient's record in JavaScript Object Notation format is stored as a byte array forming a "value" part of a chaincode state.

Figure 4. The data structure of the metadata and electronic health record (EHR) data stored on the blockchain and cloud server, respectively. CS: cloud storage.



Access Control Metadata

The block containing the information about the permissions is organized as follows. Each permission corresponds to an ID, with which a caregiver is registered in the system. Every permission specifies the timeframe ("from: to:") during which the clinician has the right to "read the patient's data that fall into a specific 'data category' and to upload the data to the cloud repository ("write")." "Timestamp" enables the patient to update and track access control changes. For patient P to revoke the right for caregiver C to access a specific type of data provided by other caregivers, patient P has to add a new permission with another time frame. To do so, the patient needs to update the ledger by sending the corresponding transaction.

EHR Metadata

Clinical metadata is a block that contains information about all the data files uploaded to the cloud by the clinicians or the patient himself. The metadata items are categorized based on the semantics of the corresponding data files. Every item contains an ID for the clinician that uploaded the data ("doctor ID") or a patient's pseudonym, a pointer to the file that is stored in the cloud ("path to file") and the hash of the data file ("hash(file)") to ensure unforgeability of the data stored in the cloud, and the "timestamp" of the moment when the data file was uploaded. It is not necessary to use a digital signature for the file instead of the hash, as the entire content of the transaction that contains doctor ID and hash(file) is digitally signed by the doctor that uploads the file.

Web Portal for Solution Users

The web application is an implementation of a user interface that provides users easy access to the functionalities of our prototype. There are 3 views of the web portal: administrator (to be merged with the identity management system in a hospital once the prototype is fully integrated into the clinical dataflow), patients (user), and caregivers (user).

The administrator page shows the list of all the patients and doctors from the hospital that are registered in the system. It is possible to enroll a new user via the administrator page by invoking the MS and verifying that the credentials were generated for this user. Through this page, it is also possible to customize the interface (eg, to add or remove a new department or new roles; extending the functionality by adding new roles such as nurse or laboratory scientist is planned for the future development of ACTION-EHR).

When the patient logs in to the patient portal, he can access the patient-specific functionalities of the prototype (Figure 5). The patient can view his data that are currently stored on the cloud, together with the corresponding permissions, including which doctor(s) can access the data and during which period of time. When the patient adds a permission to allow the doctor access the data, the data are encrypted with the public key of this doctor and uploaded to the cloud. The patient can also download his data and modify the access control policy by adding new permissions. The patient can also see the caregivers from whom he is receiving care and a history of all the sharing transactions.

Figure 5. Web portal (patient view): p1 can add a new permission for caregiver d1 ("read" the data for the specified time interval and immediately "revoke" the permission given to the specified caregiver).

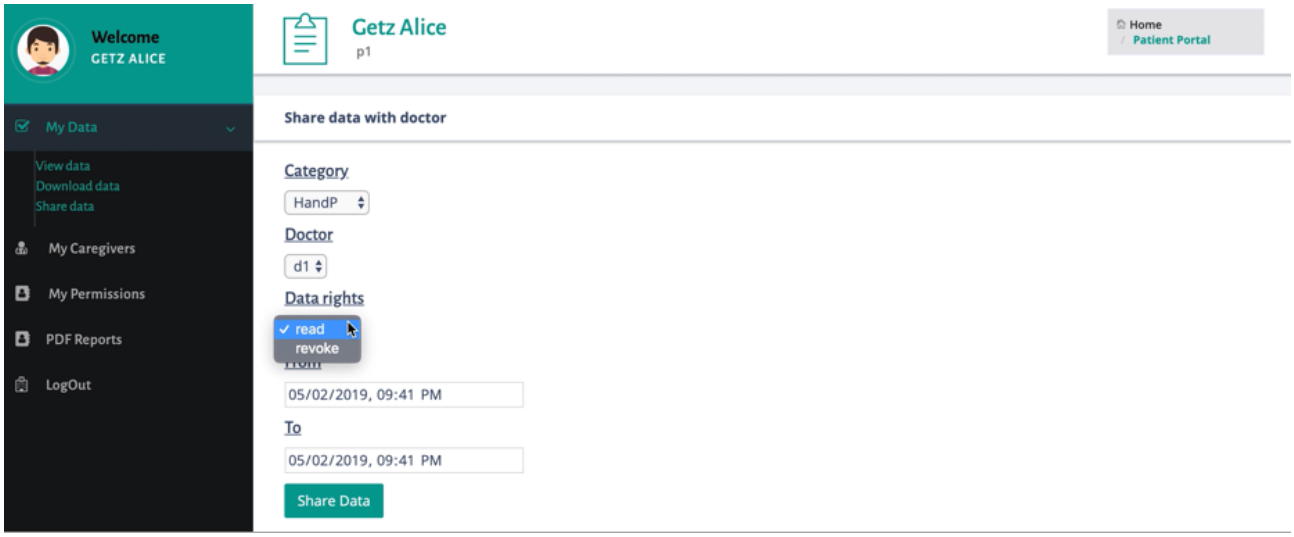
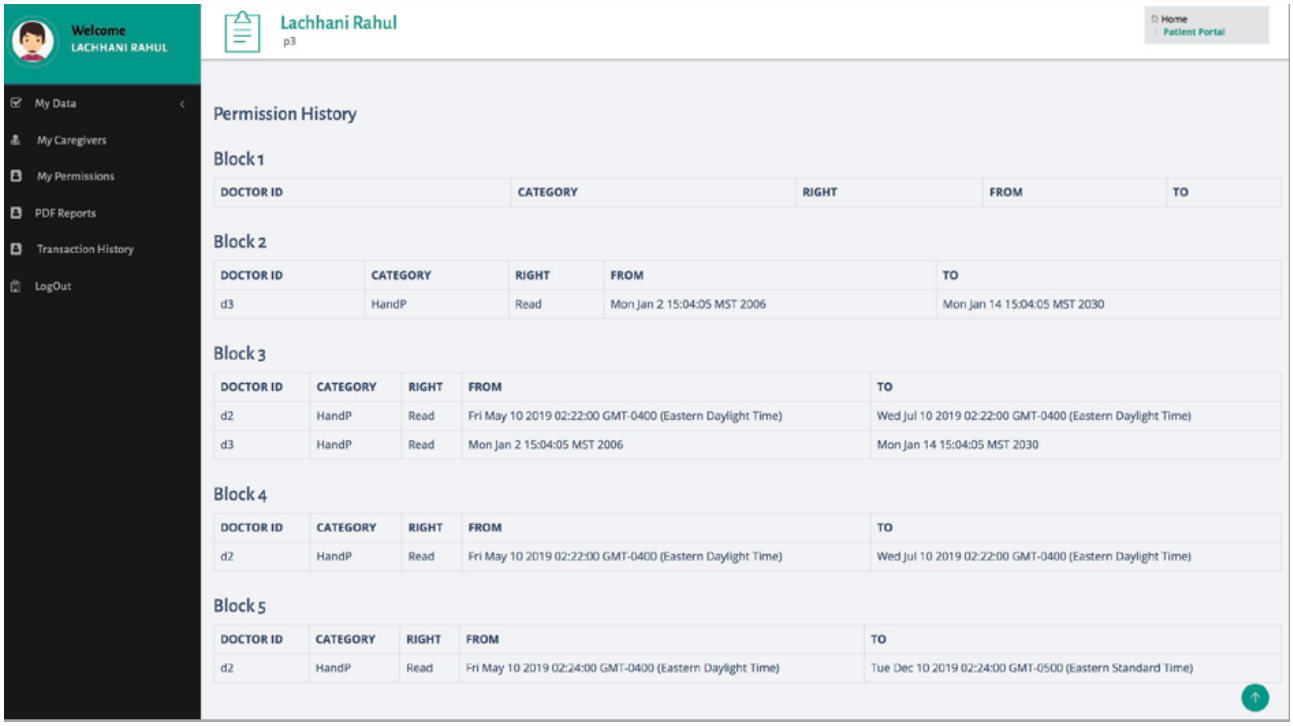


Figure 6 shows the history of all the permissions given by a specific patient. This information can be used by the patient to review his access control policy, as well as for audit purposes. In further development of the system, this information will also contribute to building the full history of the patient's data.

Figure 6. Permission history of a specific patient (p3).



The portal for caregivers (as shown in Figures 7 and 8) allows the doctors to view information for the patients receiving treatment and what data they have shared. The doctor can see which patient shared with him which type of data and during which period of time (Figure 7) as well as download the corresponding patient data files based on the permissions specified by the patient (Figure 8).

Figure 7. Web portal for the caregiver: view data.

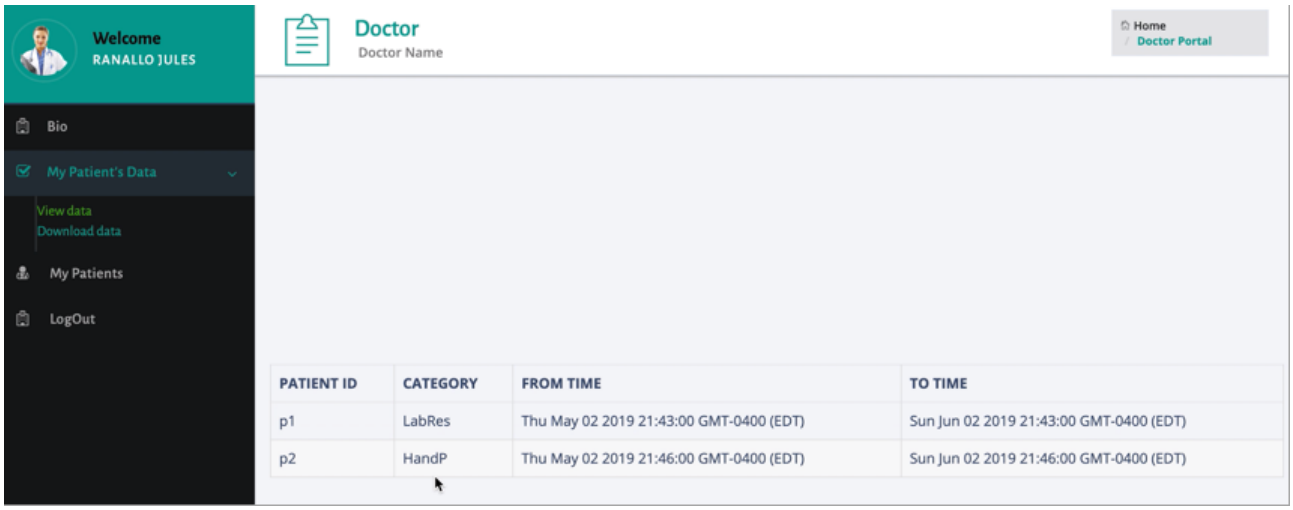
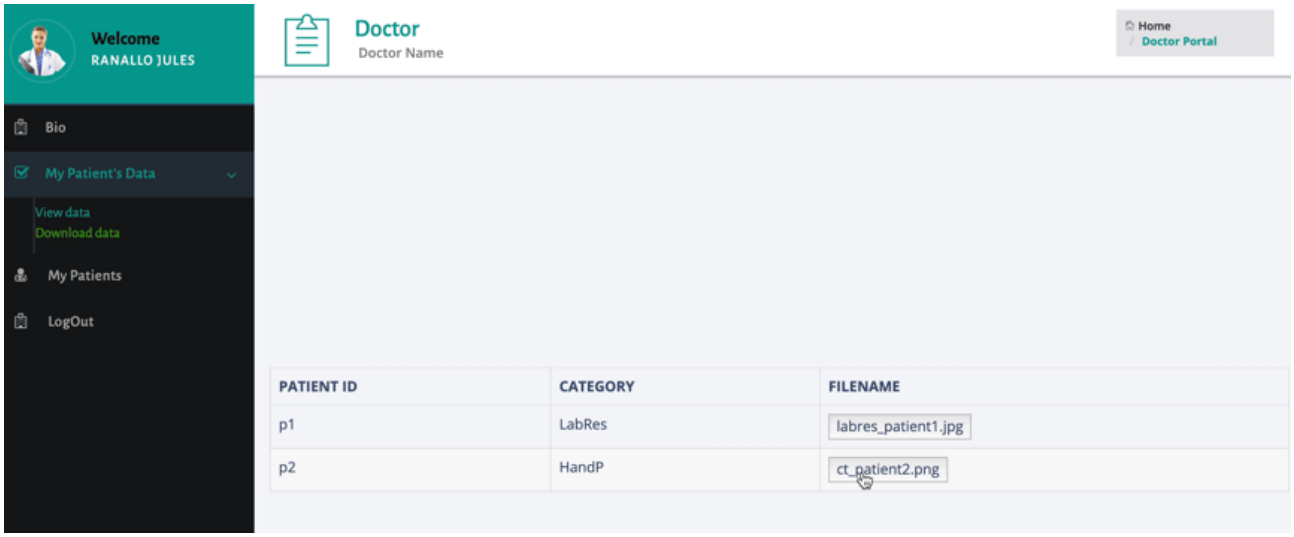


Figure 8. Web portal for the caregiver: download data.



Discussion

Principal Findings

To achieve the required functionalities of ACTION-EHR (see System Model of ACTION-EHR), we designed the system architecture and data structure of on-chain and off-chain storage. We defined and implemented the data-sharing protocol with respect to the health care scenario and developed the chaincode accordingly. We created the web application that serves as a user interface and enables interaction with the chaincode. To ensure interoperability and seamless integration of our system into the clinical dataflow, we implemented an independent pluggable module that provides conformance with the Fast Healthcare Interoperability Resources (FHIR) standard.

Given the health care environment for which our system has been developed, it is of high importance to ensure security and privacy of multiple data types of different sensitivity levels. Our system guarantees privacy rights and security properties (data integrity, availability, confidentiality, authenticity, and unlinkability) for the following types of data: EHR data, metadata (including permissions or access control policy), and cryptographic keys and user credentials.

Multimedia Appendix 1 contains a detailed description of the threat model, definition of the security properties, and security analysis.

Limitations

It is challenging to apply a relatively new technology that is not yet framed by government rules and regulations in a highly regulated health care environment. Some technical limitations of our prototype, especially related to the application domain, are highlighted in the paragraphs that follow.

The risk of having a single point of failure of the system can occur if the deployment of the system is not done correctly (if only a single orderer and single CA are employed). Using Kafka cluster and multiple CAs (such as in [27]) can address this limitation. The properties of the group signatures [31] and anonymous credentials [32] could also be explored to address this limitation in future work.

In the health care domain, emergency situations occur regularly, and data might be required urgently. If an unconscious patient arrives at a medical institution and the access control policy is defined such that that no caregiver from the medical institution has a right to access the patient's EHR, it is impossible to update the permissions and grant caregiver access to the data. Robust and secure "break-glass" mechanisms for emergency situations are required to address this limitation.

According to the new General Data Protection Regulation in Europe, a patient has "the right to be forgotten." This right might not be easily compatible with the immutability principles of the current implementations of the blockchain technology used in this work; the patient cannot delete his metadata record from the ledger. Applying different cryptographic techniques such as asymmetric encryption, threshold encryption, and proxy re-encryption, as well as principles of redactable blockchains (as proposed in [33,34]) could be used to address such limitations and will require further investigation.

Comparison With Prior Work

In this section, we describe recent related work employing blockchain technology to achieve fast, secure, and privacy-preserving sharing of EHR. We also underline the differences with our work described in this paper.

A recent review [15] provided an extensive list of studies and ongoing projects that focus on exchanging patient care data using blockchains to improve medical record management, conduct clinical studies, and support health care financing tasks. The authors describe key benefits of using blockchain technology in health care and discuss potential problems and challenges to be considered when adopting permissionless blockchain technology (eg, speed and scalability, confidentiality, the threat of a 51% attack, management of the transaction fees, and "mining").

The two most mature prototypes are MedRec and FHIRchain. MedRec [13] is a system based on Ethereum smart contracts for an intelligent representation of existing medical records that are stored within individual nodes on the network. The authors propose two incentivizing models for "mining," including the

possibility of accessing the data for research purposes. However, the authors did not propose a mechanism for generation of such anonymous data for research in a decentralized manner while ensuring patient privacy. FHIRchain [10] is a blockchain-based approach for data sharing that encapsulates the HL7 FHIR standard for clinical data. Zhang et al [10] describe a rigorous and deep study of using blockchain technology to transfer EHR. As these two systems use the permissionless blockchain technology, they both face most of the challenges listed by Kuo et al [15] and those already mentioned. Moreover, for instance, in MedRec, the pseudonymous property of transactions and use of a public key as a node address in a current prototype implementation can lead to inferring patterns of treatment from frequency analysis. Even without disclosure of name or personally identifiable information (encryption of the on-chain data and traffic), through the analysis of network communications, one could infer that some interactions have taken place. In our work, we employed permissioned blockchain technology, where only verified nodes are allowed to have access to the ledger. This prevents malicious traffic monitoring. Moreover, we employed encryption of the off-chain data and stored them in the cloud to protect against accidental or malicious violation of confidentiality and unavailability of the patient's health care data. At the same time, we removed from the original data sources the threats posed when publicly exposing interfaces for data access based on the pointers to the original data sources. In addition, employing an MS allows more flexibility of the user management processes, including rigorous verification of a new user. In our work, we also employed a pseudonymization approach that allows retrieval of the sharing history of a specific patient. In cases of the FHIR chain and other Ethereum-based implementations, public keys are employed as a form of digital identity. However, if the user loses his private key, it is impossible to authenticate this user.

In the space of applying permissioned blockchain technology in health care, the following studies share some similarities with our approach. The work presented by Liang et al [9] focuses on collecting medical data from wearable health devices, such as watches and bands. The authors proposed using the permissioned blockchain technology and storing health care data on-chain. Liang et al [9] implemented an access control scheme by utilizing the MS component and data separation via channels to protect privacy. Our approach is different, as we propose using the blockchain ledger to mainly store the metadata and permissions corresponding to the health care data, which are stored on the cloud service in encrypted form. This enables a more granular access control policy, enhances the data security and privacy, and avoids unnecessary replication of health care data. Magyar [35] used the basic principles of the HIPAA regulation and suggested a list of cryptographic tools that can be potentially applied to ensure data privacy and security, as well as potential approaches to modeling EHR blockchain–based EHR applications. While providing important insights, the work of Magyar [35] is only theoretical; no implementation is provided.

Analysis of the challenges that need to be addressed in the health care industry, as well as the potential benefits of employing blockchain technology, especially a permissioned implementation, can be found in the studies by Krawiec et al [11], Paranjape et al [36], and in a whitepaper from IBM [12]. Our work, while agreeing with the general statements, also focuses on the practical example and extends them by providing analysis of the privacy and security, as well as current limitations and approaches to address them, of a specific implementation.

Peterson et al [14] presented another system design based on the permissioned blockchain implementation (MultiChain [37]) and discussed how FHIR integration into such a system can address the interoperability issue. The proof of interoperability proposed by Peterson et al [14] is based on conformance to the FHIR protocol, which requires verification that the messages sent to the blockchain can get converted to other required formats. This work by Peterson et al focuses on data storage and data interoperability but is limited in terms of the smart-contract functionality that is not supported by the chosen underlying blockchain technology implementation. In contrast, we leveraged the smart-contract functionality to enable a dynamic access control policy definition and to ensure some of the privacy and security properties of our prototype.

Storing data on blockchain can restrict the data volumes that can be efficiently managed and can violate the rights of the patient (ie, to delete data or withdraw from participation in a research study). Motohashi et al [38] described a system design for a blockchain-based system for clinical trials that requires data aggregation from mobile devices. The authors proposed using multiple relay servers to encrypt the data before uploading it to the blockchain. While using relays helps against tampering with and takes on the complexity of data encryption on the mobile device, the relay servers (or at least the majority of them) have to be trusted. This can be acceptable only for anonymized data and if it is impossible to link the data to the real identity of a user or owner of a mobile device. Li et al [39] presented a system to share encrypted prescriptions data and used the same underlying implementation of private permissioned blockchain technology and, similar to our work, key-sharing mechanism. However, as in the study by Motohashi et al [38], Li et al [39] chose to store the data on the blockchain. In ACTION-EHR, only metadata are stored on the blockchain. Hylock et al [40] presented a patient-centric blockchain framework that supports a set of configurations (different modes related to the encryption and data-storage modes). To comply with legislation, the authors proposed an alternative approach: to use redactable blockchain [41] to build the ledger that consists of immutable and non-immutable blocks. The authors however do not provide a multinode implementation, which makes it impossible to evaluate proof-of-concept presented in the paper. The authors also proposed storing the data at the original data sources, which, as already discussed, can introduce data unavailability and security threats.

Pournaghi et al [42] proposed using both permissioned and permissionless blockchain technologies, the former to exchange the pointers to the encrypted data stored in the cloud, as well as the symmetric encryption key, encrypted with an attribute-based encryption scheme. The latter is used to distribute the description and access-control structure for the data stored in the cloud. The authors proposed using PBFT-based consensus for both blockchains, which can introduce scalability issues. In

addition, defining an access structure for an attribute-based encryption scheme requires specification of the attributes for medical and patient profiles, using a format of the tree where each inner node is a threshold gate and the leaves are the attributes. It can be difficult for a patient to construct such an access-control structure. Sharing such a structure, as well as sharing cryptographic keys (even in the encrypted form) on the blockchain, could also create a threat to patient privacy. While we use a similar concept of storing encrypted data on the cloud and sharing pointers on the blockchain, we propose exchanging the keys off-chain and enforcing an access-control policy by letting the cloud verify the consistency of the ledger prior to sharing the data. We propose a simpler yet secure approach for defining an access-control policy; permissions are defined for pseudonymized users and are stored on a private permissioned blockchain.

A plethora of existing blockchain platforms and various prototypes built on top of the technologies can aggravate the lack of interoperability between health care systems that is highly relevant due to multitude EHR systems with different interfaces. Thus, ensuring interoperability between different blockchain platforms is of high importance and shall be considered as one of the possible directions for future work. Moreover, due to custom privacy requirements and individual needs of different patients, one can think of a multiple-ledger design: a patient-specific or even case-specific ledger [43]. Data then can be replicated among multiple ledgers and locations, creating the network of networks [44].

Conclusions

In health care, a distributed ledger can be seen as a shared immutable and transparent history of all the actions performed by eHealth users; these actions include defining access control policies and sharing, accessing, and modifying the data. This work presents the architecture of the framework for the specific data sharing case for radiation oncology and the implementation of a prototype that ensures privacy, security, availability, and granular access control over highly sensitive patient data. The methodology is general and can be easily extended to support other types of patient care.

The functionality of the prototype meets the requirements from a medical practice perspective. To ease the adoption of the prototype, we implemented an independent pluggable module that conforms with the FHIR standard. Our next step is to set up a pilot network of health care institutions in the United States and Switzerland for further testing of ACTION-EHR with patient data. Once adopted by the health community, such a system will reduce the turnaround time for data sharing, improve decision making for medical care, and reduce the overall cost.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Security analysis.
[PDF File (Adobe PDF File), 99 KB - jmir_v22i8e13598_app1.pdf]

References

1. Hsiao C, King J, Hing E, Simon AE. The role of health information technology in care coordination in the United States. Med Care 2015 Feb;53(2):184-190. [doi: 10.1097/MLR.0000000000000276] [Medline: 25464164]
2. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine 2018 May 8;1(1):1-10. [doi: 10.1038/s41746-018-0029-1]
3. Health Information Privacy U.S. Department of Health & Human Services. URL: http://www.hhs.gov/hipaa/ [accessed 2019-02-01]
4. Trusted Exchange Framework and Common Agreement. URL: https://www.healthit.gov/topic/interoperability/trusted-exchange-framework-and-common-agreement [accessed 2019-02-01]
5. An update on Google Health and Google PowerMeter. 2011 Jun 24. URL: https://googleblog.blogspot.com/2011/06/update-on-google-health-and-google.html [accessed 2019-02-01]
6. Moore C, O'Neill M, O'Sullivan E, Doroz Y, Sunar B. Practical homomorphic encryption: A survey. 2014 Presented at: IEEE International Symposium on Circuits and Systems (ISCAS); 2014 Jun 1; Melbourne. [doi: 10.1109/iscas.2014.6865753]
7. Gentry C, Halevi S, Smart N. Fully homomorphic encryption with polylog overhead. 2012 Presented at: Annual International Conference on the Theory and Applications of Cryptographic Techniques; 2012 Apr 15; Berlin p. 465-482. [doi: 10.1007/978-3-642-29011-4_28]
8. Nakamoto S. 2008. URL: https://bitcoin.org/en/bitcoin-paper [accessed 2019-02-01]
9. Liang X, Zhao J, Shetty S, Liu J, Li D. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. 2018 Presented at: Personal, Indoor, and Mobile Radio Communications (PIMRC), IEEE 28th Annual International Symposium on; 2017 Oct 8; Montreal p. 1-5. [doi: 10.1109/pimrc.2017.8292361]
10. Zhang P, White J, Schmidt D, Lenz G, Rosenbloom S. FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data. Comput Struct Biotechnol J 2018;16:267-278 [FREE Full text] [doi: 10.1016/j.csbj.2018.07.004] [Medline: 30108685]

11. Krawiec R, Housman D, White M, Filipova M, Quarre F, Barr D, et al. Blockchain: Opportunities for health care. In Proc. NIST Workshop Blockchain Healthcare; 2016. URL: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/us-blockchain-opportunities-for-health-care.pdf [accessed 2019-02-01]
12. Blockchain: The chain of trust and its potential to transform healthcare our point of view; 2016. URL: https://www.healthit.gov/sites/default/files/8-31-blockchain-ibm_ideation-challenge_aug8.pdf [accessed 2019-02-01]
13. Azaria A, Ekblaw A, Vieira T, Lippman A. Using blockchain for medical data access and permission management. MedRec: Using blockchain for medical data access and permission management. In International Conference on Open and Big Data (OBD); 2016 Presented at: International Conference on Open and Big Data (OBD); 2016 Aug 22; Vienna p. 25-30. [doi: 10.1109/obd.2016.11]
14. Peterson K, Deeduvanu R, Kanjamala P, Boles K. A blockchain-based approach to health information exchange networks. 2016 Presented at: Proc. NIST Workshop Blockchain Healthcare; 2016; Gaithersburg p. 1-10.
15. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 01;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
16. Diffie W, Hellman M. New directions in cryptography. IEEE Trans. Inform. Theory 1976 Nov;22(6):644-654. [doi: 10.1109/tit.1976.1055638]
17. Katz J, Menezes A, Van OP, Vanstone S. Handbook of applied cryptography. Cleveland: CRC Press; 1996:25-27.
18. Fromknecht C, Velicanu D, Yakoubov S. A Decentralized Public Key Infrastructure with Identity Retention. IACR Cryptology ePrint Archive Nov 2014:1-16. [doi: 10.1016/b978-0-12-416681-3.00003-3]
19. Schneider FB. Implementing fault-tolerant services using the state machine approach: a tutorial. ACM Comput. Surv 1990 Dec;22(4):299-319. [doi: 10.1145/98163.98167]
20. Sompolinsky Y, Zohar A. Secure high-rate transaction processing in bitcoin. 2015 Presented at: International Conference on Financial Cryptography and Data Security; 2015 Jan 26; Puerto Rico p. 507-527. [doi: 10.1007/978-3-662-47854-7_32]
21. Borge M, Kokoris-Kogias E, Jovanovic P, Gasser L, Gailly N, Ford B. Proof-of-personhood: Redemocratizing permissionless cryptocurrencies. 2017 Presented at: IEEE European Symposium on Security and Privacy Workshops (EuroS&PW) (pp.). IEEE; 2017 Apr 26; Paris p. 23-26 URL: https://doi.org/10.1109/EuroSPW.2017.46 [doi: 10.1109/eurospw.2017.46]
22. Dubovitskaya A. Privacy-Preserving Data Exchange and Aggregation in Healthcare. EPFL. 2018. URL: https://infoscience.epfl.ch/record/256536?ln=en [accessed 2020-08-07] [WebCite Cache ID https://infoscience.epfl.ch/record/256536?ln=en]
23. Gaur N, Desrosiers L, Ramakrishna V, Novotny P, Baset S, O?Dowd A. Hands-On Blockchain with Hyperledger: Building decentralized applications with Hyperledger Fabric and Composer. Hands-On Blockchain with Hyperledger: Packt Publishing Ltd; 2018:9781788994521.
24. Androulaki E, Barger A, Bortnikov V, Cachin C, Christidis K, De CA, et al. Hyperledger fabric: a distributed operating system for permissioned blockchains. Hyperledger fabric: a distributed operating system for permissioned blockchains. In Proceedings of the Thirteenth EuroSys Conference; 2018 Presented at: Thirteenth EuroSys Conference; 2018; New York. [doi: 10.1145/3190508.3190538]
25. Perlman R. An overview of PKI trust models. IEEE Network 1999;13(6):38-43. [doi: 10.1109/65.806987]
26. Apache CouchDB. URL: http://couchdb.apache.org/ [accessed 2019-09-01]
27. Syta E, Tamas I, Visher D, Wolinsky D, Jovanovic P, Gasser L, et al. Keeping authorities ?honest or bust? with decentralized witness cosigning. 2016 Presented at: IEEE Symposium on Security and Privacy (SP); 2016 May 22; San Jose p. 526-545. [doi: 10.1109/sp.2016.38]
28. Architecting for HIPAA Security and Compliance on Amazon Web Services. URL: https://d1.awsstatic.com/whitepapers/compliance/AWS_HIPAA_Compliance_Whitepaper.pdf [accessed 2019-02-01]
29. Fujisaki E, Okamoto T. Secure integration of asymmetric and symmetric encryption schemes. 1999 Presented at: Annual International Cryptology Conference; 1999 Aug 15; Berlin p. 537-554. [doi: 10.1007/3-540-48405-1_34]
30. Hamid L. inventor; Imation Corp, assignee. Cloud-based hardware security modules. United States patent application US 13/826,353 Aug 2013:1-8.
31. Bellare M, Micciancio D, Warinschi B. Foundations of group signatures: Formal definitions, simplified requirements, and a construction based on general assumptions. 2003 Presented at: International Conference on the Theory and Applications of Cryptographic Techniques; 2003 May 4; Warsaw p. 614-629. [doi: 10.1007/3-540-39200-9_38]
32. Belenkiy M, Chase M, Kohlweiss M, Lysyanskaya A. P-signatures and noninteractive anonymous credentials. 2008 Presented at: Theory of Cryptography; 2008; New York p. 8523-8374. [doi: 10.1007/978-3-540-78524-8_20]
33. Ateniese G, Magri B, Venturi D, Andrade E. Redactable blockchain?or?rewriting history in bitcoin and friends. 2017 Presented at: IEEE European Symposium on Security and Privacy (EuroS&P); 2017 Apr 26; Paris p. 111-126. [doi: 10.1109/eurosp.2017.37]
34. Hylock RH, Zeng X. A Blockchain Framework for Patient-Centered Health Records and Exchange (HealthChain): Evaluation and Proof-of-Concept Study. J Med Internet Res 2019 Aug 31;21(8):1-28 [FREE Full text] [doi: 10.2196/13592] [Medline: 31471959]
35. Magyar G. A new disruptive technology in health data management. Blockchainolving the privacy and research availability tradeoff for EHR data; 2017 Presented at: IEEE 30th Neumann Colloquium (NC); 2017; Budapest. [doi: 10.1109/nc.2017.8263269]

36. Paranjape K, Parker M, Houlding D, Car J. Implementation Considerations for Blockchain in Healthcare Institutions. BHTY 2019 Jul 04;2:1-15. [doi: 10.30953/bhty.v2.114]
37. Steffan DN. MultiChain: A cybercurrency for cooperation. MS thesis, TU Delft, Delft University of Technology. 2015. URL: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi3s_-gvYnrAhVFrosKHd-ECbwQFjAAegQIARAB&url=https%3A%2F%2Frepository.tudelft.nl%2Fislandora%2Fobject%2Fuuid%3A59723e98-ae48-4fac-b258-2df99d11012c%2Fdatastream%2FOBJ%2Fdownload&usg=AOvVaw2e WDuIcmKItxYHBw3wlsi0 [accessed 2019-02-01]
38. Motohashi T, Hirano T, Okumura K, Kashiyama M, Ichikawa D, Ueno T. Secure and Scalable mHealth Data Management Using Blockchain Combined With Client Hashchain: System Design and Validation. J Med Internet Res 2019 May 16;21(5):e13385. [doi: 10.2196/13385]
39. Li P, Nelson SD, Malin BA, Chen Y. DMMS: A Decentralized Blockchain Ledger for the Management of Medication Histories. BHTY 2018 Dec 31;2:1-15. [doi: 10.30953/bhty.v2.38]
40. Hylock RH, Zeng X. A Blockchain Framework for Patient-Centered Health Records and Exchange (HealthChain): Evaluation and Proof-of-Concept Study. J Med Internet Res 2019 Aug 31;21(8):e13592. [doi: 10.2196/13592]
41. Ateniese G, Magri B, Venturi D, Andrade E. Redactable blockchain?or?rewriting history in bitcoin and friends. 2017 Presented at: IEEE European Symposium on Security and Privacy (EuroS&P); 2017 Apr 26; Saarbrucken p. 111-126. [doi: 10.1109/eurosp.2017.37]
42. Pournaghi SM, Bayat M, Farjami Y. MedSBA: a novel and secure scheme to share medical data based on blockchain technology and attribute-based encryption. J Ambient Intell Human Comput 2020 Jan 21:1-18. [doi: 10.1007/s12652-020-01710-y]
43. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. How blockchain could empower ehealth: An application for radiation oncology. 2016 Presented at: VLDB Workshop on Data Management and Analytics for Medicine and Healthcare. Springer, Cham; 2017 Sep 1; Munich p. 3-6. [doi: 10.1007/978-3-319-67186-4_1]
44. Hardjono T, Lipton A, Pentland A. Towards a design philosophy for interoperable blockchain systems. arXiv preprint arXiv.05934 May 2018:15.

Abbreviations

AWS: Amazon Web Services
CA: certificate authority
CS: cloud storage
ECA: enrollment certificate authority
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
HIE: health information exchange
HIPAA: Health Insurance Portability and Accountability Act
MS: membership service
PBFT: Practical Byzantine Fault Tolerance
PKI: public key infrastructure
SU: solution user
TCA: transaction certificate authority
TLS-CA: transport layer security certificate authority
UI: user interface

Edited by P Zhang, K Clauson; submitted 01.02.19; peer-reviewed by G Lizarraga, R Hylock, M Görges; comments to author 27.04.19; revised version received 05.10.19; accepted 30.05.20; published 21.08.20.

Please cite as:
Dubovitskaya A, Baig F, Xu Z, Shukla R, Zambani PS, Swaminathan A, Jahangir MM, Chowdhry K, Lachhani R, Idnani N, Schumacher M, Aberer K, Stoller SD, Ryu S, Wang F
ACTION-EHR: Patient-Centric Blockchain-Based Electronic Health Record Data Management for Cancer Care
J Med Internet Res 2020;22(8):e13598
URL: http://www.jmir.org/2020/8/e13598/
doi:10.2196/13598
PMID:32821064

Letter to the Editor

Comment on “Internet-Based Cognitive Behavioral Therapy With Real-Time Therapist Support via Videoconference for Patients With Obsessive-Compulsive Disorder, Panic Disorder, and Social Anxiety Disorder: Pilot Single-Arm Trial”

David Zargaran[1], BSc, MBBS, MRCS; Caoimhe Walsh[2], BSc, MBBS, MRCS; Foteini Stefania Koumpa[2], BSc, MBBS, MRCS; Muhammad Arsalan Ashraf[2], BSc, MBBS, MRCS; Amelia Jayne White[2], BSc, MBBS; Nikhil Patel[3], BSc, MBBS, MRCS; Ravina Tanna[4], BSc, MBBS, MRCS; Anna Trepekli[5], BSc, MBBS; Alexander Zargaran[6], BSc, MBBS

[1]St Thomas' Hospital, London, United Kingdom
[2]London Deanery, London, United Kingdom
[3]Oxford Deanery, Oxford, United Kingdom
[4]Luton and Dunstable Hospital, London, United Kingdom
[5]Univesity College Hospital, London, United Kingdom
[6]St George's, University of London, London, United Kingdom

Corresponding Author:
David Zargaran, BSc, MBBS, MRCS
St Thomas' Hospital
Westminster Bridge Rd,
London, SE1 7PH
United Kingdom
Phone: 44 02071887188
Email: Davidzargaran@gmail.com

Related Article:

Comment on: https://www.jmir.org/2018/12/e12091

(J Med Internet Res 2020;22(8):e13234) doi:10.2196/13234

KEYWORDS

internet; CBT; cognitive behavioral therapy; telemedicine; telehealth

We read with great interest the pilot single arm trial by Matsumoto et al [1], which examined the role of internet-based cognitive behavioral therapy (ICBT) via video conference. This paper is both timely and exciting as it assesses the impact of the delivery of remote cognitive behavioral therapy (CBT) to increase access to health care by leveraging improved telecommunication infrastructure.

The authors begin by explaining that difficulties with accessing CBT are due to the expense, number, and uneven urban distribution of therapists, and feel that telemental health has resolved these issues. We would comment that whilst distribution will have been addressed, the use of video conferencing alone would likely be insufficient to address the expense and the paucity of therapists. Therefore, this represents an apparent limitation on the scalability of such interventions. Varying the level of interaction with the therapist through the use of computer-assisted CBT (an empirically developed computer program that builds on CBT skills) has already been described with positive results both in terms of outcomes and cost-effectiveness in a population of primary care patients undergoing treatment for depression [2]. Future work could look to offer a combination of both computer-assisted CBT and video conferencing as a newly derived treatment regimen to assess the efficacy of this combination.

The equivalence of face-to-face CBT and ICBT in a meta-analysis of 1418 patients [3] further strengthens the validity of the conclusions drawn by Matsumoto et al [1]. However, with regards to the methodology employed, the pretrial use of medication duration did not appear to be controlled for, and the article did not state the timing of the interventions. Knowledge of concomitant use of medication could help further delineate the role of ICBT alone in assessing

the overall impact. Furthermore, an appreciation of the timing of the CBT sessions could offer therapeutic information as a potential datapoint for the efficacy of CBTs. With specific knowledge of the time an ICBT session took place, one could assess the efficacy of the ICBT session in working hours relative to out-of-hours. Further, the compliance to treatment regimens could be assessed given the increased flexibility of the session appointment times.

Although the single-arm nature of the trial limits the applicability of the conclusions, the authors describe a preference (83% of participants) for video conferencing over face-to-face CBT. More information as to the rationale belying this preference would help inform future intervention strategies. Furthermore, previous experience and perceptions toward face-to-face CBT before an intervention could also help further elucidate the validity of participants' preference toward video conferencing. One potential explanation could be that the method of participant recruitment and selection could introduce a bias of intrinsic preference of trial subjects toward seeking video-based CBT, given that one recruitment strategy employed by the authors was web-based advertising.

Although a need for further clarification exists, the authors are to be commended on their unique contribution to the field, and it is with great excitement that we anticipate future works to further describe the potential benefits of ICBT and its applications.

Conflicts of Interest

None declared.

References

1. Matsumoto K, Sutoh C, Asano K, Seki Y, Urao Y, Yokoo M, et al. Internet-Based Cognitive Behavioral Therapy With Real-Time Therapist Support via Videoconference for Patients With Obsessive-Compulsive Disorder, Panic Disorder, and Social Anxiety Disorder: Pilot Single-Arm Trial. J Med Internet Res 2018 Dec 17;20(12):e12091 [FREE Full text] [doi: 10.2196/12091] [Medline: 30559094]
2. Antle BF, Owen JJ, Eells TD, Wells MJ, Harris LM, Cappiccie A, et al. Dissemination of computer-assisted cognitive-behavior therapy for depression in primary care. Contemp Clin Trials 2019 Mar;78:46-52. [doi: 10.1016/j.cct.2018.11.001] [Medline: 30572162]
3. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]

Abbreviations

CBT: cognitive behavioral therapy
ICBT: internet-based cognitive behavioral therapy

Edited by C Eichenberg; submitted 23.12.18; peer-reviewed by K Matsumoto, S Zheng; comments to author 24.01.19; revised version received 16.02.19; accepted 17.02.19; published 12.08.20.

Please cite as:
Zargaran D, Walsh C, Koumpa FS, Ashraf MA, White AJ, Patel N, Tanna R, Trepekli A, Zargaran A
Comment on "Internet-Based Cognitive Behavioral Therapy With Real-Time Therapist Support via Videoconference for Patients With Obsessive-Compulsive Disorder, Panic Disorder, and Social Anxiety Disorder: Pilot Single-Arm Trial"
J Med Internet Res 2020;22(8):e13234
URL: https://www.jmir.org/2020/8/e13234
doi:10.2196/13234
PMID:32784172

Corrigenda and Addenda

Correction: The Relationship Between Health Management and Information Behavior Over Time: A Study of the Illness Journeys of People Living With Fibromyalgia

Annie T Chen[1], MSIS, PhD

University of Washington School of Medicine, Department of Biomedical Informatics and Medical Education, Seattle, WA, United States

Corresponding Author:
Annie T Chen, MSIS, PhD
University of Washington School of Medicine
Department of Biomedical Informatics and Medical Education
850 Republican St., Box 358047
Seattle, WA, 98109
United States
Phone: 1 206 221 9218
Email: atchen@uw.edu

Related Article:

Correction of: https://www.jmir.org/2016/10/e269/

(J Med Internet Res 2020;22(8):e23597) doi:10.2196/23597

In "The Relationship Between Health Management and Information Behavior Over Time: A Study of the Illness Journeys of People Living With Fibromyalgia" (JMIR 2016;18(10):e269) the author noted an error in Table 2. Some values under the section "Employment status" were listed incorrectly.

Values for "Student" were originally listed as:

n=11, 47.8%

The correct values for "Student" are:

n=3, 13.0%

Values for "Not employed" were originally listed as:

n=3, 13.0%

The correct values for "Not employed" are:

n=1, 4.3%

The correction will appear in the online version of the paper on the JMIR Publications website on August 20, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 17.08.20; this is a non–peer-reviewed article;accepted 17.08.20; published 20.08.20.

Please cite as:
Chen AT
Correction: The Relationship Between Health Management and Information Behavior Over Time: A Study of the Illness Journeys of People Living With Fibromyalgia
J Med Internet Res 2020;22(8):e23597
URL: http://www.jmir.org/2020/8/e23597/
doi:10.2196/23597
PMID:32816686

Corrigenda and Addenda

Correction: Online Guide for Electronic Health Evaluation Approaches: Systematic Scoping Review and Concept Mapping Study

Tobias N Bonten[1,2*], MD, PhD; Anneloek Rauwerdink[3*], MD; Jeremy C Wyatt[4], MD, PhD; Marise J Kasteleyn[1,2], PhD; Leonard Witkamp[5,6], MD, PhD; Heleen Riper[7], PhD; Lisette JEWC van Gemert-Pijnen[8], PhD; Kathrin Cresswell[9], PhD; Aziz Sheikh[9], PhD; Marlies P Schijven[3], MD, PhD; Niels H Chavannes[1,2], MD, PhD; EHealth Evaluation Research Group[10]

[1]Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, Netherlands
[2]National eHealth Living Lab, Leiden, Netherlands
[3]Department of Surgery, Amsterdam Gastroenterology and Metabolism, Amsterdam UMC, Amsterdam, Netherlands
[4]Wessex Institute, University of Southampton, Southampton, United Kingdom
[5]Department of Medical Informatics, Amsterdam UMC, Amsterdam, Netherlands
[6]Ksyos Health Management Research, Amstelveen, Netherlands
[7]Department of Clinical, Neuro and Developmental Psychology, Vrije Universiteit, Amsterdam, Netherlands
[8]Department of Psychology, Health and Technology, Centre for eHealth and Wellbeing Research, University of Twente, Enschede, Netherlands
[9]Centre of Medical Informatics, Usher Institute, The University of Edinburgh, Medical School, Edinburgh, United Kingdom
[10]Please see acknowledgements section for list of collaborators
[*]these authors contributed equally

Corresponding Author:
Tobias N Bonten, MD, PhD
Department of Public Health and Primary Care
Leiden University Medical Centre
Department of Public Health & Primary Care, Room V6-22
PO Box 9600
Leiden, 2300 RC
Netherlands
Phone: 31 715268433
Email: t.n.bonten@lumc.nl

Related Article:

Correction of: https://www.jmir.org/2020/8/e17774

(J Med Internet Res 2020;22(8):e23642) doi:10.2196/23642

In “Online Guide for Electronic Health Evaluation Approaches: Systematic Scoping Review and Concept Mapping Study (“J Med Internet Res 2020;22(8):e17774”) the authors noted one error.

The metadata erroneously listed the group author "EHealth Evaluation Research Group" as having contributed equally. This has been corrected, and the group author is no longer listed as having contributed equally.

The correction will appear in the online version of the paper on the JMIR Publications website on August 21, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories

Submitted 18.08.20; this is a non–peer-reviewed article;accepted 18.08.20; published 21.08.20.

Please cite as:
Bonten TN, Rauwerdink A, Wyatt JC, Kasteleyn MJ, Witkamp L, Riper H, van Gemert-Pijnen LJEWC, Cresswell K, Sheikh A, Schijven MP, Chavannes NH, EHealth Evaluation Research Group
Correction: Online Guide for Electronic Health Evaluation Approaches: Systematic Scoping Review and Concept Mapping Study
J Med Internet Res 2020;22(8):e23642
URL: http://www.jmir.org/2020/8/e23642/
doi:10.2196/23642
PMID:32822315

Corrigenda and Addenda

Correction: How Can Artificial Intelligence Make Medicine More Preemptive?

Usman Iqbal[1,2], PharmD, MBA, PhD; Leo Anthony Celi[3,4,5], MD, MSc; Yu-Chuan Jack Li[2,6,7,8], MD, PhD

[1]Master Program in Global Health & Development, PhD Program in Global Health & Health Security, College of Public Health, Taipei Medical University, Taipei, Taiwan

[2]International Center for Health Information Technology, Taipei Medical University, Taipei, Taiwan

[3]Laboratory for Computational Physiology, Massachusetts Institute of Technology, Cambridge, MA, United States

[4]Division of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States

[5]Department of Biostatistics, Harvard TH Chan School of Public Health, Harvard University, Boston, MA, United States

[6]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[7]Department of Dermatology, Taipei Municipal Wan-Fang Hospital, Taipei, Taiwan

[8]International Medical Informatics Association, Geneva, Switzerland

Corresponding Author:
Yu-Chuan Jack Li, MD, PhD
Graduate Institute of Biomedical Informatics
College of Medical Science and Technology
Taipei Medical University
No 172-1, Sec 2, Keelung Rd
Daan District
Taipei, 10675
Taiwan
Phone: 886 6638 2736 ext 7601
Email: jack@tmu.edu.tw

Related Article:

Correction of: https://www.jmir.org/2020/8/e17211

(J Med Internet Res 2020;22(8):e23645) doi:10.2196/23645

In “How Can Artificial Intelligence Make Medicine More Preemptive?” (J Med Internet Res 2020;22(8):e17211) the authors noted one error.

The first affiliation for author Usman Iqbal originally read:

> *College of Public Health, Taipei Medical University, Taipei, Taiwan*

This was incomplete, and has been updated to read:

> *Master Program in Global Health & Development, PhD Program in Global Health & Health Security, College of Public Health, Taipei Medical University, Taipei, Taiwan*

The correction will appear in the online version of the paper on the JMIR Publications website on August 26, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories

Submitted 18.08.20; this is a non–peer-reviewed article;accepted 18.08.20; published 26.08.20.

Please cite as:
Iqbal U, Celi LA, Li YCJ
Correction: How Can Artificial Intelligence Make Medicine More Preemptive?
J Med Internet Res 2020;22(8):e23645
URL: http://www.jmir.org/2020/8/e23645/
doi:10.2196/23645
PMID:32845851

Corrigenda and Addenda

Correction: Technology-Based Interventions in Oral Anticoagulation Management: Meta-Analysis of Randomized Controlled Trials

Hengfen Dai[1*], MS; Caiyun Zheng[2*], MS; Chun Lin[3*], PhD; Yan Zhang[1], PhD; Hong Zhang[1], BS; Fan Chen[1], BS; Yunchun Liu[1], BS; Jingwen Xiao[1], MS; Chaoxin Chen[4], MS

[1]Affiliated Fuzhou First Hospital of Fujian Medical University, Fuzhou, China

[2]Fuqing City Hospital, Fuzhou, China

[3]School of Basic Medical Sciences, Fujian Medical University, Fuzhou, China

[4]School of Pharmacy, Fujian Medical University, Fuzhou, China

[*]these authors contributed equally

Corresponding Author:
Caiyun Zheng, MS
Fuqing City Hospital
267 Qingrong Avenue
Fuqing
Fuzhou, 350300
China
Phone: 86 059185153395
Email: saiun1992@163.com

Related Article:

Correction of: https://www.jmir.org/2020/7/e18386/

(J Med Internet Res 2020;22(8):e22761) doi:10.2196/22761

In "Technology-Based Interventions in Oral Anticoagulation Management: Meta-Analysis of Randomized Controlled Trials" (JMIR 2020;22(7):e18386) the authors noted an error in the order of authorship. The authors wish to change the order of the first and second authors so that Hengfen Dai is first author and Caiyun Zheng is second author.

The previous order of authorship was as follows:

> *Caiyun Zheng, Hengfen Dai, Chun Lin, Yan Zhang, Hong Zhang, Fan Chen, Yunchun Liu, Jingwen Xiao, Chaoxin Chen*

The correct order of authorship is as follows:

> *Hengfen Dai, Caiyun Zheng, Chun Lin, Yan Zhang, Hong Zhang, Fan Chen, Yunchun Liu, Jingwen Xiao, Chaoxin Chen*

Additionally, affiliations have been reordered in accordance with the correct authorship order. Affiliations were previously listed as:

> *[1]Fuqing City Hospital, Fuzhou, China*
>
> *[2]Affiliated Fuzhou First Hospital of Fujian Medical University, Fuzhou, China*
>
> *[3]School of Basic Medical Sciences, Fujian Medical University, Fuzhou, China*
>
> *[4]School of Pharmacy, Fujian Medical University, Fuzhou, China*

The revised order of affiliations is:

> *[1]Affiliated Fuzhou First Hospital of Fujian Medical University, Fuzhou, China*
>
> *[2]Fuqing City Hospital, Fuzhou, China*
>
> *[3]School of Basic Medical Sciences, Fujian Medical University, Fuzhou, China*
>
> *[4]School of Pharmacy, Fujian Medical University, Fuzhou, China*

Hengfen Dai, Caiyun Zheng, and Chun Lin remain credited with equal contribution.

The correction will appear in the online version of the paper on the JMIR Publications website on August 4, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 22.07.20; this is a non–peer-reviewed article;accepted 24.07.20; published 04.08.20.

Please cite as:
Dai H, Zheng C, Lin C, Zhang Y, Zhang H, Chen F, Liu Y, Xiao J, Chen C
Correction: Technology-Based Interventions in Oral Anticoagulation Management: Meta-Analysis of Randomized Controlled Trials
J Med Internet Res 2020;22(8):e22761
URL: http://www.jmir.org/2020/8/e22761/
doi:10.2196/22761
PMID:32750007

Corrigenda and Addenda

Correction: Characterizing the Rural Opioid Use Environment in Kentucky Using Google Earth: Virtual Audit

Natalie Danielle Crawford[1], PhD; Regine Haardörfer[1], PhD; Hannah Cooper[1], SCD; Izraelle McKinnon[2], MPH; Carla Jones-Harrell[1], MUEP, MPH; April Ballard[3], MPH; Sierra Shantel von Hellens[4], BA; April Young[3], PhD

[1]Department of Behavioral Sciences and Health Education, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[2]Department of Epidemiology, Emory University, Atlanta, GA, United States

[3]Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[4]Morehead State University, Morehead, KY, United States

Corresponding Author:
Natalie Danielle Crawford, PhD
Department of Behavioral Sciences and Health Education
Rollins School of Public Health
Emory University
1518 Clifton Road
Atlanta, GA
United States
Phone: 1 4047429445
Email: ndcrawford@emory.edu

Related Article:

Correction of: https://www.jmir.org/2019/10/e14923/

(J Med Internet Res 2020;22(8):e21442) doi:10.2196/21442

The authors of "Characterizing the Rural Opioid Use Environment in Kentucky Using Google Earth: Virtual Audit" (J Med Internet Res 2019;21(10):e14923), the authors noticed four errors in the contributing authors' names and degrees.

Firstly, the name of the second author was misspelled and has been revised from

Regine Haardöerfer

to

Regine Haardörfer.

Next, the degree listed for Hannah Cooper has been revised from:

Hannah Cooper, PHD

to

Hannah Cooper, SCD

Then, the degree listed for April Ballard has been revised from:

April Ballard, BA

to

April Ballard, MPH

Lastly, the academic degrees listed for Carla Jones-Harrell were incorrect. Her degree listing has been revised from:

Carla Jones-Harrell, MPhil

to

Carla Jones-Harrell, MUEP, MPH.

These corrections will appear in the online version of the paper on the JMIR website together with the publication of this correction notice. Because this change was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 15.06.20; this is a non–peer-reviewed article;accepted 24.06.20; published 12.08.20.

Please cite as:
Crawford ND, Haardörfer R, Cooper H, McKinnon I, Jones-Harrell C, Ballard A, von Hellens SS, Young A
Correction: Characterizing the Rural Opioid Use Environment in Kentucky Using Google Earth: Virtual Audit
J Med Internet Res 2020;22(8):e21442
URL: https://www.jmir.org/2020/8/e21442
doi:10.2196/21442
PMID:32784194

Original Paper

Cross-Country Comparison of Public Awareness, Rumors, and Behavioral Responses to the COVID-19 Epidemic: Infodemiology Study

Zhiyuan Hou[1,2], PhD; Fanxing Du[1], MPH; Xinyu Zhou[1], BSc; Hao Jiang[1], BSc; Sam Martin[3], PhD; Heidi Larson[4,5], PhD; Leesa Lin[4], PhD

[1]School of Public Health, Fudan University, Shanghai, China

[2]Global Health Institute, NHC Key Laboratory of Health Technology Assessment, Fudan University, Shanghai, China

[3]Centre for Clinical Vaccinology and Tropical Medicine, University of Oxford, Oxford, United Kingdom

[4]Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom

[5]Department of Health Metrics Sciences, University of Washington, Seattle, WA, United States

Corresponding Author:
Zhiyuan Hou, PhD
School of Public Health
Fudan University
Mailbox 250, 138# Yixueyuan Road, Xuhui District
Shanghai
China
Phone: 86 21 33563935
Email: zyhou@fudan.edu.cn

Abstract

Background: Understanding public behavioral responses to the coronavirus disease (COVID-19) epidemic and the accompanying infodemic is crucial to controlling the epidemic.

Objective: The aim of this study was to assess real-time public awareness and behavioral responses to the COVID-19 epidemic across 12 selected countries.

Methods: Internet surveillance was used to collect real-time data from the general public to assess public awareness and rumors (China: Baidu; worldwide: Google Trends) and behavior responses (China: Ali Index; worldwide: Google Shopping). These indices measured the daily number of searches or purchases and were compared with the numbers of daily COVID-19 cases. The trend comparisons across selected countries were observed from December 1, 2019 (prepandemic baseline) to April 11, 2020 (at least one month after the governments of selected countries took actions for the pandemic).

Results: We identified missed windows of opportunity for early epidemic control in 12 countries, when public awareness was very low despite the emerging epidemic. China's epidemic and the declaration of a public health emergency of international concern did not prompt a worldwide public reaction to adopt health-protective measures; instead, most countries and regions only responded to the epidemic after their own case counts increased. Rumors and misinformation led to a surge of sales in herbal remedies in China and antimalarial drugs worldwide, and timely clarification of rumors mitigated the rush to purchase unproven remedies.

Conclusions: Our comparative study highlights the urgent need for international coordination to promote mutual learning about epidemic characteristics and effective control measures as well as to trigger early and timely responses in individual countries. Early release of official guidelines and timely clarification of rumors led by governments are necessary to guide the public to take rational action.

(J Med Internet Res 2020;22(8):e21143) doi:10.2196/21143

KEYWORDS

COVID-19; internet; surveillance; infodemic; infodemiology; infoveillance; Google Trends; public response; behavior; rumor; trend

Introduction

In early December 2019, the then-unnamed novel coronavirus disease (COVID-19) emerged in Wuhan City and spread rapidly across China [1,2]. On January 23, 2020, the Chinese government placed Wuhan and several nearby cities under quarantine and implemented containment measures to slow community transmission [3]. The World Health Organization (WHO) declared the outbreak to be a public health emergency of international concern (PHEIC) on January 30, when there were almost 8000 confirmed cases worldwide; all but 98 of these cases, along with all 170 COVID-19–related deaths, were in China. The PHEIC declaration stressed the risk the virus posed to countries beyond China and the need for a more coordinated international response to the outbreak. On March 11, the WHO declared COVID-19 a global pandemic, with a worldwide confirmed case count of 118,000 in 114 countries and the death toll reaching more than 4200. At that time, more than 90% of cases were localized in four countries, including China and South Korea, where local epidemics had significantly declined [4]. By the end of April, the epidemic had spread to 213 countries with more than 3.02 million confirmed cases, leading to at least 208,000 deaths [5]. The United States reported the highest numbers of confirmed cases and deaths, followed by Spain, Italy, the United Kingdom, Germany, and France; all these countries reported over 120,000 cases and 24,000 deaths except for Germany, which reported approximately 6000 deaths [5].

During an epidemic, it is crucial to understand how critical information about the health threat is disseminated and how the public processes and responds to this information [6,7]. The risks and uncertainties of emerging infectious diseases may arouse public awareness and prompt either constructive behavior (eg, employing personal hand hygiene and avoiding mass gatherings) or disruptive behavior (eg, panic buying and adopting unproven treatments) [7,8]. COVID-19 has triggered the spread of rumors and misinformation through social media regarding unproven remedies, which has induced public stress and panic [9,10]. In addition, the public may respond differently to an epidemic across countries. In the early stage of the COVID-19 epidemic, people in Asian countries immediately began to wear face masks; however, Europeans and North Americans opposed this practice [11]. Messages from health authorities and evidence of the effectiveness of masks against COVID-19 are conflicting [12]. It is necessary to understand why and how the public responds to COVID-19–related information, which will further inform government risk communication and appropriate official guidelines [13].

The power of internet search data is being increasingly recognized in public health emergencies [14,15]. In contrast with traditional surveys, internet surveillance can systematically track public responses to epidemics in real time and is less likely to be affected by recall bias [16]. Despite these benefits, the role of internet surveillance (also called infoveillance or infodemiology) in monitoring public behavioral responses and rumors during an epidemic is still underexplored [17-19]. Using internet surveillance data from China and worldwide, this study aimed to assess the public awareness and behavioral responses in real time during the first 100 days of the COVID-19 epidemic. This study compares the governmental and public responses across selected countries and provides insights on the control of COVID-19 and future epidemics.

Methods

Study Setting

In this study, we conducted internet surveillance in 12 countries. In addition to China, three countries in East and Southeast Asia that were affected by the first wave of the pandemic were selected, namely Japan, South Korea, and Singapore, followed by four European countries (Italy, France, Spain, and the United Kingdom) and the United States. Additionally, Brazil, South Africa, and India, where internet surveillance data (Google Trends) are available, were selected from Latin America, Africa, and South Asia, respectively. All data are publicly available.

Data Collection

Internet surveillance was used to collect real-time data from the general public to assess public awareness and rumors (China: Baidu; worldwide: Google Trends) and behavior responses (China: Ali Index; worldwide: Google Shopping). The data cover the period from December 1, 2019 (prepandemic baseline) to April 11, 2020 (at least one month after the governments of selected countries took actions to address the COVID-19 pandemic). Table 1 lists the keywords used to measure public awareness, rumors, and behavioral responses to the COVID-19 epidemic.

Table 1. Keywords searched for public awareness, behavioral responses, and rumors regarding the COVID-19 epidemic.

Domain	Keywords
Awareness of COVID-19[a]	*coronavirus* (冠状病毒), *Wuhan pneumonia* (武汉肺炎)
Behavioral response to protection measures	*mask* (口罩), *hand sanitizer* (洗手液), *disinfectant* (消毒液/消毒剂/消毒水), *thermometer* (体温计)
Rumors	
Worldwide	*chloroquine*, *hydroxychloroquine*
China	*radix isatidis* (板蓝根), *Shuanghuanglian* (双黄连), *garlic* (大蒜)

[a]COVID-19: coronavirus disease.

Google Trends

Google Trends can provide insight into the relative search volumes of search terms on Google on a daily basis [20]. Depending on the source of the search, Google Trends can be further divided into web search trends and Google Shopping trends, and these trends were highly correlated (Multimedia Appendix 1). The Google Trends index (web search) on the topic of *coronavirus* was used to assess the awareness of COVID-19 among the general public, whereas the Google Shopping indices on two topics, *mask* and *hand sanitizer*, were used to assess the adoption of personal protection measures. In countries where Google Shopping data were limited for assessing public purchasing behavior (including Japan, Singapore, South Korea, Italy, and Spain), Google web search data were used as an alternative. Google Trends presents relative search volumes ranging from 0 to 100 (the maximum daily search volume on specific terms is standardized as 100%).

The antimalarial drugs chloroquine and hydroxychloroquine have been repeatedly mentioned by world-leading politicians as treatments for COVID-19 without clinical data to support their efficacy; these drugs have potentially deadly side effects [21,22]. By analyzing Google Trends data on both drugs, we assessed the public behavioral response towards misinformation worldwide.

Baidu Index

The Baidu search engine, which is Google's equivalent in China, has more than 1 billion Chinese users. The Baidu Index, which stems from the frequency of searches using the Baidu search engine, is powered by Baidu statistics and exhibited as preset keywords. The Baidu Index reflects the daily number of searches on specific keywords, thereby assessing public awareness and intended behavior. We manually scanned and identified all Baidu Index keywords that included coronavirus, recommended personal protection measures, and rumors and misinformation. For comparison, we also gathered Baidu Index data from the same time period in the previous year as the baseline: December 1, 2018, to April 11, 2019.

Rumors and misinformation circulated widely in China regarding certain herbal medicines for personal health protection. From the Sina Weibo "Hot Search" ranking, we identified three keywords related to these rumors and misinformation: *radix isatidis*, *Shuanghuanglian*, and *garlic*. The Baidu Index on these terms was used to detect public behavioral responses regarding rumors and misinformation in China.

Ali Index

The Ali platform, Amazon's equivalent in China, is a powerful and popular web-based electronic commerce marketplace with more than one billion users, and its purchasing index reflects the number of specific products purchased on the internet. The Ali Index was used to collect behavioral data during the epidemic. The National Health Commission of China issued a series of guidelines recommending that residents adopt personal protection measures consisting of four main aspects: respiratory protection, hand hygiene, home disinfection, and health monitoring [3]. Accordingly, we identified *mask*, *hand sanitizer*, *disinfectant*, and *thermometer* as keywords to reflect the above four measures. We employed these four keywords relevant to recommended personal protection measures to assess public behavioral responses to the COVID-19 epidemic (the Ali platform did not generate an index for rumor-related items).

Data Analysis

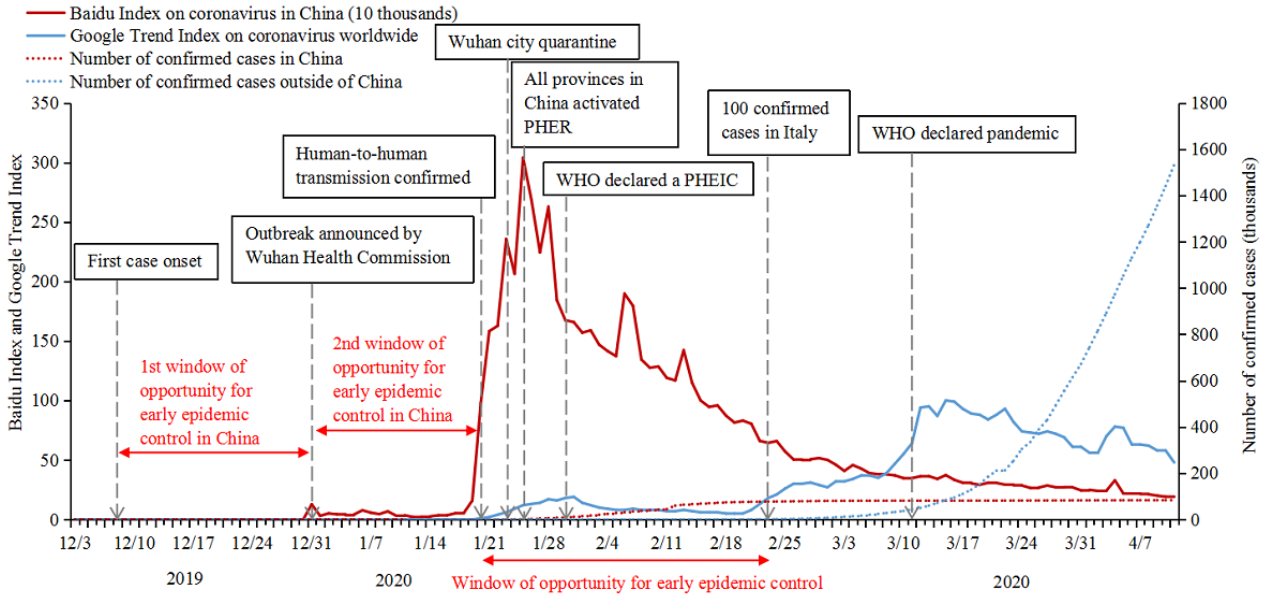
Public awareness on the epidemic was assessed using the Baidu daily indices in China and Google Trends indices worldwide. We assessed the intentions (Baidu Index) and behaviors (Ali Index) regarding the adoption of recommended personal protection measures as well as rumors about ineffective treatments in China; Spearman rank correlation analyses between the Baidu and Ali indices were employed to detect the consistency of intention and behaviors. Google Trends was used to assess the intended behavior regarding personal protection measures and antimalarial drugs worldwide. These indices were used to measure the daily number of searches or purchases related to specific keywords, which we then compared with daily data on numbers of confirmed COVID-19 cases. The Spearman rank correlation was used to examine the correlations between Google Trends for *mask* and *hand sanitizer* and the numbers of new COVID-19 cases in the selected countries. The indices were input into charts to observe the trends along the period of observation and facilitate the trend comparisons across countries.

Results

Public Awareness and Searches Related to COVID-19

The first COVID-19 case emerged on December 8, 2019, and it was first announced by the Wuhan Health Commission on December 31. As shown in Figure 1, the Baidu search index for *coronavirus* showed no results throughout most of December but jumped to 127,336 searches on December 31, 2019, when the Wuhan Health Commission first reported 27 patients with pneumonia of unknown cause. However, this number immediately decreased to 30,504 searches the next day and remained low between January 1 and 19, 2020. Starting on January 20, when the National Health Commission of China confirmed human-to-human transmission, this number increased sharply, and it peaked at 3,039,324 on January 25. Wuhan City implemented a series of quarantine measures starting January 23, followed by 30 provinces (autonomous regions and municipalities) across mainland China (excluding Tibet). After that, public searches steadily decreased with minor fluctuations. Figure 1 highlights two missed windows of opportunity for early epidemic control in China.

Figure 1. Public awareness and searches related to COVID-19 by Baidu Index in China and Google Trends worldwide from December 2019 to April 2020. PHEIC: public health emergency of international concern; PHER: public health emergency response.



Outside of China, although there was a slight increase in public searches following the confirmation of human-to-human transmission of COVID-19, the Google Trends index for *coronavirus* began to decline on January 31, the day after the WHO declared the outbreak a PHEIC. This index continued to decline and remained at a low level until late February, when COVID-19 started to spread in Italy. Figure 1 shows that the world missed an additional 1-month window of opportunity for early epidemic control, even when the number of COVID-19 cases had reached almost 80,000 in China.

Behavioral Responses of Adopting Recommended Personal Protection Measures

The National Health Commission of China first proposed respiratory protection and hand hygiene as protection guidelines on January 21, 2020, and the use of home disinfection and health monitoring on January 22 and 25, respectively [3]. As shown in Figure 2, the Baidu and Ali indices of the four recommended personal protection measures, except for the Ali Index of *thermometer*, all started to increase sharply on January 21, with an exceptional drop on January 24 (Chinese New Year's Eve). The Ali Index of *thermometer* increased rapidly from January 25. The Baidu Index, which indicated an intention to adopt all four measures, increased earlier than the actual behaviors presented in the Ali Index. Both indices remained steadily low during the same period in the previous year (baseline). These results indicate that the Chinese public responded quickly to the issuing of specific guidelines, and both intended and actual purchasing behavior dramatically increased accordingly. Correlation analysis (Multimedia Appendix 2) showed strong positive correlations between the Baidu and Ali indices for all recommended measures (correlation coefficient range of 0.676-0.860, $P<.001$).

Figure 2. Trends of Baidu and Ali indices for recommended personal protection measures in China from December 2019 to April 2020: A. Face mask. B. Hand sanitizer. C. Disinfectant. D. Thermometer.

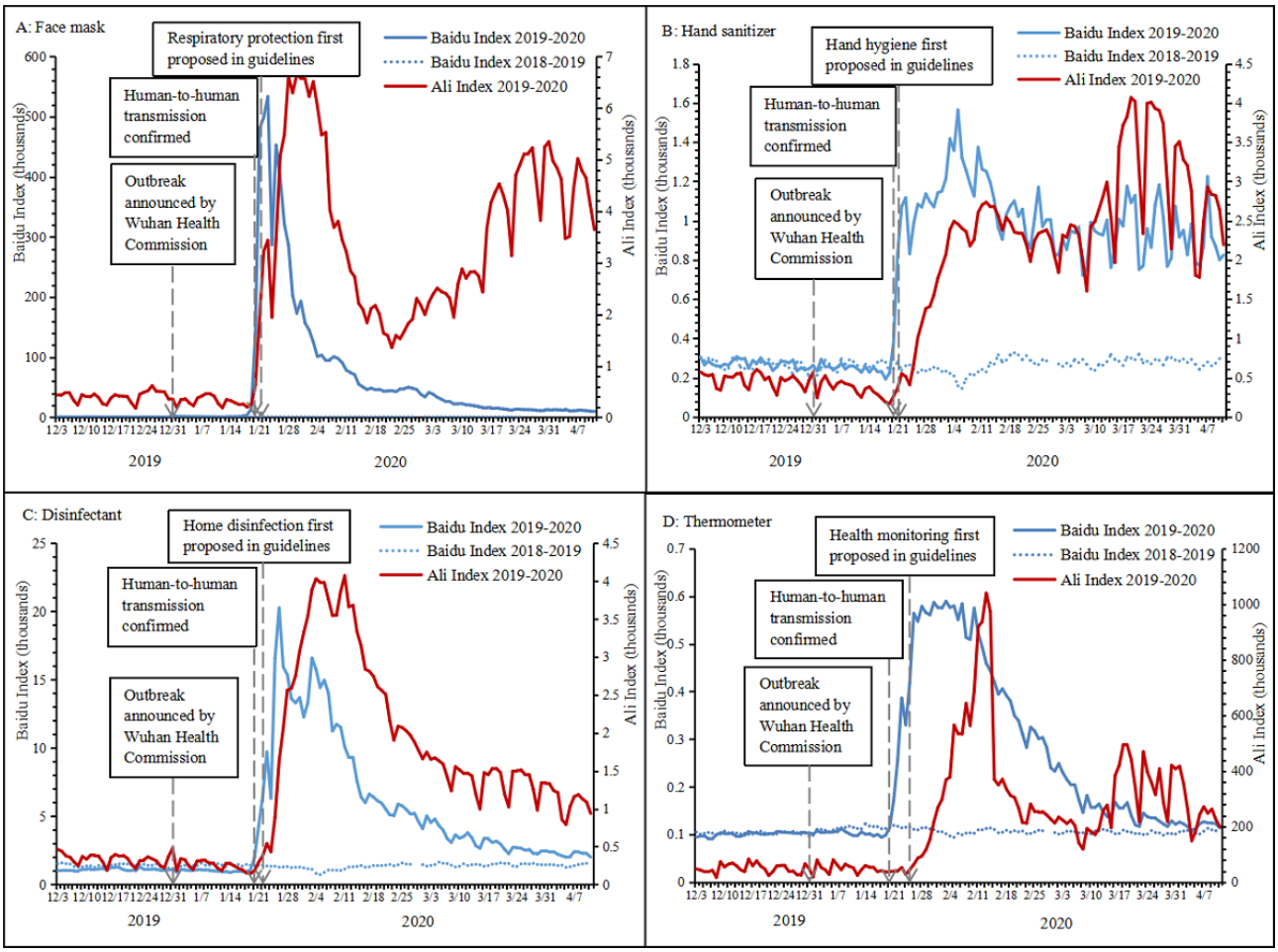
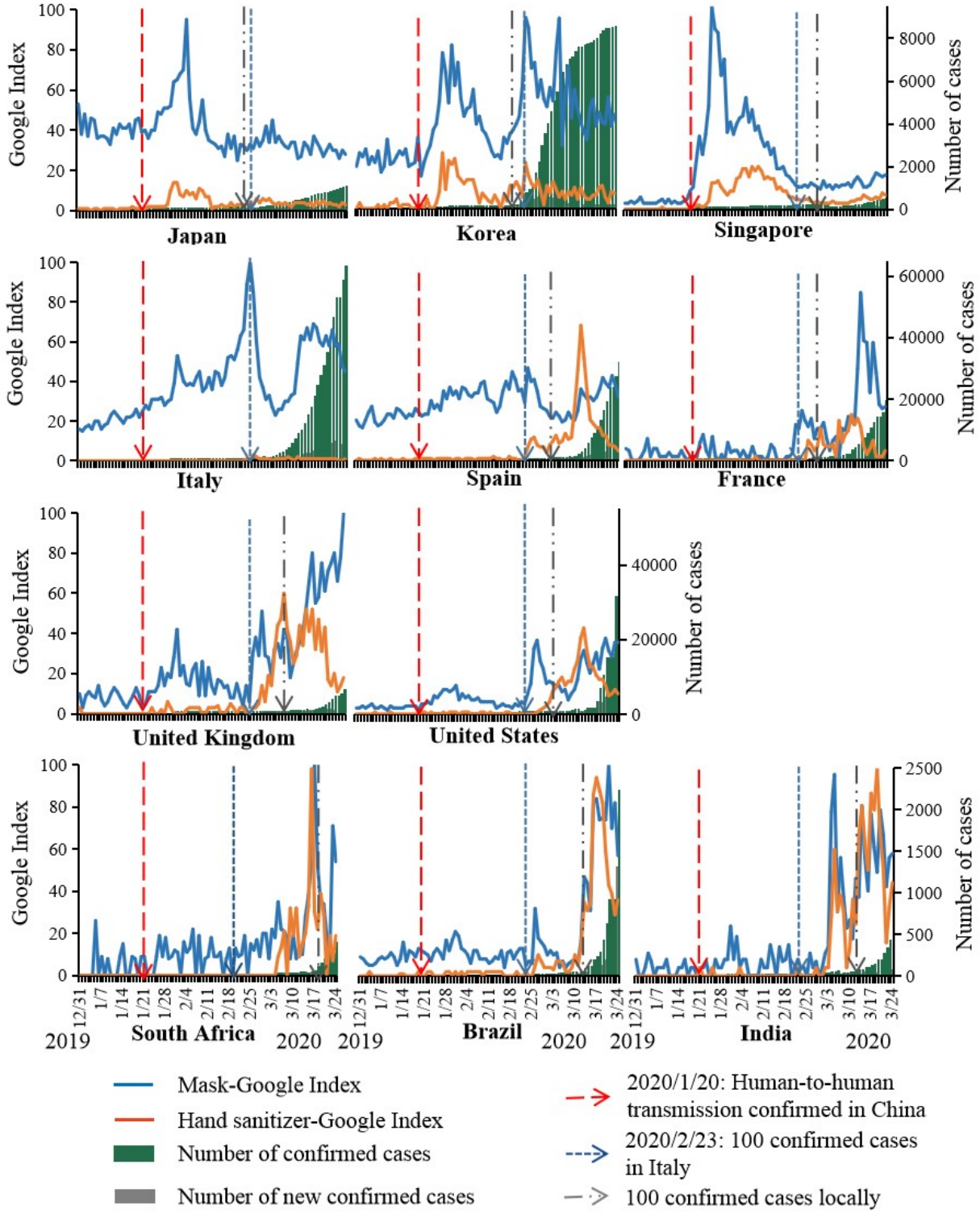


Figure 3 presents a cross-country comparison of Google Trends on mask and hand sanitizer with numbers of COVID-19 cases, and their Spearman rank correlation analyses are shown in Multimedia Appendix 3. All listed countries, except for Brazil and South Africa, started to report COVID-19 cases in late January 2020, whereas Brazil and South Africa reported their first cases on February 26 and March 5, respectively. In Japan, South Korea, and Singapore, Google Trends indices on mask and hand sanitizer quickly increased after human-to-human transmission in China and reached a peak around the end of January. The Google indices in Japan and Singapore then decreased and remained steady while the number of COVID-19 cases in these countries was small; meanwhile, with the rapid increase in the number of cases in South Korea, the Google index sharply increased and reached its second peak in late February. Correlation analysis (Multimedia Appendix 3) showed that the Google Trends for *mask* and *hand sanitizer* in the three Asian countries were significantly correlated with new COVID-19 cases in China rather than local cases.

In addition, the Google indices in European countries and the United States remained low after human-to-human transmission in China and started to increase when the number of COVID-19 cases increased from 11 to 123 in Italy on February 23 (Figure 3). In Italy and Spain, where the epidemic first spread among European countries, the Google Trends data for *hand sanitizer* were highly correlated with the number of local cases, whereas the purchase of masks lagged behind the epidemic for 2 to 3 weeks (Multimedia Appendix 3 and Multimedia Appendix 4). The public in India, Brazil, and South Africa did not respond to the spread of COVID-19 in China and Italy or to the PHEIC, with their Google indices remaining low until early March. Even if there was a small increase in the public response to the epidemic in other countries, the index then quickly decreased and did not start to rise again until the epidemic spread to those countries. Overall, the Google Trends data for *mask* and *hand sanitizer* were in line with the trends of the number of cases in each country (Figure 3 and Multimedia Appendix 3), and these trends reached their own peaks when COVID-19 spread locally. From early April, people around the world started to search for and buy face masks.

Figure 3. Cross-country comparison of Google Trends data for *mask* and *hand sanitizer* with numbers of COVID-19 cases from December 2019 to April 2020.



Public Responses to Rumors and Misinformation on Remedies

Figure 4 shows the trends in the Baidu indices on rumors, indicating intended behavior. The Baidu Index of *radix isatidis*, a traditional Chinese medicine used to treat fever, started to increase when the outbreak was first announced. It further increased sharply on January 20, 2020, when human-to-human transmission of COVID-19 and 291 cases were confirmed, and it reached a peak on January 21. The index started to decline from January 21 to 24, during which the newspaper *People's Daily* issued three reports in an attempt to refute rumors of the effectiveness of radix isatidis against COVID-19. On January 31, *People's Daily* reported that Shuanghuanglian, a Chinese medicine, could inhibit COVID-19, and the Baidu indices of both *Shuanghuanglian* and *radix isatidis* rapidly reached their

peaks. Both indices decreased rapidly on February 2, when *People's Daily* clarified that Shuanghuanglian cannot prevent COVID-19. A rumor that garlic could prevent COVID-19 started to spread on January 21, and the Baidu Index of *garlic* increased accordingly. This index reached a peak on January 27 and declined after January 28, when *People's Daily* first refuted rumors about the protective function of garlic. Representing the baseline, these indices remained steadily low during the same period in the previous year.

Figure 4. Trends of Baidu indices for rumors related to herbal remedies in China from December 2019 to April 2020.

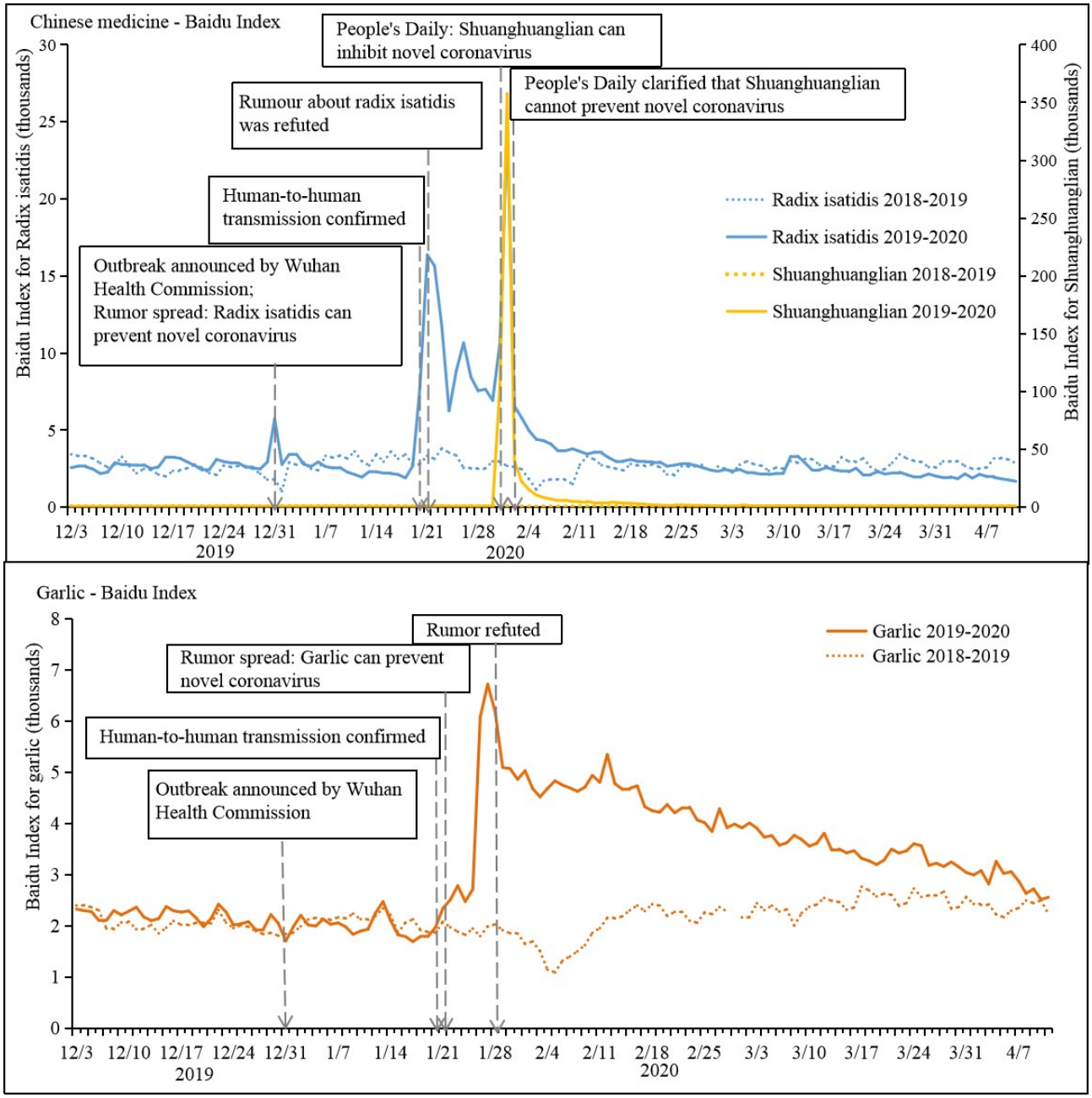
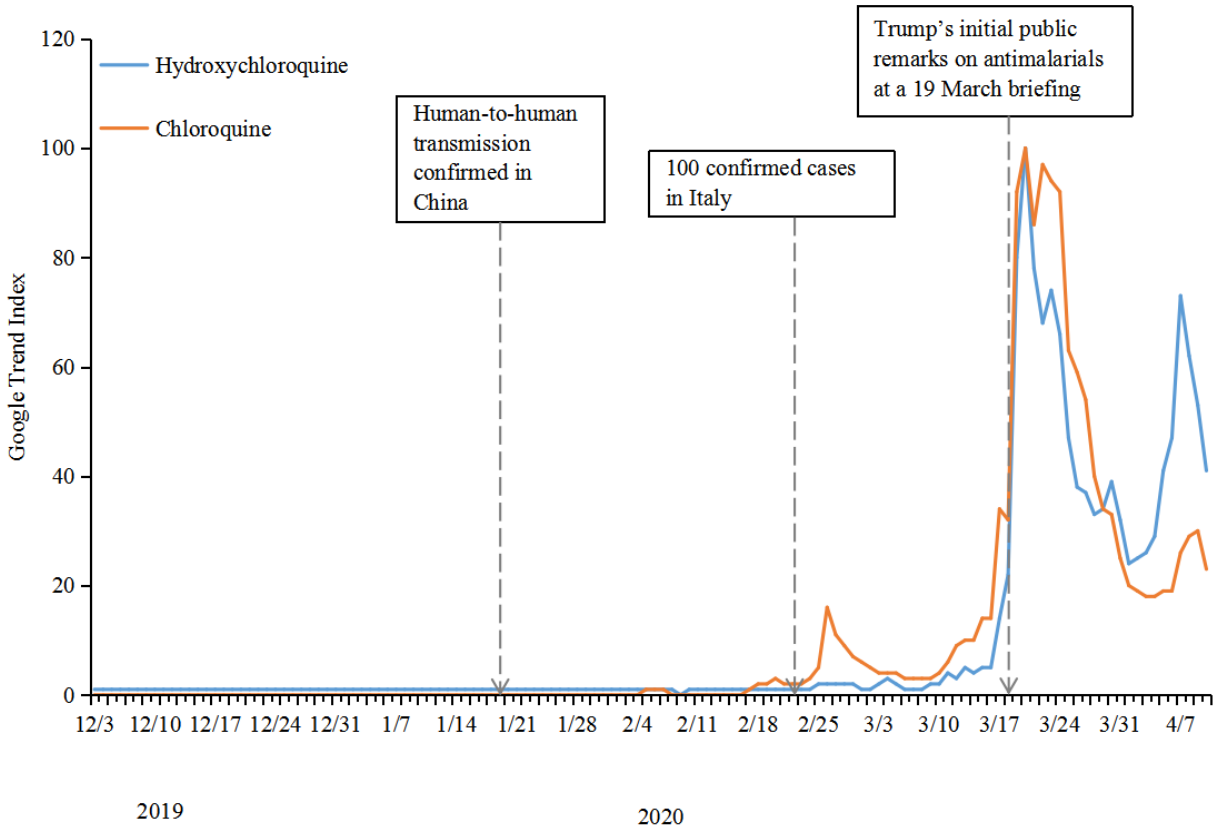


Figure 5 shows the Google Trends data for the antimalarial drugs chloroquine and hydroxychloroquine worldwide. Public searches for *chloroquine* started to increase and reached a small peak in late February, when COVID-19 started to spread in Italy. The increase in searches for *chloroquine* was observed during this time only in Asian and European countries (Japan, South Korea, Singapore, Italy, France, Spain, and the United Kingdom); it was not found in the United States, India, Brazil, or South Africa. Public searches on both drugs increased after mid-March and quickly peaked on March 20 after US President Donald Trump's initial public remarks on the drugs at a briefing on March 19; this increase was observed in all countries.

Figure 5. Google Trends data on rumors related to antimalarial drugs worldwide from December 2019 to April 2020.



Discussion

Principal Findings

Using internet surveillance data, we conducted a cross-country comparison of real-time public awareness and behavioral responses to epidemic information during the first 100 days of the COVID-19 pandemic. We identified squandered windows of opportunity for early epidemic control in 12 countries. The epidemic in China and the PHEIC did not prompt a worldwide public reaction to adopt public health protective measures; instead, most countries and regions only responded to the epidemic after their own case counts mounted. Even if there was a worldwide reaction, the public response to the epidemic in other countries would quickly fall without government leadership and communication. The public responded quickly to official announcements and adopted personal protection behaviors such as buying hand sanitizers; however, rumors and misinformation were found to have led to a surge in sales of herbal remedies in China and antimalarial drugs worldwide. Chinese data showed that the timely governmental clarification of rumors mitigated the rush to buy unproven remedies to treat or prevent COVID-19. This comparative study highlighted the importance of governmental leadership and international coordination in directing epidemic control.

The lack of transparent, timely, and effective risk communication by health authorities around an emerging infectious disease in its early stages failed to bring about appropriate levels of public awareness and behavioral responses, such as avoidance of mass gatherings and personal protection in China, Europe, and the United States. In China, the government did not provide actionable advice for personal protection until January 21. There were two missed windows of opportunity for early epidemic control in China. First, the first COVID-19 case emerged on December 8, 2019, more than three weeks before December 31, when the outbreak was first announced; second, between December 31, 2019, and January 19, 2020, the Wuhan Health Commission made four public announcements with no obvious evidence of human-to-human transmission [1-3]. This series of government announcements kept public awareness low, which prevented the public from realizing the risk of the disease and from taking personal protection measures. This period overlapped with the celebration of the arrival of 2020 and the preparation for the Chinese New Year (January 25), when public attention centered around family reunions. Similarly, our data show an additional window of 4 to 8 weeks of missed opportunity for early epidemic control in Europe and the United States between the PHEIC declaration in late January and the outbreaks in Europe and the United States in March. Transmission of COVID-19 appears to be possible even among people who show no symptoms of the disease; thus, travel restrictions have limited impact on stemming the spread. However, other than flight cancellations and evacuation from China in late January, governments in Europe and the United States did not implement epidemic prevention or control measures such as testing, surveillance, or contact/case tracing [23,24]. Despite reports of local transmissions, population movement between Italy and the rest of the world and public

gatherings in these countries were largely unchecked even after the number of cases surged in Italy [25]. Additionally, some political leaders and media outlets repeatedly called the COVID-19 pandemic a hoax [26] and downplayed the threat the pandemic posed to society, leading to a delay in public response.

Further, due to variations in societal and cultural paradigms, face masks are commonly used as a hygienic practice in many Asian countries but are only used by the unwell in European and North American countries [11]. In these countries, authorities discouraged citizens from using face masks, and local communities reacted with stigmatization and racial aggravations against members of East Asian communities who wore masks [12]. Emerging evidence about the efficacy of wearing face masks ("community transmission might be reduced if everyone, including people who have been infected but are asymptomatic and contagious, wear face masks" [11]) finally shifted public opinion. In April 2020, governments in Europe and the United States changed their mask-wearing guidelines and mandated universal face mask use outdoors until an effective vaccine becomes available. Our data show that the use of masks increased substantially after local COVID-19 epidemics began. The sudden shift of government guidelines led to a worldwide surge in the demand for face masks, which have been in severely short supply since the beginning of the epidemic, even for frontline health care professionals. This phenomenon highlights the importance of governmental response in early epidemic preparedness for mobilization and surge capacity, effective control measures, and timely and clear communication to cue public action. Future research is needed to evaluate the impact of mask-wearing policies on awareness and behaviors.

Evidence has proven that early implementation of containment measures can effectively control the COVID-19 epidemic [27]. However, the epidemic in China only caught the public attention of East and Southeast Asia, and the lessons learned in this region did not trigger appropriate epidemic responses in the rest of the world. Similarly, the epidemic in Italy only resulted in public awareness and reaction in Europe and the United States. Due to the dense populations and fragile health systems in South Asian, Latin American, and African regions, COVID-19 spread is of great concern in these countries [28]. This study indicates a need for strengthened international partnerships and coordination to combat the COVID-19 epidemic and future epidemics. The WHO should be empowered to take a leading role in guiding more preparedness actions than solely making statements that a disease constitutes a PHEIC or pandemic.

Our data showed that the public is highly responsive to governmental risk communication during epidemics, suggesting a need for real-time media surveillance. The early release of official guidelines by China's National Health Commission was effective in guiding the Chinese public to take personal protection measures; however, rumors also triggered several incidents of panic buying (eg, masks, Shuanghuanglian, and garlic), which were quickly calmed after official clarifications. Likewise, references to antimalarial drugs by US President Donald Trump triggered panic searches worldwide. Governments should engage early in public risk communication about epidemic control, detect misinformation, confront inaccurate messages, and clarify rumors in real time [29,30]. Working with the private sector to ensure sufficient stock and reasonable pricing for recommended personal protection products, such as hand sanitizer and masks, is critical for effective epidemic prevention and control and to avoid public panic.

Compared with traditional surveys, internet surveillance tools provide real-time, longitudinal, and dynamic data for capturing public awareness, rumors, and behavioral reactions, and they can be an effective means to evaluate public response towards epidemic information and rumors [19].

Limitations

This study has several limitations. First, Google Trends are not applicable in all countries, including China and most countries in Africa; therefore, we could only compare countries with available Google Trends data. For countries in which the volumes of Google Shopping data were low (Japan, Singapore, South Korea, Italy, and Spain), we used Google web search data instead. This may have led to inconsistency in the measurements. Second, there is an inherent bias in internet data because the population may be skewed towards younger people. Moreover, Google Trends data related to personal protection measures represent intended behaviors instead of real shopping behaviors. Third, we are unable to conduct segmentation analysis these data to inform audience-tailored communication strategies.

Conclusion

Our comparative study identified missed windows of opportunity for early epidemic control in 12 countries; it also highlighted the urgent need for international coordination to promote mutual learning regarding epidemic characteristics as well as effective control measures and to trigger early and timely responses in individual countries. During an epidemic, early release of official guidelines and timely governmental clarification of rumors are necessary to guide the public to make rational responses.

Acknowledgments

ZH acknowledges financial support from the National Natural Science Foundation of China (No. 71874034). The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Authors' Contributions

ZH and LL conceptualized the study design. XZ, FD, and HJ collected and analyzed the data. ZH, LL, and FD interpreted the results and wrote the manuscript. LL, HL, and SM revised the manuscript. All authors approved the final draft of the manuscript. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflicts of Interest

HL is on the Merck Vaccine Confidence Advisory Board. Her research group, the Vaccine Confidence Project, received research grants from GSK and Merck on vaccine confidence issues. None of those research grants are related to this paper.

Multimedia Appendix 1
Correlations between Google Trends and Google Shopping indices for behavioral responses.
[DOCX File , 19 KB - jmir_v22i8e21143_app1.docx]

Multimedia Appendix 2
Correlations between Baidu and Ali indices for behavioral responses in China.
[DOCX File , 13 KB - jmir_v22i8e21143_app2.docx]

Multimedia Appendix 3
Correlations between Google Trends data on the search terms "mask" and "hand sanitizer" and the numbers of new COVID-19 cases.
[DOCX File , 14 KB - jmir_v22i8e21143_app3.docx]

Multimedia Appendix 4
Lag correlations between Google Trends dada on the search term "mask" and numbers of new COVID-19 cases in Italy and Spain.
[PNG File , 36 KB - jmir_v22i8e21143_app4.png]

References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
2. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb;395(10223):507-513. [doi: 10.1016/s0140-6736(20)30211-7]
3. National Health Commission of the People's Republic of China. Webpage in Chinese. URL: http://www.nhc.gov.cn [accessed 2020-03-02]
4. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-03-11]
5. Coronavirus disease 2019 (COVID-19) situation report-100. World Health Organization. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200429-sitrep-100-covid-19.pdf?sfvrsn=bbfbf3d1_2 [accessed 2020-04-29]
6. Lin L, Savoia E, Agboola F, Viswanath K. What have we learned about communication inequalities during the H1N1 pandemic: a systematic review of the literature. BMC Public Health 2014 May 21;14(1):484 [FREE Full text] [doi: 10.1186/1471-2458-14-484] [Medline: 24884634]
7. Lin L, Jung M, McCloud RF, Viswanath K. Media Use and Communication Inequalities in a Public Health Emergency: A Case Study of 2009–2010 Pandemic Influenza a Virus Subtype H1N1. Public Health Rep 2014 Nov;129(6_suppl4):49-60. [doi: 10.1177/00333549141296s408]
8. Hsu Y, Chen Y, Wei H, Yang Y, Chen Y. Risk and Outbreak Communication: Lessons from Taiwan's Experiences in the Post-SARS Era. Health Secur 2017 Apr;15(2):165-169 [FREE Full text] [doi: 10.1089/hs.2016.0111] [Medline: 28418746]
9. The Lancet. COVID-19: fighting panic with information. Lancet 2020 Feb;395(10224):537. [doi: 10.1016/s0140-6736(20)30379-2]
10. Larson HJ. Blocking information on COVID-19 can fuel the spread of misinformation. Nature 2020 Apr;580(7803):306. [doi: 10.1038/d41586-020-00920-w] [Medline: 32231320]
11. Feng S, Shen C, Xia N, Song W, Fan M, Cowling BJ. Rational use of face masks in the COVID-19 pandemic. Lancet Respir Med 2020 May;8(5):434-436. [doi: 10.1016/s2213-2600(20)30134-x]
12. Cheng KK, Lam TH, Leung CC. Wearing face masks in the community during the COVID-19 pandemic: altruism and solidarity. Lancet 2020 Apr:online ahead of print. [doi: 10.1016/s0140-6736(20)30918-1]

13. Betsch C, Wieler LH, Habersaat K. Monitoring behavioural insights related to COVID-19. Lancet 2020 Apr;395(10232):1255-1256. [doi: 10.1016/s0140-6736(20)30729-7]
14. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
15. Majumder MS, Santillana M, Mekaru SR, McGinnis DP, Khan K, Brownstein JS. Utilizing Nontraditional Data Sources for Near Real-Time Estimation of Transmission Dynamics During the 2015-2016 Colombian Zika Virus Disease Outbreak. JMIR Public Health Surveill 2016 Jun 01;2(1):e30 [FREE Full text] [doi: 10.2196/publichealth.5814] [Medline: 27251981]
16. Milinovich GJ, Williams GM, Clements ACA, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis 2014 Feb;14(2):160-168. [doi: 10.1016/s1473-3099(13)70244-5]
17. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 2020 Mar 27;25(10):30522 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
18. Lin Y, Liu C, Chiu Y. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun 2020 Jul;87:30-32 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.020] [Medline: 32283286]
19. Husnayain A, Fuad A, Su EC. Applications of Google Search Trends for risk communication in infectious disease management: A case study of the COVID-19 outbreak in Taiwan. Int J Infect Dis 2020 Jun;95:221-223 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.021] [Medline: 32173572]
20. Arora VS, McKee M, Stuckler D. Google Trends: Opportunities and limitations in health and health policy research. Health Policy 2019 Mar;123(3):338-341. [doi: 10.1016/j.healthpol.2019.01.001] [Medline: 30660346]
21. Robertson A. Trump's chloroquine hype is a misinformation problem bigger than social media. The Verge. 2020 Apr 09. URL: https://www.theverge.com/2020/4/9/21209797/trump-chloroquine-hydroxychloroquine-medication-social-media-misinformation [accessed 2020-04-09]
22. Jaffe S. Regulators split on antimalarials for COVID-19. Lancet 2020 Apr;395(10231):1179. [doi: 10.1016/s0140-6736(20)30817-5]
23. Anna PA, Hale T, Phillips T, Webster S. Variation in government responses to COVID-19. Blavatnik School of Government, University of Oxford. 2020 May 28. URL: https://www.bsg.ox.ac.uk/research/publications/variation-government-responses-covid-19 [accessed 2020-07-24]
24. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet 2020 Mar;395(10228):931-934. [doi: 10.1016/s0140-6736(20)30567-5]
25. Legido-Quigley H, Asgari N, Teo YY, Leung GM, Oshitani H, Fukuda K, et al. Are high-performing health systems resilient against the COVID-19 epidemic? Lancet 2020 Mar;395(10227):848-850. [doi: 10.1016/s0140-6736(20)30551-1]
26. Sullivan B. Fox News Faces Lawsuit For Calling COVID-19 A 'Hoax'. Forbes. 2020 Apr 10. URL: https://www.forbes.com/sites/legalentertainment/2020/04/10/covid-19-lawsuit-against-fox-news/ [accessed 2020-04-10]
27. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. JAMA 2020 Apr 10:1915-1923 [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]
28. Hopman J, Allegranzi B, Mehtar S. Managing COVID-19 in Low- and Middle-Income Countries. JAMA 2020 Mar 16;323(16):1549. [doi: 10.1001/jama.2020.4169] [Medline: 32176764]
29. Risk Communication and Community Engagement (RCCE) Action Plan Guidance COVID-19 Preparedness and Response. World Health Organization. 2020 Mar 16. URL: https://www.who.int/publications/i/item/risk-communication-and-community-engagement-(rcce)-action-plan-guidance [accessed 2020-07-24]
30. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 May 18;27(3):taaa031 [FREE Full text] [doi: 10.1093/jtm/taaa031] [Medline: 32125413]

Abbreviations

COVID-19: coronavirus disease
PHEIC: public health emergency of international concern
WHO: World Health Organization

Edited by G Eysenbach; submitted 06.06.20; peer-reviewed by S Song, J Brož; comments to author 13.07.20; revised version received 20.07.20; accepted 23.07.20; published 03.08.20.

Please cite as:
Hou Z, Du F, Zhou X, Jiang H, Martin S, Larson H, Lin L
Cross-Country Comparison of Public Awareness, Rumors, and Behavioral Responses to the COVID-19 Epidemic: Infodemiology Study
J Med Internet Res 2020;22(8):e21143
URL: https://www.jmir.org/2020/8/e21143
doi:10.2196/21143
PMID:32701460

Original Paper

The Anatomy of the SARS-CoV-2 Biomedical Literature: Introducing the CovidX Network Algorithm for Drug Repurposing Recommendation

Lyndsey Elaine Gates[1*], MSN, DNP; Ahmed Abdeen Hamed[2*], MS, PhD

[1]School of Nursing, Norwich University, Northfield, VT, United States

[2]School of Cybersecurity, Data Science, and Computing, Norwich University, Northfield, VT, United States

[*]all authors contributed equally

Corresponding Author:
Ahmed Abdeen Hamed, MS, PhD
School of Cybersecurity, Data Science, and Computing
Norwich University
158 Harmon Drive
Northfield, VT, 05663
United States
Phone: 1 802 485 2276
Email: ahamed@norwich.edu

Abstract

Background: Driven by the COVID-19 pandemic and the dire need to discover an antiviral drug, we explored the landscape of the SARS-CoV-2 biomedical publications to identify potential treatments.

Objective: The aims of this study are to identify off-label drugs that may have benefits for the coronavirus disease pandemic, present a novel ranking algorithm called CovidX to recommend existing drugs for potential repurposing, and validate the literature-based outcome with drug knowledge available in clinical trials.

Methods: To achieve such objectives, we applied natural language processing techniques to identify drugs and linked entities (eg, disease, gene, protein, chemical compounds). When such entities are linked, they form a map that can be further explored using network science tools. The CovidX algorithm was based upon a notion that we called "diversity." A diversity score for a given drug was calculated by measuring how "diverse" a drug is calculated using various biological entities (regardless of the cardinality of actual instances in each category). The algorithm validates the ranking and awards those drugs that are currently being investigated in open clinical trials. The rationale behind the open clinical trial is to provide a validating mechanism of the PubMed results. This ensures providing up to date evidence of the fast development of this disease.

Results: From the analyzed biomedical literature, the algorithm identified 30 possible drug candidates for repurposing, ranked them accordingly, and validated the ranking outcomes against evidence from clinical trials. The top 10 candidates according to our algorithm are hydroxychloroquine, azithromycin, chloroquine, ritonavir, losartan, remdesivir, favipiravir, methylprednisolone, rapamycin, and tilorone dihydrochloride.

Conclusions: The ranking shows both consistency and promise in identifying drugs that can be repurposed. We believe, however, the full treatment to be a multifaceted, adjuvant approach where multiple drugs may need to be taken at the same time.

(J Med Internet Res 2020;22(8):e21169) doi:10.2196/21169

KEYWORDS

health; informatics; COVID-19 treatment; drug repurposing; network algorithm; ranking; drug; biomedical; antiviral; COVID-19

Introduction

Background

In December 2019, the first known cases of the coronavirus disease (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), were identified in humans in Wuhan, China. Transmission was thought to have occurred through contact with animals at a wet market, although the exact animal-to-human transmission incidence is still unknown at the time of this publication [1]. Since then, SARS-CoV-2 infections have been reported worldwide, with

the greatest numbers of infections in the United States, Russia, and Europe [2]. With no immunity among humans, at the time of this writing, SARS-CoV-2 has infected over 4.5 million individuals and killed over 300,000 individuals worldwide with the potential for many more fatalities without a known cure or vaccination [3].

With high mortality rates thus far resulting from the SARS-CoV-2 infection, true morbidity rates are still unknown. Studies have shown that cardiac injury occurred in up to 28% of patients hospitalized with a SARS-CoV-2 infection [4], which can increase the risk of death later on. The systemic inflammatory response that occurs in severe SARS-CoV-2 infection may result in hypoxemia and increased cardiac demand on an already taxed cardiovascular system. Previous research has shown that viral respiratory infections may increase the risk of myocardial infarctions in patients, prompting some researchers to focus on infection prevention initiatives as strategies for lowering risk of cardiovascular disease [4].

Due to the uncertainty of the processes involved in SARS-CoV-2 infection and treatment, novel approaches are needed. One approach is drug repurposing, where drugs that are approved for other diseases are investigated to determine whether they are safe and effective for different conditions. Drug repurposing efforts have resulted in a great number of treatment options for disorders such as Sildenafil treatment for erectile dysfunction (previously intended for hypertension and angina), Itraconazole for lung and prostate cancer treatment (previously used only for antifungal purposes), and Saracatinib therapy for investigation into Alzheimer disease reversal (previously thought to be a failure drug for anticancer purposes) [5].

The purpose of this study is to identify drugs in circulation that could be repurposed to offer potential benefits of use in treating patients infected with SARS-CoV-2.

Computational Methods for Drug Repurposing

The idea of investigating drugs that are already designed for a given indication to serve a new purpose is not new to science [6]. This idea has become recently more prominent due to the human advancement in computational sciences and the possibilities of processing large data sets on a high-performance computing unit [7]. Drug repurposing can provide effective therapies, such as in the case of invasive fungal infections [8]. Computational repurposing offers the following two approaches: (1) target-based repurposing, where a drug is interacting with a gene or a protein, and (2) disease-based, where a drug is found connected to new indications [9].

Over the past decades, various approaches have been proposed using machine learning, network science, and text mining [10]. Cheng et al [11] introduced a bipartite network-based approach for a drug-target inference method that was proven useful for detecting drug-target interactions in the molecular polypharmacological domain. Wu et al [12] presented a heterogeneous network clustering approach for associating drugs with disease. The nodes of the network presented nodes as drugs and diseases, while edges represented a shared gene or biological process, etc [12].

Text mining played a significant role in advancing biomedical informatics in general [13] and computational drug research in particular [14,15]. Wu et al [16] developed computational models to retrieve drug-drug interaction and drug-gene interaction evidence from PubMed abstracts.

Andronis et al [17] presented a review on the application of literature mining using semantic ontology and controlled vocabularies as tools to explore the drug repurposing research. Loging et al [18] presented literature databases such as Medline as knowledge-driven systems to identify novel uses for drugs that span the therapeutic pipeline. Deftereos et al [19] demonstrated the importance of both text mining and networks to discover new drug uses and adverse drug reactions. The purpose of text mining is used to extract the names of the genes, disease, and biological processes. Networks presented the associations among the names to allow the discovery of the molecular mechanisms underlying disease and links to existing drugs that may be used [19]. Nabirotchkin et al [20] presented a combined networks and text mining approach to advance the multi-keyword search for drug repurposing (eg, two drugs and one target).

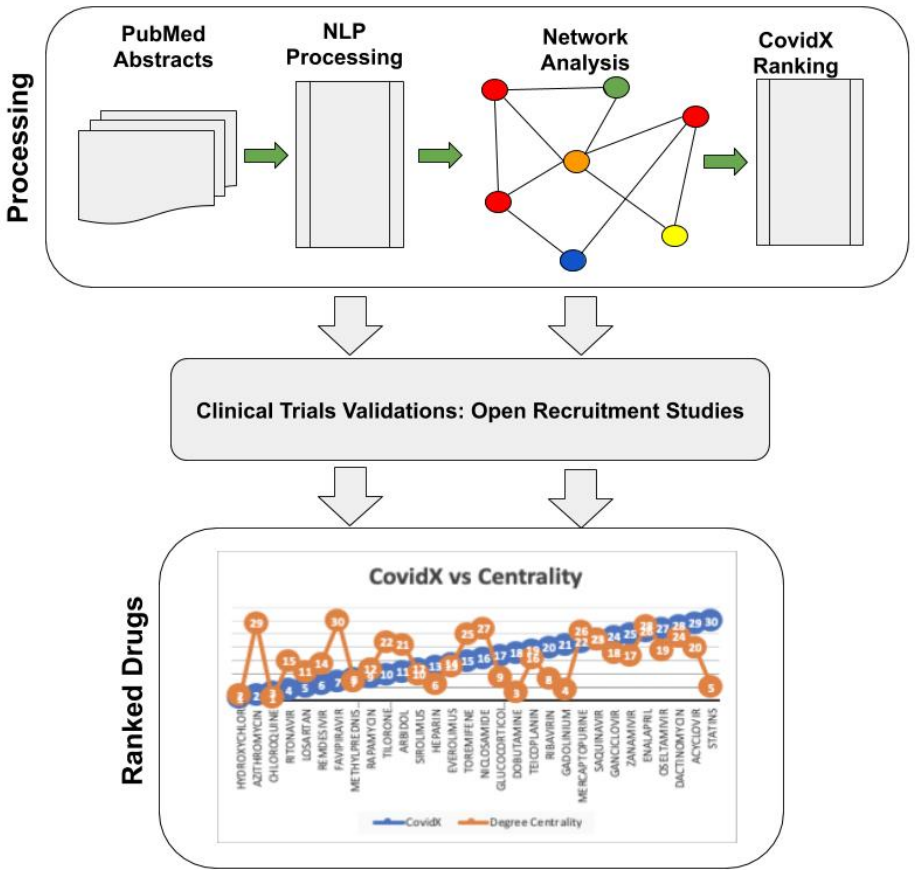
As search capabilities advanced, ranking algorithms also emerged. Luo et al [21] presented a ranking algorithm to produce a drug repurposing recommendation system. Karatszas et al [22] presented a web-based tool, which they called composite drug reranking scoring, to identify the most promising drugs and chemical substances to test. Hamed et al [23] combined both literature text mining and networks to rank existing drugs based on the biological specificity of a drug to explore their similarity and whether they can be repurposed. Similarly, we present a new ranking algorithm that is based on a new notion, which they called *variability*. The methods section describes the details of each component of this research.

Methods

Overview

The method that we used in this exploratory study is described as follows: named entity recognition [24], drug-entities network construction, and drug scoring using an algorithm that we titled CovidX. Before the algorithm produced the final ranking, it factored in whether a drug was under current investigation in clinical trials. The algorithm awarded those drugs by a given weight based on the frequency of studies that were open for recruitment. Each step is presented along with the data set used in the analysis onward. Figure 1 shows the data workflow and the various steps of data processing and making drug repurposing recommendations.

Figure 1. Workflow diagram showing how the processing of PubMed abstracts using NLP produces a network of drugs (symbolized in red) and various biological features (in various colors). The workflow shows the clinical trials validation steps before it produces the final ranking results of the CovidX algorithm. NLP: natural language processing.



Data Sets

Starting from the PubMed web portal, we searched for "SARS-CoV-2" keywords. This search process produced 941 publications captured in the abstracts and metadata. Since the focus of this work is a computational study about mining literature, the focus of the data was the text embedded in the abstracts. The study kept track of each article ID, known as PMID, and the abstract text. The abstracts were the data grounds for text analysis, and the PMID ensured that the extracted entities of the same article were linked together. Another complementary data set that we extracted from the clinical trials web portal included those studies related to COVID-19 and that were also recruiting participants. This data set was comprised of 433 records, which were analyzed for their title, conditions, and intervention sections for COVID-19, and the coexisting drugs in the PubMed data set.

Instruments and Experiment's Environment

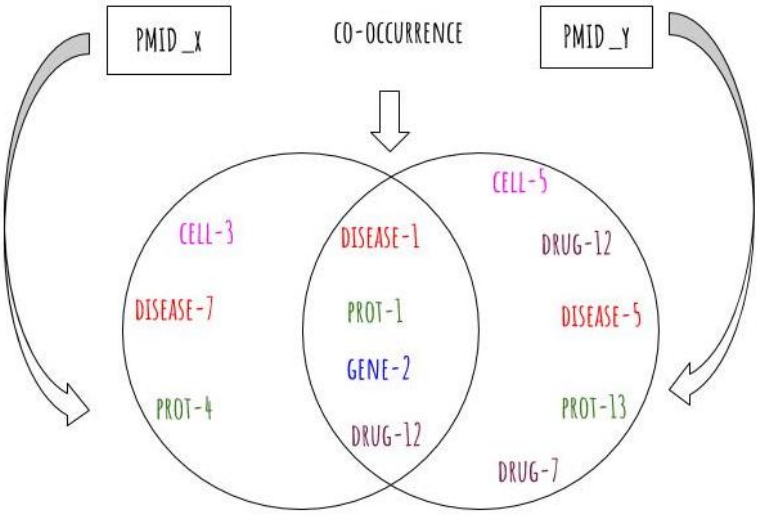
The data preparing, analysis, experiment, and implementation used here were done entirely using Python Jupyter Notebook [25] to facilitate the reproducibility of this study. We used specific packages to do the work: TextBlob [26] for text processing, NetworkX [27] for network analysis and algorithm implementation, and the Python version iGraph [28] for network visualization.

Entity Recognition and Network Construction

In the first step, we applied known natural language techniques [29] to identify and extract the entities mentioned in the biomedical abstracts. Named entity recognition [24] provides the means of identifying and extracting entities such as drug names, genes, proteins, cell lines, organism, and chemical compounds. Some of these tools are ontology-driven and identify entities if they exist in biological ontology [30]. Biological ontology, such as the Gene Ontology [31-33], offer great tools since they have been manually vetted by domain experts. They also play a significant role in the identification of the biological entities of this paper. All identified names are qualified by their ontology (GO: membrane, CHEBI: chloroquine) or feature type (disease: obstructive coronary disease).

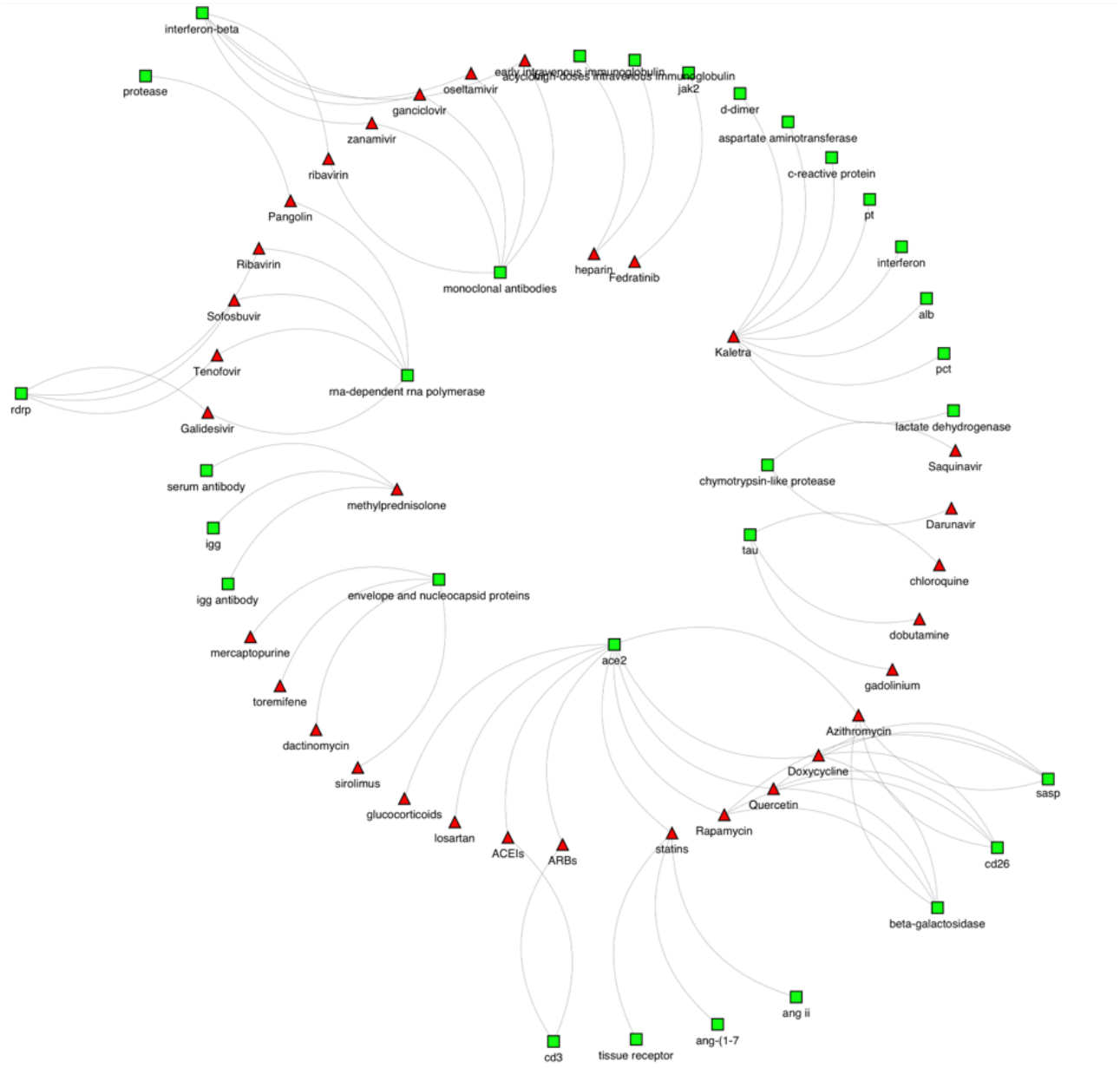
In the second step, we constructed a (drug-entity) map using the biological entities that have been identified in the first step. The map is constructed using a heuristic: "if a set of targets belongs to a given abstract, and co-occurring with a drug, then they must be related. How they are related is demonstrated in the scope of this work." This offers a connectivity mechanism between the drugs and the various targets. As new drugs are extracted and connected with their corresponding targets, the map grows naturally. Figure 2 demonstrates the extraction of various feature types and links with its original PMID.

Figure 2. Two hypothetical PubMed Abstracts (PMIDx and PMIDy). Each abstract contributes some biological entities. The entities in the intersection namely (disease-1, prot-1, gene-2, and drug-12) contribute to a network of co-occurring entities that become the grounds for further exploration. In the presence of a larger data set of abstracts and more biological features, the network becomes an interesting ground for exploration and potential discoveries.



This provides us with an interesting landscape in terms of structure and connectivity for exploration and discovery. The network naturally grows when a drug, for example, in one abstract is mentioned in another abstract. This is the case when a drug target is addressed in various publications by the same or different authors. The more the abstracts have features in common, the more the network grows in links and structure. Here we represent the drug-entity map as a graph, G=(V; E), where (V) represents the set of drugs on one hand and the features on the other hand as nodes, and (E) denotes the connections from drugs to the entities co-occurring in the data set. Figure 3 presents drug-protein as a subset network of the entire network of drug-feature extracted in this process.

Figure 3. The connections between drugs and protein targets. The drugs nodes are represented as red triangles and the protein nodes are represented by green squares. Here we note that various drugs have targeted one specific protein (ACE2), while other drugs have been found linked to more than one protein as in the case with methylprednisolone. ACEIs: angiotensin-converting enzyme inhibitors; ace2: angiotensin-converting enzyme 2; ang: angiotensin; ANG II: angiotensin II; ARBs: angiotensin II receptor blockers; igg: immunoglobulin G; jak2: Janus kinase 2; RdRp: RNA-dependent RNA polymerase.



CovidX Ranking Algorithm

The CovidX ranking algorithm we present here is based on a simple yet powerful heuristics: the more knowledge known about the drug, the higher the likelihood that it can be repurposed. The following criteria were used to determine whether the knowledge could be considered relevant:

- The diversity of links to the named entities identified using the natural language processing framework used earlier (eg, gene, protein, cell line, chemical, disease)
- Whether the drug is currently being investigated in an ongoing clinical trial [34] specialized in COVID-19

The notion of diversity is quantified by the degree of drug to each unique feature type. Diversity is captured by a binary matrix where a value of 1 represents a connection and 0 represents no connection to the category. For instance, if a drug is found linked to a given protein instance, this registers a value of 1. On the other hand, if a drug is not linked to a gene, this places a value of a 0 in the corresponding column. In the case where a drug is found linked to various instances of the same category (eg, a drug is linked to three different proteins), this still contributes a value of 1 not 3.

The scoring mechanism is based on two contributing factors: (1) the log of the diversity score for each drug mentioned in the PubMed data set and (2) the normalized frequency of each drug in the clinical trial data set [34]. The justification for this heuristic is that a drug can be promising but has not been investigated in a clinical trial yet. On the other hand, it is not

the case that every clinical trial results in a successful outcome. Some trials may fail in late phases.

The given weight of these two factors addresses both the issues of promise or failure. The following are the algorithmic steps necessary to produce a ranking score:

1. Generate an incident matrix for all drugs in the network
2. Convert the incident matrix into a diversity matrix, one that is based on the node category (eg, drug), not the actual label (eg, chloroquine)
3. For each category a drug is linked with, place a value of 1; otherwise, place a value of 0
4. Calculate the column variability score: log of the value for total number of categories divided by the sum of the 1 values
5. If a drug is active in a clinical trial, add the drug name and its frequency score to the list of potential candidates
6. Rank drugs in an ascending order based on the variability score, the lower the value the higher the rank

Algorithm Formalism

Suppose N denotes a matrix of drugs and their corresponding neighbors. The first column G is a set of drugs instances {g1, g2, g3, g4}, while the remaining number of columns denote neighboring nodes of each drug in C. Each instance of G has a set of biological neighbors of type (genes, proteins, cell type, cell line). This can be mathematically represented with a matrix where the first column is the drug instances. Each row is linked to the actual instances of feature type. For simplification, equation 1 purposes only instances of four different feature types {w, x, y, z}.

Here we note that each feature in the A matrix corresponds to a feature type with an index. This alphabetical symbol represents the type, and the index represents a unique instance of that feature. For example, drug g1 has two instances of type w,

namely, w1 and w2. Similarly, drug g4 has two instances of type z, namely, z3 and z4. This A matrix can be transformed into a new binary matrix, which we call the “diversity” matrix denoted as D. Table 1 shows a concrete representation of each drug and the expression of the diversity notion. Due to the page size limitation, we included only the most significant five features, namely, cell line, cell type, disease, DNA, and protein.

This matrix is described with the following set of columns: {g1, g2, g3, g4}. Here the rows are reserved for the feature types (ie, {w, x, y, z}), which we introduced earlier. For each drug linked to a specific instance (of a feature type), a value of 1 is placed in the corresponding row; otherwise, a value of 0 is recorded. Equation 2 shows the newly transformed matrix D.

To summarize the D matrix and produce a rank for each drug, we calculated its transpose R according to the following equation $R=D^T$. Each rank is derived from the sum of each g_i column. Equation 3 shows the new matrix R and the sum of each column. For better readability, we also list the column and row labels.

Using the column sums calculated, we ranked the drugs according to equation 4. It is to be noted that s is the column sum and n is the total number of features. This formula is inspired by the inverse term frequency56. The second part of the equation takes into consideration the frequency weight of each drug studied in clinical trials by dividing the frequencies q by the total number of clinical trials m. For those drugs that were not studied in clinical trials, we assigned a random weight in the range of zero and the lowest frequency of drugs.

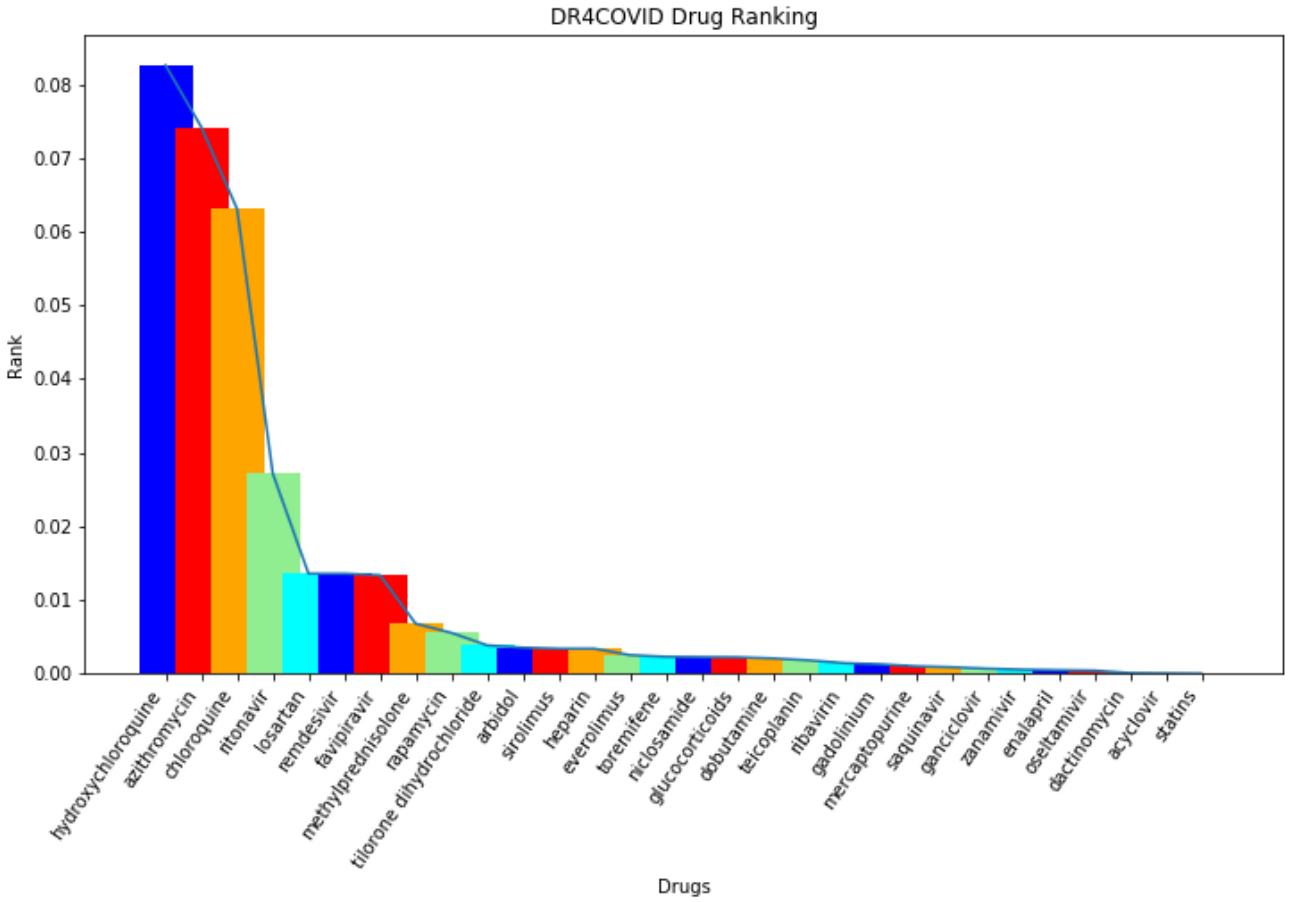
Figure 4 shows the final ranking results produced by our algorithm. It is striking to see that hydroxychloroquine is ranked at the top on the list despite the various news article about the side effects. We believe this is because it is the most investigated in clinical trial.

Table 1. Summary of the binary matrix of drugs-entity resulted from the natural language procressing.[a]

Drug	Cell line	Cell type	Disease	DNA	Protein
Hydroxychloroquine	1	1	1	0	1
Chloroquine	1	0	1	1	1
Sirolimus	1	0	1	0	1
Rapamycin	0	0	1	0	1
Everolimus	0	0	1	0	1
Methylprednisolone	0	0	1	0	1
Statins	0	0	1	0	1
Ribavirin	0	1	1	1	1
Dobutamine	0	0	1	1	1
Gadolinium	0	0	1	1	1
Tilorone dihydrochloride	0	0	1	0	0
Enalapril	0	0	1	0	1
Losartan	0	0	1	1	1
Glucocorticoids	0	0	1	1	1
Saquinavir	0	0	1	0	1
Heparin	1	0	1	1	1
Azithromycin	0	0	1	1	0
Dactinomycin	1	0	1	0	0
Toremifene	1	0	1	0	0
Mercaptopurine	1	0	1	0	0
Remdesivir	0	0	1	1	1
Ritonavir	0	1	1	1	0
Teicoplanin	0	1	1	1	1
Arbidol	0	0	1	0	1
Zanamivir	0	0	1	0	1
Ganciclovir	0	0	1	0	1
Oseltamivir	0	0	1	0	1
Acyclovir	0	0	1	0	1
Favipiravir	0	0	1	0	1
Niclosamide	0	0	1	0	1

[a]The table is a representation of the diversity matrix. When a drug is found connected to a feature type, it scores a value of 1 regardless of how many instances of the feature. If no instances are found, then a value of 0 is registered in the corresponding type of the given feature.

Figure 4. The drugs that are found in the literature and the derived ranking using the CovidX algorithm. The algorithm takes into consideration the diversity notion introduced earlier and the frequency of a drug that is currently being studied in a clinical trial. The combination of these two factors has produced a ranking mechanism to enable the selection of candidate drugs for repurposing. The diagram shows the top-10 listed according to the significance.



Results

Pathophysiology of SARS-CoV-2

It is speculated that SARS-CoV-2 enters the body by binding to the angiotensin-converting enzyme 2 (ACE2) coreceptors in a host cell, which are found in tissues of the lung, heart, kidney, brain, and gastrointestinal system. The actual ACE2 enzyme is involved with the renin-angiotensin-aldosterone system (RAAS), where it breaks down angiotensin II (ANG II) and generates angiotensin 1-7, thereby decreasing blood pressure in the body [35]. There is also speculation that SARS-CoV-2 enters the body through the cluster of differences (CD)26 receptor, as this is the host receptor for Middle East respiratory syndrome–related coronavirus (MERS-CoV) [36].

Current research is focused on determining the exact pathogenesis of SARS-CoV-2. One study from China showed that in patients hospitalized with a SARS-CoV-2 infection, common disease complications included acute respiratory distress syndrome (ARDS), acute cardiac injury, and secondary infection [37]. Clinical features of the disease thus far have shown to be an initial increase in the secretion of interleukin (IL)–4 and IL-10, which are T helper (Th)–2 cytokines and suppress inflammation ([37] as cited in [38]), and a Th-1 cell hyper-response that is thought to lead to the ARDS associated with severe acute respiratory syndrome (SARS; [39] as cited in [38]).

In less symptomatic patients, symptoms may involve gastrointestinal dysfunction or no symptoms at all. In those who are more highly symptomatic, early symptoms include fever, cough, nasal congestion, and fatigue. These symptoms have been reported to progress to dyspnea and pneumonia, which can be confirmed with computed tomography (CT) imaging. The sequelae of SARS-CoV-2 pneumonia includes decreased oxygen saturation, changes in blood gas composition, ground glass opacities on CT imaging, and other alveolar abnormalities. Laboratory markers include lymphopenia, elevated C-reactive protein, and elevated levels of proinflammatory cytokines [40]. Some research has shown significant lower T cell and B cell levels; natural killer levels; and CD3, CD4, and CD8 lymphocyte levels [41]. Liver dysfunction has been reported with abnormal levels of alanine aminotransferase and aspartate aminotransferase [42]. Procalcitonin levels have shown to be normal in patients admitted to the hospital for SARS-CoV-2 infection and then increased in those admitted to intensive care units [3]. Higher plasma levels of lactate dehydrogenase have also been reported [43].

Additionally, some patients with a SARS-CoV-2 infection have progressed to even more severe respiratory distress, sepsis, and septic shock [44]. Related laboratory markers have included an elevated D-dimer and elevated fibrinogen levels [45]. Some

cases have shown complications related to coagulopathy and disseminated intravascular coagulation (DIC) [46].

Nucleocapsids are structures that consist of a protein that encloses a piece of RNA. With a virus, the nucleocapsid can be targeted by drugs that can open this type of envelope for access and eventual modification or destruction.

Cardiovascular Drugs

Due to the viral binding of ACE2 receptors, concerns have been raised regarding the use of antihypertensive and cholesterol-lowering agents in patients. For example, the use of ACE inhibitors and ANG II receptor blockers, in addition to statin medications, have been shown to enhance the effects of ACE2 and, therefore, the risk of a SARS-CoV-2 infection. On the other hand, discontinuing RAAS blockade has the potential to increase the cardiovascular and respiratory complications that result from SARS-CoV-2 infection related to increased blood pressure, inflammation, and fibrosis [35].

Anticoagulant Drugs

Patients on anticoagulant therapy and particularly those on vitamin K antagonists (VKA) have shown worse clinical outcomes when hospitalized with SARS-CoV-2 infection related to the variability in metabolic processes during the diseased state. Direct oral anticoagulant therapy is often discouraged due to potential drug interactions with proposed antiviral and other therapy. As a result, it has been suggested for patients to be switched to low-molecular weight heparin or unfractionated heparin to avoid complications of variable VKA therapy. In those patients for whom VKA therapy is required, there should be close monitoring of prothrombin time and international normalized ratio [47].

Additionally, for prophylactic treatment of coagulopathy and DIC, and for the prevention of clots related to long-term bed rest of patients being treated for a SARS-CoV-2 infection, heparin therapy has been proposed; however, the effective dosing regimen has yet to be determined [46].

Drugs With Antiviral Activity

Outbreaks of coronavirus, prior to that of SARS-CoV-2, have largely been treated with chemical agents that target enzymes involved with DNA replication and synthesis (helicase and polymerase) and the protein catalysis needed for functional viruses (protease), as well as agents that modify the immune response, such as interferon and corticosteroids ([48], as cited in [49]). In particular, drugs that target chymotrypsin-like protease have been attractive due to its significance in the viral replication process [50].

SARS-CoV-2 is 97% genetically identical to SARS infections seen in the past. The RNA-dependent RNA polymerase (RdRp) involved with SARS-CoV-2 has become a target for several drugs used to treat hepatitis, HIV, and the Ebola virus. One modeling study suggested that drugs such as ribavirin, remdesivir, sofosbuvir, galidesivir, and tenovir, which tightly bind to RdRp, could be effective in treating SARS-CoV-2 infections [51].

Recent studies have shown agents that typically interfere with protease, such as lopinavir, ritonavir, and saquinavir, may interact with the main protease involved in SARS-CoV-2 [49]. Darunavir is a more potent protease inhibitor and currently used for HIV infection in conjunction with ritonavir, an antiretroviral agent [52].

Arbidol, a broad-spectrum antiviral medication used to treat influenza, has been shown in an in vitro study to inhibit SARS-CoV-2 infection [53]. However, no conclusive efficacy of arbidol in human use has been reported to date. Oseltamivir, a neuramidase inhibitor and another medication used to treat influenza, has not shown any effectiveness in the prevention or treatment of SARS-CoV-2 infection [54]. Similarly, zanamivir, another neuramidase inhibitor, has not demonstrated efficacy against SARS-related coronavirus (SARS-CoV) infections in the past [55].

Several medications have been hypothesized to have a reduction in the senescence that occurs with SARS-CoV-2 infection. Rapamycin, a protein synthesis inhibitor, when combined with azithromycin, has shown promise with preventing the onset of senescence and inhibiting viral replication. Doxycycline, which is an antibiotic medication, has shown antiviral activity in mammals, as well as a reduction of IL-6 serum levels and an antisenescent effect. Quercetin, which is a pigment found in food and plants, and found in supplement form, has been proposed as a potential binder of SARS-CoV-2 to reduce virus-host interactions with ACE2 receptors [36].

Tilorone is a broad-spectrum antiviral medication that has shown promise for MERS-CoV, Chikungunya, Ebola, and Marburg infections. It has been shown to induce interferon [56]. Teicoplanin, which is a glycopeptide antibiotic drug, has shown to be effective against SARS-CoV, Ebola, influenza, hepatitis C, HIV, and MERS-CoV in vitro [57].

Acyclovir and ganciclovir are nucleoside analog antiviral drugs. They are typically used to target DNA viruses such as herpes simplex virus and varicella-zoster viruses but have not shown efficacy against older SARS-CoV infections [55].

Antimalarial Drugs

Hydroxychloroquine is a drug that is currently used for malarial infection, rheumatoid arthritis (RA), and systemic lupus erythematosus (SLE). Its anti-inflammatory and immunomodulatory mechanisms of action in RA and SLE are largely unknown, and its antimalarial activity is thought to be related to inhibiting the polymerization of heme molecules within parasites [58]. Additionally, chloroquine has been shown to prevent the induction and accumulation of beta-galactosidase, which is a biomarker of cellular senescence [36].

Both in vitro studies and physiologically based pharmacokinetic models have recently shown that both chloroquine and hydroxychloroquine have inhibited the growth of SARS-CoV-2, with hydroxychloroquine being more potent and with a greater safety profile than chloroquine. The mechanism is unknown but thought to be related to the immunomodulatory mechanisms found inherent to these drugs [59].

An open-label, nonrandomized clinical trial from France showed a significant effect of hydroxychloroquine and an even greater drug potency when it was combined with azithromycin, an antibiotic medication [54].

Corticosteroid Drugs

Methylprednisolone is a steroid medication largely used for anti-inflammatory purposes in endocrine disorders, rheumatic disorders, collagen diseases, dermatologic disease, allergic states, ophthalmic diseases, respiratory diseases, hematologic disorders, neoplastic diseases, edematous states, gastrointestinal diseases, nervous system disorders, and others [60]. Methylprednisolone, used in conjunction with tocilizumab and lopinavir therapy, has shown promise in the treatment of SARS-CoV-2 infection through decreased oxygen requirements, decrease in C-reactive protein, increased lymphocyte levels, decreased fever, and improved chest tightness, potentially related to depression of inflammatory cytokine involvement [61].

Convalescent Plasma Therapy

Patients who have contracted SARS-CoV-2 infection have shown to seroconvert for immunoglobulin G and immunoglobulin M antibodies, which may serve to be useful for diagnosis of either infection or exposure [62]. Current research is being conducted on the administration of antibodies to patients with SARS-CoV-2 infection with the intent of neutralizing the virus. Options for this treatment include monoclonal antibodies, synthetic preparations grown in animal hosts, or administration of antibodies harvested from patients who have recovered from SARS-CoV-2 infection. This therapy has shown efficacy in past treatment of SARS coronavirus (SARS-CoV-1) and Middle East respiratory syndrome infections [63].

Other Therapy

The use of type 1 interferons including interferon alpha and interferon beta is being investigated for treatment of SARS-CoV-2. This therapy is typically used for treatment of multiple sclerosis or other disorders impacted by immunomodulatory drugs. However, multiple in vitro and in vivo studies have been conducted on the effect of interferons on MERS-CoV and SARS-CoV with either inconclusive results or a mild improvement of symptoms [64].

One drug used for myeloproliferative neoplasms, Fedratinib, targets the receptors for Janus kinase 2, which is a gene involved with the growth and division of cells. There is some speculation that this could be used for modulation of the inflammatory processes involved in a SARS-CoV-2 infection [65].

Some researchers have suggested using serum albumin as a delivery vehicle for potential drugs to ensure their uptake [66].

Discussion

Principal Findings

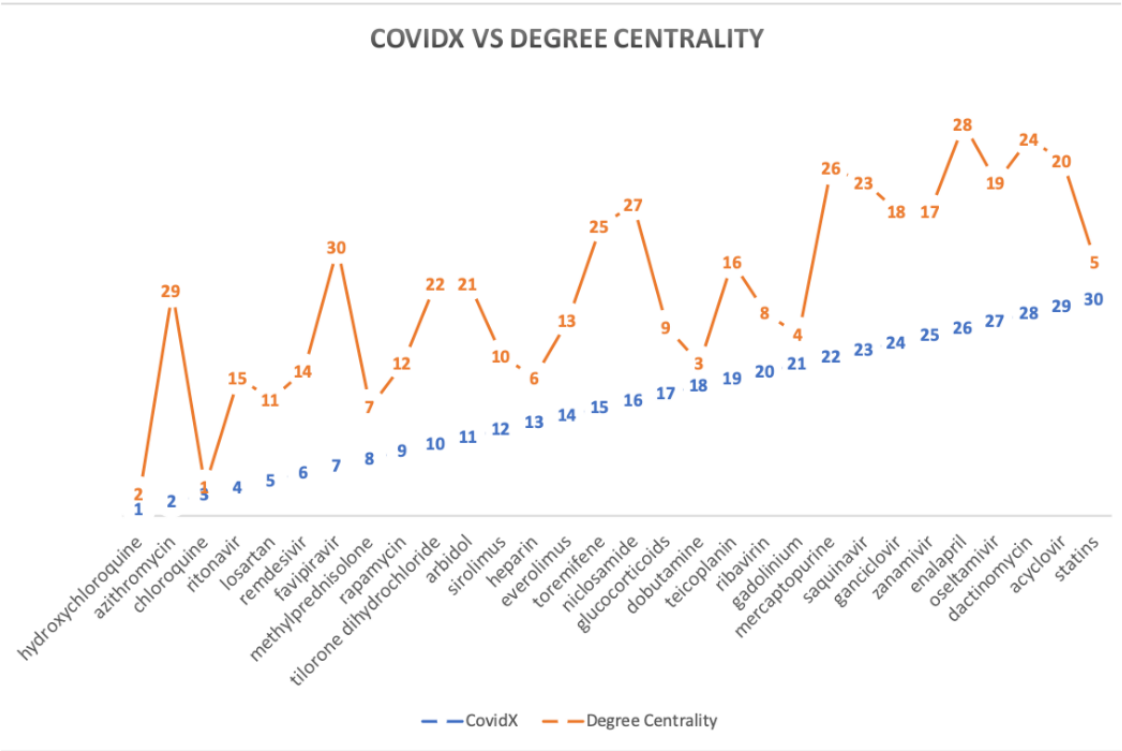
Here we discuss the results of the CovidX algorithm and how such results are compared against counterparts produced using common centrality methods (eg, the degree and eigenvector centrality). The research identified 30 drugs found in the SARS-CoV-2 literature abstracts publicly available in PubMed. According to CovidX, each drug is ranked based on two factors: (1) the variability factor of the known information about each drug and (2) whether the drug is currently under investigation in ClinicalTrials.gov [34]. The following is a list of the top 10 according to the algorithm: hydroxychloroquine, azithromycin, chloroquine, ritonavir, losartan, remdesivir, favipiravir, methylprednisolone, rapamycin, and tilorone dihydrochloride. Given the top 10 ranking generated, it is promising to see what the algorithm produced and highlighted. It is a well-known fact that drugs such as hydroxychloroquine, chloroquine, and remdesivir are occupying the headlines of major news networks and websites. It is a valid argument to claim that the frequency of the drug being investigated in a clinical trial has impacted the ranking mechanism. In fact, we believe the contribution of the clinical trial factor has certainly provided a measurable notion of support and guided the algorithm to stay within an acceptable range of being promising. The clinical trials alone, however, masked the importance of other drugs that have been studied in publications but not in clinical trials.

To better our understanding of the CovidX rankings, we compared the result of the algorithm against closely related centrality measures (more specifically, the degree of centrality, which studies the drug and the number of connected neighbors regardless of their type, and the eigenvector centrality, which is comparable to the sum of the columns that CovidX's variability measure is based upon). Table 2 shows the list of drugs and their degree centrality and eigenvector centrality. Comparing the results of produced ranks using the two measures have provided a common overlapping theme. Both measures show that only three drugs among the top 10 were found in common with the ones produced by CovidX, namely, hydroxychloroquine, chloroquine, and methylprednisolone. Figure 5 shows how the CovidX ranking is compared to the degree centrality of the network.

Table 2. Summary of the network centrality analysis.

Drug	Degree centrality	Eigenvector centrality
Acyclovir	0.363636	0.069286
Arbidol	0.189394	0.030948
Azithromycin	0.151515	0.002197
Chloroquine	0.151515	0.496042
Dactinomycin	0.121212	0.015459
Dobutamine	0.098485	0.256032
Enalapril	0.090909	0.0101
Everolimus	0.090909	0.034237
Favipiravir	0.083333	0.014957
Gadolinium	0.068182	0.256032
Ganciclovir	0.060606	0.069286
Glucocorticoids	0.05303	0.030512
Heparin	0.05303	0.09582
Hydroxychloroquine	0.037879	0.182739
Losartan	0.037879	0.014163
Mercaptopurine	0.037879	0.015459
Methylprednisolone	0.037879	0.118721
Niclosamide	0.037879	0.021294
Oseltamivir	0.037879	0.069286
Rapamycin	0.037879	0.034237
Remdesivir	0.030303	0.063039
Ribavirin	0.022727	0.153115
Ritonavir	0.015152	0.063039
Saquinavir	0.015152	0.016058
Sirolimus	0.015152	0.049697
Statins	0.015152	0.05124
Teicoplanin	0.015152	0.063039
Tilorone dihydrochloride	0.007576	0.017456
Toremifene	0.007576	0.015459
Zanamivir	0.007576	0.069286

Figure 5. How the degree of centrality as a ranking mechanism is compared to the CovidX algorithm. The horizontal axis is the drugs listed in alphabetical order, and the vertical axis shows the ranking.



Our algorithm shows that hydroxychloroquine is our top candidate, but it was scored as second in the degree of centrality. It is surprising to see that azithromycin scored second according to the CovidX algorithm, but it was scored 29th according to the degree. We also notice some close terms in the rank of chloroquine (third vs first) and methylprednisolone (eighth vs seventh).

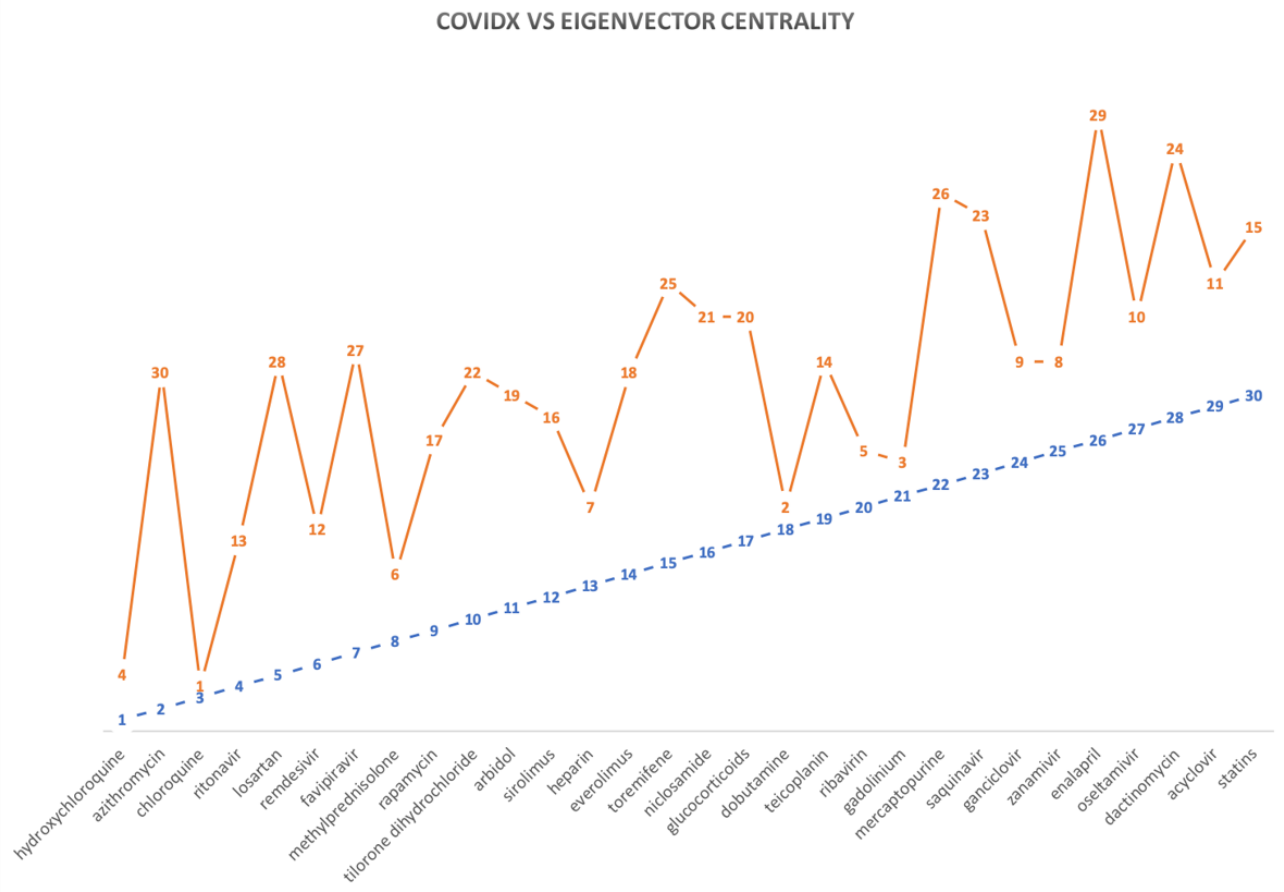
We also noted that other promising drugs such as remdesivir scored 14th in the degree and 12th in the eigenvector centrality, while it scored sixth according to CovidX. By observing the lower-ranked drugs produced by CovidX, we also noticed that statin drugs scored at the bottom of our list and were also ranked much higher according to both degree and eigenvector centrality (fifth and 15th, respectively). Another similar observation was also noticed for a drug known as gadolinium where it was ranked fourth and third according to degree and eigenvector centrality, respectively, while CovidX suppressed its rank to the 21st position in the list. Figure 6 depicts the ranking comparison between CovidX and the eigenvector network centrality. The figure shows a slight overlap in the ranking, as in the case of tilorone dihydrochloride, which was scored in the 10th position of each measure. We also observed that an overlap in the ranking of the top 10, where chloroquine is in the third position versus the first position in degree of centrality. In addition, hydroxychloroquine is ranked as the top drug in CovidX, while it is ranked in the second position according to the eigenvector centrality.

Although studying the biomedical literature stored in PubMed has clearly paved the way to highlight the centrality of drugs and hence their significance to treat COVID-19, we also believe the role of the clinical trial factor, as an external factor, has transformed the ranking and provided real-world insights that the CovidX algorithm has rightfully used. Another significant aspect of incorporating the clinical trials factor is that it can be frequently updated to produce new ranks to the drugs as new clinical trials are launched.

It is important to note that the ranking algorithm we provide here is only a first step on the path of identifying drugs that can be repurposed. However, some of the findings of this paper have already elaborated on the importance of adjuvant therapy (more than one drug to work together) to treat the symptoms. Some drugs may focus on enhancing the immune system, another may enhance the lung function, and another could work against the virus. To achieve these therapies, further clustering analysis must be performed against the antiviral drugs listed in this paper as a starting point. Based on the outcome of this future study, another ranking mechanism might be needed to identify the most promising adjuvants.

In conclusion, this research presents a network analysis method to explore the SARS-CoV-2 literature to provide a comprehensive map of the drug mentions and their co-occurring biological features (eg, gene, protein, chemical compounds, cell). Such a knowledge by itself is significant and may enable more advanced science to identify not only a drug but also a biologic therapy or a vaccine. We have taken this research further by incorporating real-world evidence from clinical trials, and we made recommendations that are subject to change given new evidence. In the future, we hope to expand our network to look beyond publications that are not only concerned with SARS-CoV-2. We also hope to study other real-world evidence factors that may contribute to the future version of the CovidX algorithm.

Figure 6. A comparison of the ranks computed by the CovidX algorithm vs the eigenvector centrality of each drug. The horizontal axis is the drugs listed in alphabetical order, and the vertical axis shows the ranking.



Limitations

The results of this research were strictly derived from the basis of existing research embedded in publications available in the PubMed archive.

Conclusions

Given how dynamic the rapid development of this pandemic is, we conclude that our ranking here, though promising, is temporary and contingent upon future evidence. We also believe that it could be the case that a few drugs will be used as an adjuvant treatment as opposed to a single drug. Based on new evidence from the literature and clinical trials, the ranking will be a step toward a full treatment. Future efforts will invest in clustering the ranked drugs in an attempt to predict how adjuvant therapy may contribute to the full treatment of the disease.

Acknowledgments

This research is partly supported by Norwich University Applied Research Institute. The authors wish to thank Osama Harraz, Yomna Gohar, Tamer Fandy, and Page Spiess for their input from a pharmaceutical and biochemistry point of view, and Laila Hamed for her inspiring ideas related to some figures in this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study input files, study intermediate output file, study results, and source code in Jupyter Notebooks.
[ZIP File (Zip Archive), 3306 KB - jmir_v22i8e21169_app1.zip]

References

1. Cohut M. Novel coronavirus: what we know so far. Medical News Today. 2020. URL: https://www.medicalnewstoday.com/articles/novel-coronavirus-your-questions-answered [accessed 2020-08-14]

2. COVID-19 coronavirus pandemic. Worldometer. URL: https://www.worldometers.info/coronavirus/#countries [accessed 2020-07-08]
3. Interim clinical guidance for management of patients with confirmed coronavirus disease (COVID-19). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html [accessed 2020-08-14]
4. Ho JSY, Tambyah PA, Sia C. A call for vaccine against COVID-19: implications for cardiovascular morbidity and healthcare utilization. Cardiovasc Drugs Ther 2020 Aug;34(4):585-587 [FREE Full text] [doi: 10.1007/s10557-020-06985-z] [Medline: 32385785]
5. Gns HS, Gr S, Murahari M, Krishnamurthy M. An update on drug repurposing: re-written saga of the drug's fate. Biomed Pharmacother 2019 Feb;110:700-716 [FREE Full text] [doi: 10.1016/j.biopha.2018.11.127] [Medline: 30553197]
6. Sanseau P, Koehler J. Editorial: computational methods for drug repurposing. Brief Bioinform 2011 Jul;12(4):301-302. [doi: 10.1093/bib/bbr047] [Medline: 21768130]
7. Issa N, Byers S, Dakshanamurthy S. Drug repurposing: translational pharmacology, chemistry, computers and the clinic. Curr Top Med Chem 2013;13(18):2328-2336. [doi: 10.2174/15680266113136660163] [Medline: 24059462]
8. Katragkou A, Roilides E, Walsh TJ. Can repurposing of existing drugs provide more effective therapies for invasive fungal infections? Expert Opin Pharmacother 2016 Jun;17(9):1179-1182. [doi: 10.1080/14656566.2016.1186647] [Medline: 27217005]
9. Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. A survey of current trends in computational drug repositioning. Brief Bioinform 2016 Jan;17(1):2-12 [FREE Full text] [doi: 10.1093/bib/bbv020] [Medline: 25832646]
10. Sadeghi SS, Keyvanpour MR. An analytical review of computational drug repurposing. IEEE/ACM Trans Comput Biol Bioinform 2019 Aug 12:PP. [doi: 10.1109/TCBB.2019.2933825] [Medline: 31403439]
11. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Comput Biol 2012;8(5):e1002503 [FREE Full text] [doi: 10.1371/journal.pcbi.1002503] [Medline: 22589709]
12. Wu C, Gudivada RC, Aronow BJ, Jegga AG. Computational drug repositioning through heterogeneous network clustering. BMC Syst Biol 2013;7(Suppl 5):S6. [doi: 10.1186/1752-0509-7-s5-s6]
13. Chen H, Fuller S, Friedman C, Hersh W. Medical Informatics Knowledge Management and Data Mining in Biomedicine. Boston, MA: Springer; 2005.
14. Zheng S, Dharssi S, Wu M, Li J, Lu Z. Text Mining for Drug Discovery. New York, NY: Humana Press; 2019:231-252.
15. Balakin KV. Pharmaceutical Data Mining: Approaches and Applications for Drug Discovery. Hoboken, New Jersey: Wiley; 2009.
16. Wu H, Shendre A, Zhang S, Zhang P, Wang L, Zeruesenay D, et al. Translational knowledge discovery between drug interactions and pharmacogenetics. Clin Pharmacol Ther 2020 Apr;107(4):886-902. [doi: 10.1002/cpt.1745] [Medline: 31863452]
17. Andronis C, Sharma A, Virvilis V, Deftereos S, Persidis A. Literature mining, ontologies and information visualization for drug repurposing. Brief Bioinform 2011 Jul;12(4):357-368. [doi: 10.1093/bib/bbr005] [Medline: 21712342]
18. Loging W, Rodriguez-Esteban R, Hill J, Freeman T, Miglietta J. Cheminformatic/bioinformatic analysis of large corporate databases: application to drug repurposing. Drug Discovery Today: Therapeutic Strategies 2011 Dec;8(3-4):109-116. [doi: 10.1016/j.ddstr.2011.06.004]
19. Deftereos SN, Andronis C, Friedla EJ, Persidis A, Persidis A. Drug repurposing and adverse event prediction using high-throughput literature analysis. Wiley Interdiscip Rev Syst Biol Med 2011;3(3):323-334. [doi: 10.1002/wsbm.147] [Medline: 21416632]
20. Nabirotchkin S, Peluffo AE, Rinaudo P, Yu J, Hajj R, Cohen D. Next-generation drug repurposing using human genetics and network biology. Curr Opin Pharmacol 2020 Jan 22:A [FREE Full text] [doi: 10.1016/j.coph.2019.12.004] [Medline: 31982325]
21. Luo H, Li M, Wang S, Liu Q, Li Y, Wang J. Computational drug repositioning using low-rank matrix approximation and randomized algorithms. Bioinformatics 2018 Jun 01;34(11):1904-1912. [doi: 10.1093/bioinformatics/bty013] [Medline: 29365057]
22. Karatzas E, Bourdakou MM, Kolios G, Spyrou GM. Drug repurposing in idiopathic pulmonary fibrosis filtered by a bioinformatics-derived composite score. Sci Rep 2017 Oct 03;7(1):12569. [doi: 10.1038/s41598-017-12849-8] [Medline: 28974751]
23. Hamed A, Leszczynska A, Schreiber M. MolecRank: a specificity-based network analysis algorithm. 2019 Presented at: International Conference on Advanced Machine Learning Technologies and Applications; 28 March, 2019; Egypt. [doi: 10.1007/978-3-030-14118-9_16]
24. Nadeau D, Satoshi S. A survey of named entity recognition and classification. Lingvisticae Investigationes 2007 Aug 10;30(1):3-26. [doi: 10.1075/li.30.1.03nad]
25. Project Jupyter. 2020 Jul 08. URL: https://www.jupyter.org [accessed 2020-07-08]
26. TextBlob. Google Search. URL: https://www.google.com/search?q=TextBlob&oq=TextBlob&aqs=chrome.69i57j35i39j69i59j0l5.2340j0j4&sourceid=chrome&ie=UTF-8 [accessed 2020-08-14]

27. NetworkX: network analysis in Python. GitHub. URL: https://networkx.github.io/ [accessed 2020-08-14]
28. Python iGraph. URL: https://igraph.org/python/ [accessed 2020-08-14]
29. Christopher M, Hinrich S. Foundations of Statistical Natural Language Processing. Cambridge: MIT Press; May 1999.
30. Müller HM, Kenny EE, Sternberg PW. Textpresso: an ontology-based information retrieval and extraction system for biological literature. PLoS Biol 2004 Nov;2(11):e309 [FREE Full text] [doi: 10.1371/journal.pbio.0020309] [Medline: 15383839]
31. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res 2015 Jan;43(Database issue):D1049-D1056 [FREE Full text] [doi: 10.1093/nar/gku1179] [Medline: 25428369]
32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000 May;25(1):25-29 [FREE Full text] [doi: 10.1038/75556] [Medline: 10802651]
33. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 2004 Jan 01;32(Database issue):D258-D261 [FREE Full text] [doi: 10.1093/nar/gkh036] [Medline: 14681407]
34. ClinicalTrials.gov. URL: https://clinicaltrials.gov/ [accessed 2020-08-14]
35. South AM, Diz DI, Chappell MC. COVID-19, ACE2, and the cardiovascular consequences. Am J Physiol Heart Circ Physiol 2020 May 01;318(5):H1084-H1090. [doi: 10.1152/ajpheart.00217.2020] [Medline: 32228252]
36. Sargiacomo C, Sotgia F, Lisanti MP. COVID-19 and chronological aging: senolytics and other anti-aging drugs for the treatment or prevention of corona virus infection? Aging (Albany NY) 2020 Mar 30;12(8):6511-6517 [FREE Full text] [doi: 10.18632/aging.103001] [Medline: 32229706]
37. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
38. Guastalegname M, Vallone A. Could Chloroquine /Hydroxychloroquine be harmful in coronavirus disease 2019 (COVID-19) treatment? Clin Infect Dis 2020 Jul 28;71(15):888-889 [FREE Full text] [doi: 10.1093/cid/ciaa321] [Medline: 32211771]
39. Savarino A, Boelaert J, Cassone A, Majori G, Cauda R. Effects of chloroquine on viral infections: an old drug against today's diseases. Lancet Infect Dis 2003 Nov;3(11):722-727. [doi: 10.1016/S1473-3099(03)00806-5]
40. Velavan TP, Meyer CG. The COVID-19 epidemic. Trop Med Int Health 2020 Mar;25(3):278-280. [doi: 10.1111/tmi.13383] [Medline: 32052514]
41. He R, Lu Z, Zhang L, Fan T, Xiong R, Shen X, et al. The clinical course and its correlated immune status in COVID-19 pneumonia. J Clin Virol 2020 Jun;127:104361 [FREE Full text] [doi: 10.1016/j.jcv.2020.104361] [Medline: 32344320]
42. Zhang C, Shi L, Wang F. Liver injury in COVID-19: management and challenges. Lancet Gastroenterol Hepatol 2020 May;5(5):428-430 [FREE Full text] [doi: 10.1016/S2468-1253(20)30057-1] [Medline: 32145190]
43. Wang Z, Yang B, Li Q, Wen L, Zhang R. Clinical features of 69 cases with coronavirus disease 2019 in Wuhan, China. Clin Infect Dis 2020 Jul 28;71(15):769-777 [FREE Full text] [doi: 10.1093/cid/ciaa272] [Medline: 32176772]
44. Rotzinger D, Beigelman-Aubry C, von Garnier C, Qanadli S. Pulmonary embolism in patients with COVID-19: time to change the paradigm of computed tomography. Thromb Res 2020 Jun;190:58-59 [FREE Full text] [doi: 10.1016/j.thromres.2020.04.011] [Medline: 32302782]
45. Han H, Yang L, Liu R, Liu F, Wu KL, Li J, et al. Prominent changes in blood coagulation of patients with SARS-CoV-2 infection. Clin Chem Lab Med 2020 Jun 25;58(7):1116-1120. [doi: 10.1515/cclm-2020-0188] [Medline: 32172226]
46. Tang N, Bai H, Chen X, Gong J, Li D, Sun Z. Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. J Thromb Haemost 2020 May;18(5):1094-1099. [doi: 10.1111/jth.14817] [Medline: 32220112]
47. Testa S, Paoletti O, Giorgi-Pierfranceschi M, Pan A. Switch from oral anticoagulants to parenteral heparin in SARS-CoV-2 hospitalized patients. Intern Emerg Med 2020 Aug;15(5):751-753 [FREE Full text] [doi: 10.1007/s11739-020-02331-1] [Medline: 32297089]
48. Zumla A, Chan JFW, Azhar EI, Hui DSC, Yuen KY. Coronaviruses - drug discovery and therapeutic options. Nat Rev Drug Discov 2016 May;15(5):327-347 [FREE Full text] [doi: 10.1038/nrd.2015.37] [Medline: 26868298]
49. Ortega J, Serrano M, Pujol F, Rangel H. Unrevealing sequence and structural features of novel coronavirus using approaches: the main protease as molecular target. EXCLI J 2020;19:400-409 [FREE Full text] [doi: 10.17179/excli2020-1189] [Medline: 32210741]
50. Khan SA, Zia K, Ashraf S, Uddin R, Ul-Haq Z. Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 integrated computational approach. J Biomol Struct Dyn 2020 Apr 13:1-10. [doi: 10.1080/07391102.2020.1751298] [Medline: 32238094]
51. Elfiky AA. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): a molecular docking study. Life Sci 2020 Jul 15;253:117592 [FREE Full text] [doi: 10.1016/j.lfs.2020.117592] [Medline: 32222463]
52. Pasquau Liaño J, Hidalgo Tenorio C. [Chemical characteristics, mechanism of action and antiviral activity of darunavir]. Enferm Infecc Microbiol Clin 2008 Oct;26 Suppl 10:3-9. [doi: 10.1016/s0213-005x(08)76547-9] [Medline: 19195453]

53. Dong L, Hu S, Gao J. Discovering drugs to treat coronavirus disease 2019 (COVID-19). Drug Discov Ther 2020;14(1):58-60. [doi: 10.5582/ddt.2020.01012] [Medline: 32147628]
54. Tobaiqy M, Qashqary M, Al-Dahery S, Mujallad A, Hershan A, Kamal M, et al. Therapeutic management of patients with COVID-19: a systematic review. Infect Prev Pract 2020 Sep;2(3):100061-100063. [doi: 10.1016/j.infpip.2020.100061]
55. Tan EL, Ooi EE, Lin C, Tan HC, Ling AE, Lim B, et al. Inhibition of SARS coronavirus infection in vitro with clinically approved antiviral drugs. Emerg Infect Dis 2004 Apr;10(4):581-586 [FREE Full text] [doi: 10.3201/eid1004.030458] [Medline: 15200845]
56. Ekins S, Lane T, Madrid P. Tilorone: a broad-spectrum antiviral invented in the USA and commercialized in Russia and beyond. Pharm Res 2020 Mar 25;37(4):71 [FREE Full text] [doi: 10.1007/s11095-020-02799-8] [Medline: 32215760]
57. Baron SA, Devaux C, Colson P, Raoult D, Rolain J. Teicoplanin: an alternative drug for the treatment of COVID-19? Int J Antimicrob Agents 2020 Apr;55(4):105944 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105944] [Medline: 32179150]
58. Plaquenil® Hydroxychlorioquine sulfate tablets, USP. U.S. Food & Drug Administration. URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/009768s037s045s047lbl.pdf [accessed 2020-08-14]
59. Yao X, Ye F, Zhang M, Cui C, Huang B, Niu P, et al. In vitro antiviral activity and projection of optimized dosing design of Hydroxychloroquine for the treatment of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Clin Infect Dis 2020 Jul 28;71(15):732-739 [FREE Full text] [doi: 10.1093/cid/ciaa237] [Medline: 32150618]
60. Methylprednisolone tablets. U.S. Food & Drug Administration. URL: https://www.accessdata.fda.gov/drugsatfda_docs/anda/98/40183_Methylprednisolone_Prntlbl.pdf [accessed 2020-08-14]
61. Buonaguro F, Puzanov I, Ascierto P. Anti-IL6R role in treatment of COVID-19-related ARDS. J Transl Med 2020 Apr 14;18(1):165 [FREE Full text] [doi: 10.1186/s12967-020-02333-9] [Medline: 32290847]
62. Long Q, Liu B, Deng H, Wu G, Deng K, Chen Y, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. Nat Med 2020 Jun;26(6):845-848. [doi: 10.1038/s41591-020-0897-1] [Medline: 32350462]
63. Casadevall A, Pirofski L. The convalescent sera option for containing COVID-19. J Clin Invest 2020 Apr 01;130(4):1545-1548. [doi: 10.1172/JCI138003] [Medline: 32167489]
64. Sallard E, Lescure F, Yazdanpanah Y, Mentre F, Peiffer-Smadja N. Type 1 interferons as a potential treatment against COVID-19. Antiviral Res 2020 Jun;178:104791 [FREE Full text] [doi: 10.1016/j.antiviral.2020.104791] [Medline: 32275914]
65. Wu D, Yang XO. TH17 responses in cytokine storm of COVID-19: an emerging target of JAK2 inhibitor Fedratinib. J Microbiol Immunol Infect 2020 Jun;53(3):368-370 [FREE Full text] [doi: 10.1016/j.jmii.2020.03.005] [Medline: 32205092]
66. Mani Mishra P, Uversky VN, Nandi CK. Serum albumin-mediated strategy for the effective targeting of SARS-CoV-2. Med Hypotheses 2020 Apr 24;140:109790 [FREE Full text] [doi: 10.1016/j.mehy.2020.109790] [Medline: 32353740]

Abbreviations

ACE2: angiotensin-converting enzyme 2
ANG II: angiotensin II
ARDS: acute respiratory distress syndrome
CD: cluster of differences
COVID-19: coronavirus disease
CT: computed tomography
DIC: disseminated intravascular coagulation
IL: interleukin
MERS-CoV: Middle East respiratory syndrome–related coronavirus
RA: rheumatoid arthritis
RAAS: renin-angiotensin-aldosterone system
RdRp: RNA-dependent RNA polymerase
SARS: severe acute respiratory syndrome
SARS-CoV: severe acute respiratory syndrome–related coronavirus
SARS-CoV-1: severe acute respiratory syndrome coronavirus
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
SLE: systemic lupus erythematosus
Th: T helper
VKA: vitamin K antagonists

Edited by G Eysenbach; submitted 11.06.20; peer-reviewed by T Fandy, Y Li, K Koshechkin; comments to author 06.07.20; revised version received 09.07.20; accepted 24.07.20; published 20.08.20.

Please cite as:
Gates LE, Hamed AA
The Anatomy of the SARS-CoV-2 Biomedical Literature: Introducing the CovidX Network Algorithm for Drug Repurposing Recommendation
J Med Internet Res 2020;22(8):e21169
URL: http://www.jmir.org/2020/8/e21169/
doi:10.2196/21169
PMID:32735546

Original Paper

Effects of Internet Hospital Consultations on Psychological Burdens and Disease Knowledge During the Early Outbreak of COVID-19 in China: Cross-Sectional Survey Study

Lin Li[1*], MD; Gang Liu[2*], MD; Weiguo Xu[1], MD; Yun Zhang[3], MD; Mei He[4], MD

[1]Department of Respiratory and Critical Care Medicine, Mianyang Central Hospital, Mianyang, China

[2]Department of Orthopedics, Mianyang Central Hospital, Mianyang, China

[3]Department of Neurology, Mianyang Central Hospital, Mianyang, China

[4]Department of President's Office, Mianyang Central Hospital, Mianyang, China

[*]these authors contributed equally

Corresponding Author:
Mei He, MD
Department of President's Office
Mianyang Central Hospital
12# Changjia Lane, Jingzhong Street
Fucheng District
Mianyang, 621000
China
Phone: 1 13981110800
Email: hemeimy@163.com

Abstract

Background: Coronavirus disease (COVID-19) has become a global threat to human health. Internet hospitals have emerged as a critical technology to bring epidemic-related web-based services and medical support to the public. However, only a few very recent scientific literature reports have explored the effects of internet hospitals on psychological burden and disease knowledge in major public health emergencies such as the COVID-19 pandemic.

Objective: The aim of this study was to explore the role of internet hospitals in relieving psychological burden and increasing disease knowledge during the early outbreak of the COVID-19 pandemic.

Methods: This survey was conducted from January 26 to February 1, 2020, during the early outbreak of COVID-19 in China. The platform used for the consultation was the WeChat public account of our hospital. To participate in the study, the patient was required to answer a list of questions to exclude the possibility of COVID-19 infection and confirm their willingness to participate voluntarily. Next, the participant was directed to complete the self-report questionnaire. After the internet consultation, the participant was directed to complete the self-report questionnaire again. The questionnaire included sections on general information, the General Health Questionnaire-28 (GHQ-28), and the participant's worries, disease knowledge, and need for hospital treatment.

Results: The total number of internet consultations was 4120. The consultation topics mainly included respiratory symptoms such as cough, expectoration, and fever (2489/4120, 60.4%) and disease knowledge, anxiety, and fear (1023/4120, 24.8%). A total of 1530 people filled out the questionnaires before and after the internet consultation. Of these people, 1398/1530 (91.4%) experienced psychological stress before the internet consultation, which significantly decreased after consultation (260/1530, 17.0%) (χ^2_1=1704.8, P<.001). There was no significant difference in the number of people who expressed concern about the COVID-19 pandemic before and after the internet consultation (χ^2_1=0.7, P=.43). However, the degree of concern after the internet consultation was significantly alleviated (t_{2699}=90.638, P<.001). The main worries before and after consultation were the dangers posed by the disease and the risk of infection of family members. The scores of the self-assessment risk after the internet consultation were significantly lower than those before consultation (t_{3058}=95.694, P<.001). After the consultation, the participants' knowledge of the symptoms, transmission routes, and preventive measures of COVID-19 was significantly higher than before the consultation (t_{3058}=–106.105, –80.456, and –152.605, respectively; all P<.001). The hospital treatment need score after the internet consultation decreased from 3.3 (SD 1.2) to 1.6 (SD 0.8), and the difference was statistically significant (t_{3058}=45.765, P<.001).

Conclusions: During the early outbreak of COVID-19, internet hospitals could help relieve psychological burdens and increase disease awareness through timely and rapid spread of knowledge regarding COVID-19 prevention and control. Internet hospitals should be an important aspect of a new medical model in public health emergency systems.

(J Med Internet Res 2020;22(8):e19551) doi:10.2196/19551

KEYWORDS

internet hospital; telemedicine; novel coronavirus disease; pandemic; psychological burden; disease cognition; coronavirus; COVID-19; public health; infectious disease

Introduction

From severe acute respiratory syndrome (SARS) and H_1N_1 influenza to Middle East respiratory syndrome (MERS) and the latest new coronaviruses, new infectious diseases have become a global problem that seriously threatens human health [1,2]. In contrast to other viruses, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes coronavirus disease (COVID-19), has the characteristics of a long incubation period and strong infectivity [3,4]. On January 30, 2020, the World Health Organization declared the COVID-19 outbreak to be a public health emergency of international concern; at this point, all regions of China had reported cases of infection [5]. Soon, COVID-19 began spreading worldwide. The panic caused by this worldwide public health emergency has far surpassed those caused by the Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV) pandemics [6]. Previous research has revealed a wide and profound range of impacts on people at the individual, community, and international levels during outbreaks of infection [7]. Because winter and spring are the peak seasons of influenza, people who have symptoms of upper respiratory infection, such as fever, fatigue, and cough, often fear that they are infected with COVID-19 [8]. If many people rush to the hospital in a short time, medical resources will be occupied and potential cross-infection and other problems will occur. Therefore, medical support at this time is essential. To help people seek timely professional medical advice and accurately guide them on when and how to report to a hospital, as well as to reduce the risk of cross-infection from population flow, the Chinese government has issued decrees that require domestic internet hospitals to make full use of the advantages of telemedicine to provide convenient and high-quality services in response to the epidemic [9,10]. Therefore, our hospital quickly responded and immediately set up free COVID-19 consultation services as the main form of telemedicine for the public during the epidemic. This action attracted widespread attention from all walks of life. Our internet hospital received more than 35,000 hits on in the first day, and the number of free medical consultations exceeded 1000 [11]. Internet hospitals, as a 21st-century approach to triage, enable efficient screening of patients and protect patients, clinicians, and the community from exposure during an infectious public health emergency [12]. However, only a few very recent scientific literature reports have explored the effects of internet hospitals on psychological burden and disease knowledge in major public health emergencies such as the COVID-19 pandemic. Thus, in addition to discussing symptoms and distinguishing disease, the clinicians also provide disease knowledge as well as medical or emotional support and understanding during internet consultation. A structured questionnaire containing questions regarding patients' demographics, psychological state, and disease knowledge was used to explore the changes in these items before and after consultation. The goal of this research was to explore the role of internet hospitals in relieving psychological burden and increasing disease knowledge during the early outbreak of COVID-19 in China.

Methods

Participants

This survey was conducted from January 26 to February 1, 2020, which was the early period of the COVID-19 outbreak in China. The questionnaire was distributed as an electronic form that was collected and analyzed through the internet. Inclusion criteria were people who were aged ≥18 years and who completed the questionnaire. Exclusion criteria were people aged ≤17 years and who provided illogical responses to the questionnaire.

The Ethics Committee of Mianyang Central Hospital approved our study protocol and procedures of informed consent before the formal survey. Participants were required to answer a yes or no question to confirm their willingness to participate voluntarily. After confirmation of the question, the participant was directed to complete the self-report questionnaire.

Internet Consultation Procedure

Before the internet hospital and web-based medical consultation system was set up, the project working group and engineers provided targeted training to the expert team, including the clinician-side use of internet hospitals, real-name registration, web-based consultation procedures, screening processes, and solutions to common problems. The online consultation was free to all. These free services could be obtained through the public WeChat account of our hospital by clicking Integrated Services, then clicking Internet Hospital, and finally clicking COVID-19 Consultation; then, real-time communication could be realized between clinicians and patients.

Free internet medical consultations for "fighting against COVID-19" could be accessed from 8 AM to 9 PM every day starting on January 26, 2020; clinicians talked with patients about their symptoms, helped them distinguish COVID-19 from the common cold, provided disease knowledge, and offered medical or emotional support and understanding.

The procedure of free COVID-19 consultation services was as follows. First, at the start of the conversation, the participant was required to answer a list of questions regarding their

respiratory symptoms and provide detailed travel and exposure histories for screening. Participants who were not suspected of being infected with COVID-19 could proceed to the next step. Next, the participant was required to answer a yes or no question to confirm their willingness to participate voluntarily. After answering this question affirmatively, the participant was directed to complete the self-report questionnaire. Then, a clinician would communicate with the participant regarding their symptoms and psychological state, whether they had developed COVID-19, prevention and control of the disease, and any questions they had, until the participant actively ended the online consultation. Finally, the participant was directed to complete the self-report questionnaire again.

Data Collection

The study instrument comprised a structured questionnaire packet including the following five sections.

1. General Information

The general information section included questions about the respondent's name, gender, age, education level, and occupation.

2. Psychological State

The General Health Questionnaire-28 (GHQ-28) was used to evaluate the respondents' psychological state. The GHQ-28 contains 28 items and is composed of four factors, namely physical condition, anxiety/insomnia, social dysfunction, and severe depression [13]. Each item is rated on a 4-level scale. The higher the total score, the greater the respondent's psychological stress. A total GHQ-28 score of ≤5 points indicates that the subject has no psychological stress, 6-11 points indicates mild to moderate psychological stress, and >11 points indicates severe psychological stress.

3. Concerns About COVID-19

This section contained questions regarding the respondents' concerns about the COVID-19 pandemic, including whether they worried about the COVID-19 pandemic, and to what degree; they were also asked whether they mostly worried about the dangers of the disease, the risk of infection of family members and friends, isolation from their family and/or social environment, or the effects on their functional ability. Each item was evaluated by a dichotomy (yes or no) or a 9-level score from 1 (mild) to 9 (severe).

4. Disease Knowledge

Questions about the respondents' disease knowledge included their self-assessed risk, self-perceived seriousness of infection and difficulty of treatment, sufficiency of disease knowledge, and level of preparation for the epidemic. Each item was evaluated by a dichotomy (yes or no) or a 9-level score from 1 (high/disagree) to 9 (low/agree).

5. Hospital Treatment Need

In this section, the respondents self-assessed their need to go to a hospital for further medical treatment. This need was also rated on a 9-level scale, from 1 (not needed) to 9 (needed).

The questionnaire administered before the internet consultation contained sections 1 to 5, and that administered after consultation contained sections 2 to 5.

Statistical Analysis

The data were organized and analyzed using SPSS 25.0 software (IBM Corporation). Measurement data were expressed as mean (SD), and the independent sample *t* test was adopted for comparison between groups. For categorical data, frequencies and percentages were used, and the chi-square test was adopted for comparisons between groups. A two-tailed *P* value <.05 was considered statistically significant.

Results

Characteristics of the Participants

A total of 1530 people filled out the questionnaire before and after the internet consultation. The average age of the 1530 respondents was 38.9 years (SD 9.9); 998 (65.2%) were women and 532 (34.8%) were men. Among the 1530 respondents, 1101 (72.0%) were aged 18 to 40 years and 429 (28.0%) were older than 40 years. Regarding the respondents' education levels, 96/1530 (6.3%) had a master's degree or higher, 307/1530 (20.1%) had a bachelor's degree, 678/1530 (44.3%) had a high school diploma, and 449/1530 (29.4%) had a junior high school education or below. Of the 1530 respondents, 212 (13.9%) were professionals (teacher/lawyer/student), 586 (38.3%) were employed in enterprises and institutions, 512 (33.5%) were laborers and service workers, and 220 (14.4%) had other occupations (retired, freelancer, or unemployed).

General Information About the Internet Consultations

A total of 4120 internet consultations were performed from January 26 to February 1, 2020. The number peaked on the first day and then gradually declined (Figure 1). The topics of discussion during the consultation mainly included respiratory symptoms such as cough, expectoration, and fever, accounting for 2489/4120 (60.4%) of the total consultations (Table 1). COVID-19 knowledge and anxiety and fear were the subjects of 1023/4120 consultations (24.8%) (Table 1). In the first 7 days, 5 people with fever and exposure histories were screened, and 136/1530 (8.9%) of the respondents needed to go to a hospital.

Figure 1. The change in the number of internet consultations with time.

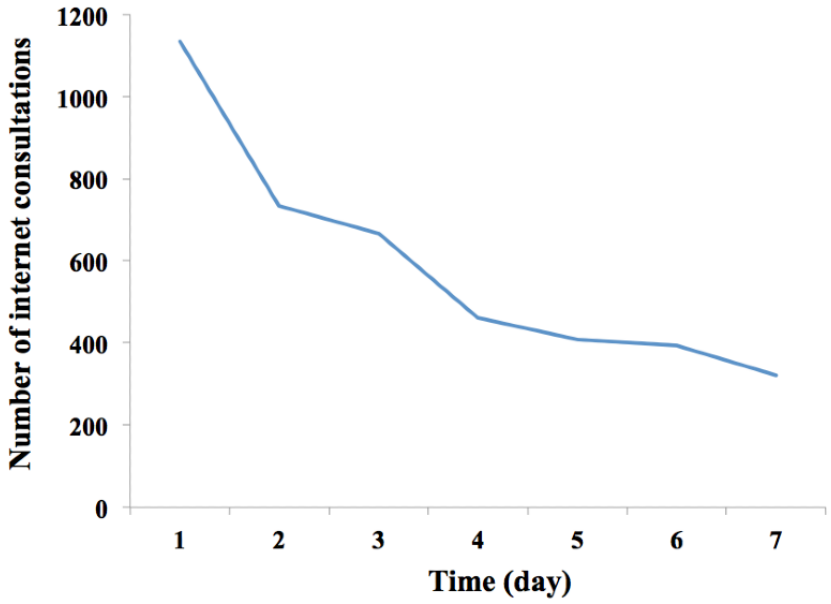


Table 1. Topics discussed during the internet consultations (N=4120).

Topic	n (%)
Respiratory symptoms, such as cough and expectoration	1628 (39.5)
Fever	861 (20.9)
COVID-19[a] knowledge	598 (14.5)
Anxiety and fear	425 (10.3)
Consultation on chronic diseases	230 (5.6)
Gastrointestinal symptoms	155 (3.8)
Other	223 (5.4)

[a]COVID-19: coronavirus disease.

Degrees of Psychological Stress Before and After the Internet Consultation

Before the internet consultation, 1398/1530 respondents (91.4%) reported experiencing psychological stress; this number was significantly higher than that after the internet consultation (260/1530, 17.0%), and the difference was statistically significant (χ^2_1=1704.8, *P*<.001). Among the 1530 respondents, 1352 (88.4%) scored 6 to 11 points on the GHQ-28, indicating light to moderate psychological stress; meanwhile, 46 (3.0%) scored >11 points on the GHQ-28, indicating severe psychological stress. The number of people with mild to moderate psychological stress before the internet consultation was significantly higher than that after consultation (χ^2_1=1693.8, *P*<.001). There was no statistically significant difference in the number of people with severe psychological stress before and after the internet consultation (*P*>.99) (Table 2).

Table 2. Degrees of psychological stress before and after the internet consultation (N=1530).

General Health Questionnaire-28 score (points)	Degree of psychological stress	Before internet consultation, n (%)	After internet consultation, n (%)	Chi-square (*df*=1)	*P* value
≤5	None	132 (8.6)	1270 (83.0)	1704.8	<.001
6-11	Mild to moderate	1352 (88.4)	214 (14.0)	1693.8	<.001
>11	Severe	46 (3.0)	46 (3.0)	<0.001	>.99

Concerns About the COVID-19 Pandemic Before and After the Internet Consultation

There was no significant difference in the percentage of people who expressed concern about the COVID-19 pandemic before and after the internet consultation (χ^2_1=0.7, *P*=.43). However, the degree of concern after the internet consultation was significantly alleviated compared to before the internet consultation (t_{2699}=90.638, *P*<.001). The main worries before and after the consultation were the dangers of the disease and the risk of infection of family members and friends. The respondents expressed fewer concerns about isolation from their family or social environment and effects on their functional ability. After the internet consultation, there were fewer concerns expressed about the dangers of the disease and the risk of infection of family members and friends than before the consultation, and the difference was statistically significant (χ^2_1=227.4 and 59.4, respectively; all *P*<.001) (Table 3).

Table 3. Concerns about the COVID-19 pandemic before and after the internet consultation (N=1530).

Question	Before internet consultation	After internet consultation	Comparison between groups	*P* value
I worry about the COVID-19[a] pandemic, n (%)	1343 (87.8)	1358 (88.8)	χ^2_1=0.7	.43
Degree of worry (1=mild, 9=severe), mean (SD)	8.0 (0.9)	3.7 (1.5)	t_{2699}=90.638	<.001
Main cause of concern, n (%)				
The dangers of the disease	1122 (73.3)	668 (43.7)	χ^2_1=227.4	<.001
The risk of infection of family and friends	948 (62.0)	736 (48.1)	χ^2_1=59.4	<.001
Isolation from family and/or my social environment	194 (12.7)	171 (11.2)	χ^2_1=1.6	.22
The effects on my functional ability	200 (13.1)	168 (11.0)	χ^2_1=3.2	.09

[a]COVID-19: coronavirus disease.

Disease Knowledge Before and After the Internet Consultation

The self-assessment risk scores before the internet consultation were significantly higher than those after the consultation (t_{3058}=95.694, *P*<.001). There was no statistically significant difference in the scores of the self-perceived seriousness of infection and difficulty of treatment before and after the internet consultation (t_{3058}=1.333, *P*=.18 and t_{3058}=0.242, *P*=.81). After consultation, the respondents' knowledge of the symptoms, transmission routes, and preventive measures of COVID-19 was higher than that before consultation, and the difference was statistically significant (t_{3058}=–106.105, –80.456, and –152.605, respectively; all *P*<.001). In terms of disease treatment and prognosis, the levels of knowledge were relatively low both before and after the internet consultation, and the respondents all reported that they were not prepared to fight the epidemic (*P*=.07) (Table 4).

Table 4. Respondents' knowledge regarding COVID-19 before and after the internet consultation.

Question	Before internet consultation, mean (SD)	After internet consultation, mean (SD)	t_{3058}	*P* value
Self-assessment risk score (1=low, 9=high)	8.1 (0.9)	4.2 ± 1.3	95.694	<.001
I think infection may seriously impair my health (1=disagree, 9=agree)	5.1 (1.2)	5.1 (1.0)	1.333	.18
I think COVID-19[a] is hard to cure after infection (1=disagree, 9=agree)	4.6 (1.1)	4.6 (1.3)	0.242	.81
I think I have sufficient information about (1=disagree, 9=agree):				
Symptoms	3.7 (1.1)	7.8 (1.0)	–106.105	<.001
Treatment	3.2 (1.1)	3.2 (1.2)	–0.765	.45
Prognosis	2.9 (1.0)	2.9 (1.4)	–1.579	.11
Transmission routes	4.7 (1.0)	7.7 (1.0)	–80.456	<.001
Preventive measures	2.9 (0.9)	8.0 (0.9)	–152.605	<.001
I think I am prepared to fight the epidemic (1=disagree, 9=agree)	3.0 (1.5)	3.1 (1.4)	–1.799	.07

[a]COVID-19: coronavirus disease

Need for Hospital Treatment Before and After the Internet Consultation

The score of hospital treatment need after the internet consultation was 1.6 (SD 0.8), which is lower than that before the consultation (3.3, SD 1.2); the difference was statistically significant (t_{3058}=45.765, P<.001).

Discussion

Principal Findings

Our results showed that after the internet consultation, the number of respondents who had psychological stress reactions decreased and their degrees of concern about the COVID-19 pandemic were alleviated. Additionally, the respondents' knowledge of the symptoms, transmission routes, and preventive measures of COVID-19 increased after the internet consultation. Our results revealed the important role of internet hospitals in the response to infectious public health emergencies through their timely and rapid spread of knowledge regarding prevention and control of COVID-19.

In recent years, our hospital has relied on the "internet + medical care" model of our internet hospital to provide people with web-based consultations, appointments, electronic prescriptions, and other services, achieving a new mode of internet medical care with a combination of online and offline medical services. Since the outbreak of the epidemic, our hospital has quickly established a multidisciplinary project working group led by the director of the hospital and senior medical experts, including physicians, nursing experts, psychologists, and health managers [11]. Regulations were established to ensure the quality and standardization of internet consultation services, including a "Notice on the Standardized Implementation of Internet COVID-19 Consultation" and "Operating Manual for COVID-19 Online Consultation." Over 7 days, from January 26 to February 1, 2020, the total number of online consultations in the internet hospital was 4120. Among these, 2489/4120 people (60.4%) consulted on respiratory symptoms such as cough, sputum, and fever, 598/4120 (14.5%) on knowledge of COVID-19, and 425/4120 (10.3%) on psychological problems related to anxiety and fear. In addition to helping to distinguish the common cold, the symptoms of which are similar to early signs of COVID-19, we noted that consultation on the knowledge of COVID-19 and psychological problems constituted nearly 25% of the topics discussed (1023/4120 consultations, 24.8%). This showed that during the epidemic, people have a strong need for disease knowledge and a high risk of psychological stress, which provides good opportunities for telemedicine.

The results of the study showed that during the early period of the outbreak, 1343/1530 (87.8%) of people before the internet consultation and 1358/1530 (88.8%) after the internet consultation expressed concern about the COVID-19 pandemic; however the degrees of concern before and after the consultation were different. The degree of concern before the consultation was generally more severe. We also found that the number of people with psychological stress or mild to moderate psychological stress before internet consultation was significantly higher than that after consultation. This was consistent with the fact that these respondents believed that the risk of disease was high before the online consultation. Thus, it can be seen that in major public health emergencies such as the COVID-19 pandemic, internet hospitals enable people to obtain timely disease and symptom consultations, gain knowledge regarding the prevention and control of COVID-19, learn about best practices during home quarantine, and receive prompt responses to their concerns [14]. All these benefits will be propitious to reduce their psychological stress and fear about the epidemic.

The outbreak of COVID-19 occurred at the Lunar New Year in China. If we had failed to raise awareness of COVID-19 and inform people about the correct prevention and control measures of the disease in the shortest time, the epidemic would probably have spread further [15]. At this time, the traditional methods of spreading knowledge about diseases, such as centralized training, lectures, and leaflet dissemination, are costly and inefficient; additionally, these methods are not sufficiently fast, and the coverage is narrow. Accordingly, the internet has become the best tool and information carrier for epidemic prevention [16]. Our research results showed that after the internet consultation, the respondents' knowledge of the symptoms, transmission routes, and preventive measures of COVID-19 was higher than that before consultation. The respondents' psychological stress and degree of concern about the epidemic were alleviated after the internet consultation, which may be related to their deeper understanding of the disease. It has been confirmed that it is feasible and effective to promptly spread knowledge regarding epidemic prevention and control through internet hospitals and to increase people's knowledge of the disease.

In contrast to rescues in natural disasters, hospitals are obviously unable to meet the needs of all people during major public health emergencies such as the COVID-19 pandemic. Even hospitals are prone to cluster disease and virus spread [17]. Therefore, a decentralized medical assistance model is emphasized in public health emergencies, and the internet plays a key role similar to that of tiered medical service in this decentralization. After online consultation, only 136/1530 (8.89%) of the survey respondents needed to go to a hospital for further examination, and the score of hospital treatment need was also significantly lower than that before the consultation; this indicates that most online consultations are only due to the common cold or mild discomfort that can be monitored at home. Through internet consultations, these patients learn that they only require observation or can simply take nonprescription drugs. The web-based internet hospital service provided patients with necessary medical guidance during the epidemic and reduced the mobility of people to the hospital, which is extremely beneficial to cut off transmission routes and reduce person-to-person transmission [18,19].

Limitations

One limitation of this study is that it is based on a web-based questionnaire. It is possible that the survey did not reach underdeveloped areas due to limited technology availability and that it omitted people who are not comfortable using the

internet. Another limitation is that self-reported levels of psychological impact are not always aligned with assessment by mental health professionals. Irrespective of the above limitations, this study provides invaluable information on the role of internet hospitals on people during the early outbreak of COVID-19 in China. Most importantly, our findings directly demonstrate that internet hospitals are perfectly suited to this pandemic situation because they help relieve psychological burdens on patients and increase disease knowledge. These effects are crucial because ensuring that the general public is well-informed about a condition such as COVID-19 can reduce unnecessary psychological burdens without increasing risk of cross-infection.

With the increasingly powerful functions of the internet and handheld smart devices, internet hospitals are an increasingly feasible way to spread disease knowledge [20,21]. Internet hospitals should be an important part of the emergency system and a novel medical model, especially during major public health emergencies. These hospitals will be indispensable not only during the COVID-19 pandemic but also in future outbreaks of infection.

Acknowledgments

This study was funded by the Special Program for Novel Coronavirus Pneumonia of the Bureau of Science and Technology of Mianyang (2020YJKY002).

Conflicts of Interest

None declared.

References

1. Ross AGP, Crowe SM, Tyndall MW. Planning for the Next Global Pandemic. Int J Infect Dis 2015 Sep;38:89-94 [FREE Full text] [doi: 10.1016/j.ijid.2015.07.016] [Medline: 26253461]
2. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature 2008 Feb 21;451(7181):990-993 [FREE Full text] [doi: 10.1038/nature06536] [Medline: 18288193]
3. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb 29;395(10225):689-697 [FREE Full text] [doi: 10.1016/S0140-6736(20)30260-9] [Medline: 32014114]
4. Chen L, Liu W, Zhang Q, Xu K, Ye G, Wu W, et al. RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. Emerg Microbes Infect 2020;9(1):313-319 [FREE Full text] [doi: 10.1080/22221751.2020.1725399] [Medline: 32020836]
5. Eurosurveillance Editorial Team. Note from the editors: World Health Organization declares novel coronavirus (2019-nCoV) sixth public health emergency of international concern. Euro Surveill 2020 Feb;25(5):200131e [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.5.200131e] [Medline: 32019636]
6. Bassetti M, Vena A, Giacobbe DR. The novel Chinese coronavirus (2019-nCoV) infections: Challenges for fighting the storm. Eur J Clin Invest 2020 Mar;50(3):e13209 [FREE Full text] [doi: 10.1111/eci.13209] [Medline: 32003000]
7. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
8. Asmundson GJG, Taylor S. Coronaphobia: Fear and the 2019-nCoV outbreak. J Anxiety Disord 2020 Mar;70:102196 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102196] [Medline: 32078967]
9. Notice of the General Office of the National Health Commission on Doing a Good Job in Internet Diagnosis and Treatment Consulting Services in the Prevention and Control of the Epidemic. Article in Chinese. The General Office of the National Health Commission. 2020 Feb 07. URL: http://www.nhc.gov.cn/yzygj/s7653p/202002/ec5e345814e744398c2adef17b657fb8.shtml [accessed 2020-07-21]
10. Notice on launching online services to further strengthen epidemic prevention and control in Hubei. Article in Chinese. The State Council's Comprehensive Team for Joint Prevention and Control of New Coronavirus Pneumonia. 2020 Feb 26. URL: http://www.gov.cn/xinwen/2020-02/27/content_5483977.htm [accessed 2020-07-21]
11. Li X, Li H, He M, Deng L. The role of Internet hospitals in the prevention and control of new coronavirus pneumonia. Article in Chinese. Chin J Mod Nurs 2020;26(8):prepublished online [FREE Full text] [doi: 10.3760/cma.j.cn115682-20200203-00393]
12. Zhou Y, Tian D, Zhou D. Application of fever clinic online consultation for Internet-based hospitals in epidemic prevention and control. Article in Chinese. Chin J Hosp Admin 2929;36(4):286-289 [FREE Full text] [doi: 10.3760/cma.j.cn112225-20200211-00110]
13. Goulia P, Mantas C, Dimitroula D, Mantis D, Hyphantis T. General hospital staff worries, perceived sufficiency of information and associated psychological distress during the A/H1N1 influenza pandemic. BMC Infect Dis 2010 Nov 09;10:322. [doi: 10.1186/1471-2334-10-322] [Medline: 21062471]

14. Xu C, Zhang X, Wang Y. Mapping of Health Literacy and Social Panic Via Web Search Data During the COVID-19 Public Health Emergency: Infodemiological Study. J Med Internet Res 2020 Jul 02;22(7):e18831 [FREE Full text] [doi: 10.2196/18831] [Medline: 32540844]
15. Zhu Y, Fu K, Grépin KA, Liang H, Fung IC. Limited Early Warnings and Public Attention to Coronavirus Disease 2019 in China, January-February, 2020: A Longitudinal Cohort of Randomly Sampled Weibo Users. Disaster Med Public Health Prep 2020 Apr 03:1-4 [FREE Full text] [doi: 10.1017/dmp.2020.68] [Medline: 32241328]
16. Portnoy J, Waller M, Elliott T. Telemedicine in the Era of COVID-19. J Allergy Clin Immunol Pract 2020 May;8(5):1489-1491 [FREE Full text] [doi: 10.1016/j.jaip.2020.03.008] [Medline: 32220575]
17. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
18. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet Hospitals Help Prevent and Control the Epidemic of COVID-19 in China: Multicenter User Profiling Study. J Med Internet Res 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]
19. Tashkandi E, Zeeneldin A, AlAbdulwahab A, Elemam O, Elsamany S, Jastaniah W, et al. Virtual Management of Patients With Cancer During the COVID-19 Pandemic: Web-Based Questionnaire Study. J Med Internet Res 2020 Jun 24;22(6):e19691 [FREE Full text] [doi: 10.2196/19691] [Medline: 32501807]
20. Sun S, Yu K, Xie Z, Pan X. China empowers Internet hospital to fight against COVID-19. J Infect 2020 Jul;81(1):e67-e68 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.061] [Medline: 32251688]
21. Chan AKM, Nickson CP, Rudolph JW, Lee A, Joynt GM. Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. Anaesthesia 2020 Mar 30:online ahead of print [FREE Full text] [doi: 10.1111/anae.15057] [Medline: 32227594]

Abbreviations

COVID-19: coronavirus disease
GHQ-28: General Health Questionnaire-28
MERS: Middle East respiratory syndrome
MERS-CoV: Middle East respiratory syndrome coronavirus
SARS: severe acute respiratory syndrome
SARS-CoV: severe acute respiratory syndrome coronavirus
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach, G Fagherazzi; submitted 22.04.20; peer-reviewed by A Rovetta, Y Liu; comments to author 25.06.20; revised version received 08.07.20; accepted 18.07.20; published 04.08.20.

Please cite as:
Li L, Liu G, Xu W, Zhang Y, He M
Effects of Internet Hospital Consultations on Psychological Burdens and Disease Knowledge During the Early Outbreak of COVID-19 in China: Cross-Sectional Survey Study
J Med Internet Res 2020;22(8):e19551
URL: https://www.jmir.org/2020/8/e19551
doi:10.2196/19551
PMID:32687061

Original Paper

The Internet Hospital Plus Drug Delivery Platform for Health Management During the COVID-19 Pandemic: Observational Study

Liang Ding[1*], PhD; Qiuru She[2*], BS; Fengxian Chen[2], BS; Zitong Chen[2], BS; Meifang Jiang[2], BS; Huasi Huang[2], BS; Yujin Li[2], MS; Chaofeng Liao[2], BS

[1]Clinical Trial and Research Center, People's Hospital of Baoan Shenzhen, Shenzhen, China

[2]Department of Pharmacy, People's Hospital of Baoan Shenzhen, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Chaofeng Liao, BS
Department of Pharmacy, People's Hospital of Baoan Shenzhen
118 Longjing 2nd Road
Shenzhen, 518101
China
Phone: 86 18926480093
Email: bayylcf@126.com

Abstract

Background: Widespread access to the internet has boosted the emergence of online hospitals. A new outpatient service called "internet hospital plus drug delivery" (IHDD) has been developed in China, but little is known about this platform.

Objective: The aim of this study is to investigate the characteristics, acceptance, and initial impact of IHDD during the outbreak of COVID-19 in a tertiary hospital in South China

Methods: The total number of and detailed information on online prescriptions during the first 2 months after work resumption were obtained. Patients' gender, age, residence, associated prescription department, time of prescription, payment, and drug delivery region were included in the analysis.

Results: A total of 1380 prescriptions were picked up or delivered between March 2 and April 20, 2020. The largest group of patients were 36-59 years old (n=680, 49.3%), followed by the 18-35 years age category (n=573, 41.5%). In total, 39.4% (n=544) of the patients chose to get their medicine by self-pickup, while 60.6% (n=836) preferred to receive their medicine via drug delivery service. The top five online prescription departments were infectious diseases (n=572, 41.4%), nephrology (n=264, 19.1%), endocrinology (n=145, 10.5%), angiocardiopathy (n=107, 7.8%), and neurology (n=42, 3%). Of the 836 delivered prescriptions, 440 (52.6%) were sent to Guangdong Province (including 363 [43.4%] to Shenzhen), and 396 (47.4%) were sent to other provinces in China.

Conclusions: The IHDD platform is efficient and convenient for various types of patients during the COVID-19 crisis. Although offline visits are essential for patients with severe conditions, IHDD can help to relieve pressure on hospitals by reducing an influx of patients with mild symptoms. Further efforts need to be made to improve the quality and acceptance of IHDD, as well as to regulate and standardize the management of this novel service.

(J Med Internet Res 2020;22(8):e19678) doi:10.2196/19678

KEYWORDS

internet hospital; drug delivery; internet hospital plus drug delivery; IHDD; health management; COVID-19

Introduction

Background

As the third largest country in the world by area, China has 34 provincial regions, over 1.4 billion people, but only 10 million licensed physicians (2.2 for every 1000 people) in 2019, according to the National Bureau of Statistics of China [1]. Since the severe acute respiratory syndrome (SARS) epidemic in 2003, the Chinese government has been rebuilding the three-tier health care system. Today, the health care system in

China consists of community health centers (CHCs), and secondary and tertiary hospitals in urban areas; and village clinics, township health centers (THCs), and county hospitals in rural areas. CHCs, village clinics, and THCs are considered core primary care providers and are expected to provide affordable first contact care while secondary and tertiary care facilities provide specialist referral services [2]. However, with no gatekeeping in the primary health care system, patients can freely choose their provider at any health facility. Although many disorders could be treated by primary care providers conveniently and at a relatively low price, many patients are unwilling to see these providers owing to their lack of confidence in the health professionals' skills and the quality of health care provided. They tend to go to high-level hospitals even for mild symptoms, effectively overcrowding those hospitals [3]. On the other hand, skilled doctors are unwilling to work at the community level and in remote rural areas for financial and professional reasons. These two problems have led to countless transprovincial patients, resulting in numerous additional economic and time costs [4].

The rapid increase in internet users (from 22.7% to 59.6% of the population between 2008 and 2018) [5] offers the Chinese government a new alternative to address these health care problems. On October 25, 2014, the first officially approved "internet hospital" went online in Guangdong Province. In the beginning, the internet hospital consisted of four clinics operated by doctors from the Second People's Hospital of Guangdong Province, an online platform operated by a medical technology company, and a network of medical consulting facilities based in clinics in rural villages, CHCs, and large pharmacy chain stores [4]. The inchoate platform usually required onsite equipment (computers, cameras, speakers, and cable network). Patients needed to go to a medical consultation facility near their home and meet through the internet with the doctor based in a top-level hospital in a big city. With the widespread adoption of smartphones and tablet computers, and the ever-increasing popularity of mobile internet communication, a mobile health (mHealth) care model was made accessible to the public. mHealth allows patients to access information, assessments, and treatments in a timely manner. In addition, it empowers doctors with another way to connect with their patients and to practice without geographical limitations [6]. Therefore, the extra costs of health care, such as those associated with travel, time, and doctor consultations, can be dramatically reduced.

During the outbreak of coronavirus disease (COVID-19) [7], the Chinese government adopted a series of administrative measures to stop the spread of the epidemic [8], including requiring domestic internet hospitals to vigorously carry out remote medical services [9]. Although the convenience and ubiquity of internet hospitals makes them a promising avenue through which to overcome geographical limitations between patients and doctors, there is still a "social distance" barrier between the patients and their prescription medicines. In order to solve this problem, many hospitals intend to cooperate with delivery companies to build a partnership for drug delivery [10]. This bundled approach could offer an omnichannel solution that can help people on their path to urgently needed health care and medicine during the epidemic.

Objective

To explore the advantages of the IHDD model for health management during public health emergencies, we analyzed the prescriptions of online outpatients at the People's Hospital of Baoan Shenzhen in Shenzhen City, Guangdong. Data from the first 2 months after work resumption were collected to reveal the characteristics, acceptance, and initial impact of the new bundled approach.

Methods

Data Sources

The total number of online prescriptions and detailed information on them were obtained automatically from the hospital information system (DTHealth, V7.0) of the People's Hospital of Baoan Shenzhen. Data from March 2 to April 20, 2020 were collected. Patients' gender, age, residence, associated prescription departments, time of prescription, payment, and drug delivery region were included in the analysis. GraphPad Prism 8.0.2 (GraphPad Software) was used to summarize and analyze the data.

Ethics Statement

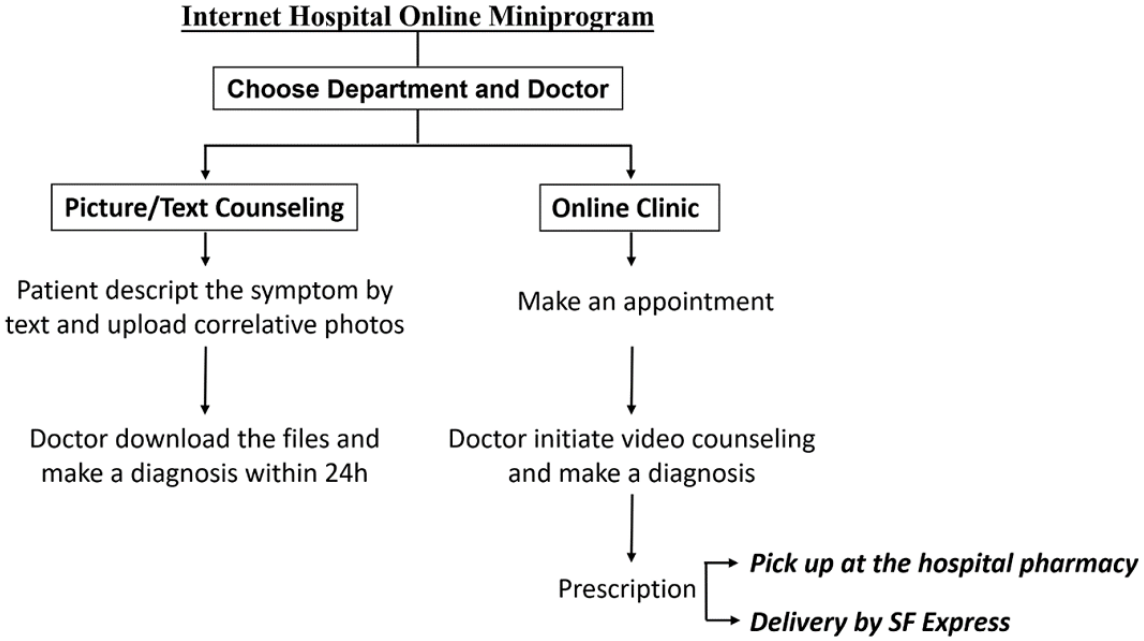
Ethics approval was obtained from the medical research ethics committee of the People's Hospital of Baoan Shenzhen before the start of the study.

Results

The Internet Hospital Workflow

Figure 1 shows the online consultation workflow of the internet hospital. The patients chose a department and doctor by self-assessment using the hospital miniprogram. There was no prescription option for picture/text counseling patients. For the online clinic, a confirmation would be sent by text message once the online clinic appointment was made. Another message would be sent to the patient 3 minutes before the video counseling session began to remind them to open the hospital miniprogram in time. The doctor then initiated a video consultation with the patient and made an online prescription, if necessary, based on the diagnosis of the patient. At the payment step, the patient could choose self-pickup of medication at the hospital's pharmacy or delivery to an assigned place.

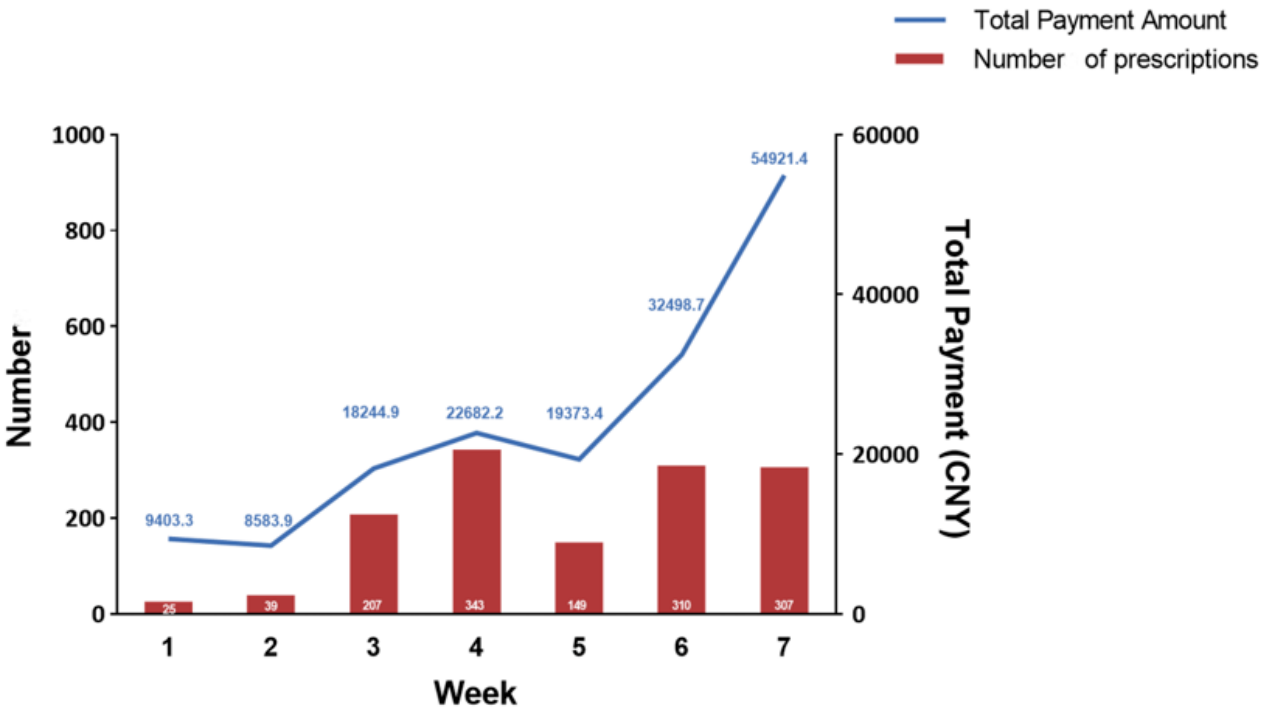
Figure 1. Online workflow of the internet hospital.



Number and Payment Amount of Online Prescriptions

A total of 1380 prescriptions were picked up or delivered between March 2 and April 20, 2020. The weekly number and payment amount pertaining to these prescriptions are summarized in Figure 2. There was an increase in the use of the online prescription service. The number and total payments of the 7th week significantly increased, by 11.3 and 4.8 times, respectively, compared with the first week.

Figure 2. The number of and total payments pertaining to online prescriptions.



Patient Characteristics

There was no sex-based differences among the patients who received prescriptions (Table 1). The patients were divided into four groups according to their age: 1-17 years old, 18-35 years old, 36-59 years old, and ≥60 years old. The largest group of patients were 36-59 years old (n=680, 49.3%), followed by those who were 18-35 years old (n=573, 41.5%). In total, 65.7% (n=907) of the patients were local residents, 5.6% (n=77) were from Guangdong cities other than Shenzhen, 28.7% (n=396) were from other provinces in China. Less than half the patients (n=544, 39.4%) chose to receive their medicine by self-pickup, while 60.6% (n=836) preferred to get their medicine by drug delivery service.

Table 1. Baseline characteristics of patients (N=1380).

Characteristic	Value
Sex, n (%)	
Male	693 (50.2)
Female	687 (49.8)
Age (years)	
Median (range)	38 (1-93)
Group, n (%)	
1-17	12 (0.9)
18-35	573 (41.5)
36-59	680 (49.3)
≥60	115 (8.3)
Residence, n (%)	
Local (Shenzhen City)	907 (65.7)
Other cities in Guangdong Province	77 (5.6)
Other provinces	396 (28.7)
Access to medicine, n (%)	
Delivery	836 (60.6)
Self-pickup	544 (39.4)

Distribution of Online Prescriptions

The top five online prescription departments were infectious diseases (n=572, 41.4%), nephrology (n=264, 19.1%), endocrinology (n=145, 10.5%), angiocardiopathy (n=107, 7.8%), and neurology (n=42, 3%). The majority of infectious disease and neurology patients chose drug delivery services, while most patients with other diagnoses preferred to pick up their medication (Table 2).

Table 2. Delivery/self-pickup preference of online prescription patients (N=1380).

Department	Delivery (n=836), n (%)	Self-pickup (n=544), n (%)
Infectious disease (n=572)	551 (96.3)	21 (3.7)
Nephrology (n=264)	86 (32.6)	178 (67.4)
Endocrinology (n=145)	48 (33.1)	97 (66.9)
Angiocardiopathy (n=107)	37 (34.6)	70 (65.4)
Neurology (n=42)	25 (59.5)	17 (40.5)

Drug Delivery Details of Online Prescriptions

For the 836 delivered prescriptions, 440 (52.6%) were sent to Guangdong Province (including 363 [43.4%] to Shenzhen) and 396 (47.4%) were sent to other provinces in China. The top 10 provinces for out-of-province deliveries were Heilongjiang, Hubei, Guangxi, Shandong, Jiangsu, Hunan, Shanxi, Henan, Anhui, and Jiangxi (Table 3). Most of them are located in the northeast, eastern, and central parts of China (Figure 3 and Table 3). The top 10 delivered medicines are listed in Table 4. Most of the medicines were used to treat infectious or chronic disease, which was consistent with the distribution of online prescription.

Table 3. Out-of-province delivery details based on the geographical regions of China.

Region	Patients (n=396), n (%)
Northeast	107 (27.0)
Eastern	102 (25.8)
Central	82 (20.7)
Southern	35 (8.8)
Northwest	35 (8.8)
Northern	21 (5.3)
Southwest	14 (3.5)
Top 10 provinces	
Heilongjiang	94 (23.7)
Hubei	36 (9.1)
Guangxi	33 (8.3)
Shandong	27 (6.8)
Jiangsu	24 (6.1)
Hunan	24 (6.1)
Shanxi	24 (6.1)
Henan	22 (5.6)
Anhui	14 (3.5)
Jiangxi	14 (3.5)

Figure 3. Regional distribution of prescription deliveries.

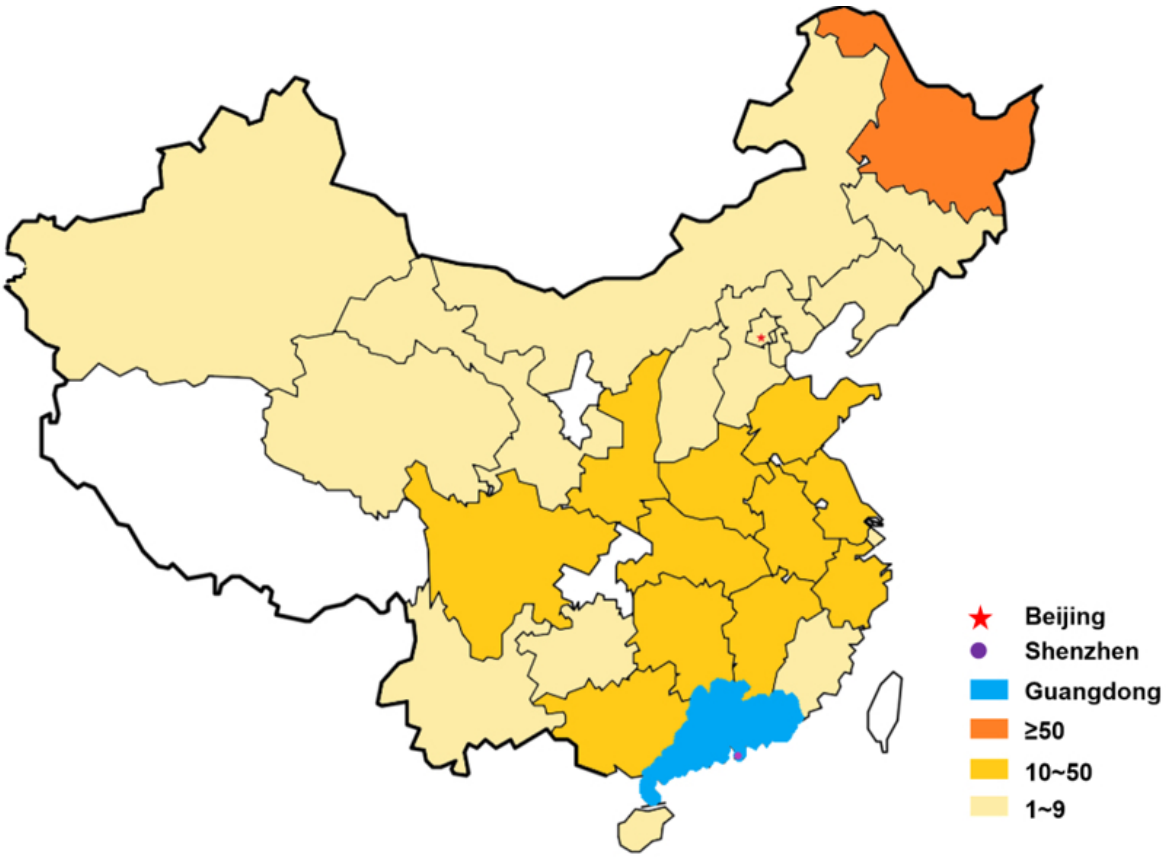


Table 4. Top 10 delivered medicines.

Name	Formulation	Manufacturer
Entecavir	Dispersible tablet	ChiaTai TianQing
Metoprolol succinate	Sustained release tablet	AstraZeneca AB
Entecavir	Dispersible tablet	Dawnrays
Metformin hydrochloride	Tablet	Bristol-Myers Squibb
Tenofovir disoproxil fumarate	Tablet	Brilliant
Atorvastatin calcium	Tablet	Pfizer
Nifedipine	Controlled release tablet	Bayer
Atorvastatin calcium	Tablet	JiaLin
Mecobalamin	Tablet	Desano
Amlodipine besylate	Tablet	Pfizer

Discussion

Principal Findings

We conducted a pilot evaluation of the IHDD health care model in a tertiary hospital in Shenzhen during the first 2 months after work resumption. The unbalanced distribution of medical resources and the outbreak of COVID-19 [11,12] promoted the growth and exploration of more convenient internet-based medical practices [13], especially in the well-developed southern and southeastern parts of China, where people use the internet more often for medical purposes [14]. Despite the advantages of internet hospitals, access to medication remained an obstacle that may have discouraged people from using this platform. The traditional internet hospital required patients to go to the hospital or drugstore for medicines, which could cause more infections during an epidemic. The drug pickup process may increase risk of acute infectious disease, particularly for patients with suppressed immune systems or disabilities, which can then lead to severe health deterioration. On the other hand, out-of-city and out-of-province patients could have problems finding the exact prescription medications they need, as those medicines might be not available at their local hospitals and drugstores. In January 2019, the General Office of the State Council of the People's Republic of China implemented the National Centralized Drug Purchasing (NCDP) and Using pilot program and selected 11 cities (including Shenzhen) in mainland China to carry out the “4+7” City-Drug-Volume-Based-Purchasing pilot project [15]. As the “frontier” of Chinese prescription medicine reform, prescription medicines that were made by the doctors of Shenzhen’s hospitals may be more affordable. Therefore, the development of IHDD could enable patients across the country to access online prescription medication in a secure and convenient way.

The Chinese government has encouraged internet hospitals to join the epidemic prevention and control efforts of the COVID-19 outbreak [16]. On March 15, 2020, the first professional standard, “Specification for Online Consultation Service for Infectious Disease Epidemic Situation” was published on the national group standard information platform of China, requiring that internet hospitals provide 24/7 online services in response to the epidemic [17]. The internet hospital of the People’s Hospital of Baoan Shenzhen has been officially online since March 2, 2020. From opening to April 20, 2020, it saw a total of 8638 patients, an average of 176 per day, with 5877 in picture/text counseling and 2761 in online clinic video counseling (including 1381 that did not result in a prescription). Most of the picture/text consultations were prehospital services such as psychological counseling and medical education. The number and payment amounts of online prescriptions increased progressively from the first investigated week to the 7th one (Figure 2), which indicates increased acceptance of IHDD. The drop in prescription numbers during the 5th week might be the result of two factors: Tomb-Sweeping holiday and a lack of antiviral medicine.

Most prescription patients were between the ages of 18 and 60 years, had no time for onsite visitations, and had greater access to new medical platforms. At present, health authorities and the government have warned older people that they are at a higher risk of more serious and possibly fatal illness associated with COVID-19. Moreover, the global recommendation for older populations includes social isolation, which involves staying at home and avoiding contact with other people for an extended period of time [18]. Our data show that only 8.3% of IHDD users were ≥60 years (Table 1). This may be due to differences in public acceptance. Older populations usually take more time to become familiarized with the operations of IHDD.

The stay-at-home order constrained people from going outside, which increased difficulties associated with health management, especially chronic disease management. Medical professionals at hospitals with fever clinics are required to participate in COVID-19 prevention, control, and treatment, which has reduced their concentration on other diseases. In fact, the management of chronic disease has become a crucial issue in cities with large outbreaks of COVID-19 [19]. The largest number of internet hospital prescriptions came from the department of infectious diseases, which includes acute and chronic viral hepatitis, fatty liver, alcoholic hepatitis, drug-induced liver damage, autoimmune liver disease, and genetic and metabolic liver disease. The second largest group was from nephrology, followed by endocrinology. Patients with chronic liver disease, kidney disease, or diabetes could easily renew their prescriptions and receive their medicine by IHDD.

A notable finding was that most patients with an infectious disease chose to receive their medicine by delivery (96.3%), whereas most patients with other diseases selected self-pickup (Table 2). This can be explained by the need for special storage of some medicines (eg, recombinant human erythropoietin for kidney disease or insulin for diabetes).

Established in 1984, the People's Hospital of Baoan Shenzhen is also the Eighth People's Hospital of Shenzhen, the Shenzhen Baoan Affiliated Hospital of Southern Medical University, and the Second Affiliated Hospital of Shenzhen University. It was recognized as a Grade A Tertiary Hospital by the Guangdong Health Department in 2012. As an important source of health care providers in Shenzhen, the hospital holds a great reputation in both basic clinical practice and diverse clinical research and training. The IHDD platform enabled patients all over the country to obtain access to its health professionals and quality-assured medicine (Figure 3). The hospital even offered medical service to the patients in Wuhan, the epicenter of COVID-19. In fact, online prescription medicines that were delivered to Hubei Province accounted for the second largest number of all out-of-province deliveries (Table 3). The top 10 provinces for IHDD delivery were located in the northeast, eastern, and central parts of China. One of the reasons is that these areas are relatively economically developed regions, and their residents are highly educated, which ensures they have a better understanding of the benefits of IHDD.

The top 10 delivered medicines were used for the treatment of hepatitis, hypertension, hyperlipidemia, diabetes, climacteric symptoms, etc (Table 4). This finding was consistent with the distribution of prescription departments. In fact, 7 of these departments were enrolled in the National Essential Drugs of China program. Within these 7, 3 belonged to the "4+7" NCDP catalog. The affordability and quality of these medications were guaranteed by the government.

One of the concerns of IHDD is the safety and security of drug delivery. Therefore, delivery service companies with good reputations were chosen by the hospitals as partners. As the industry leader, SF Express is the first logistics company to cooperate with both pharmaceutical providers and hospitals. It has a long history of ambient temperature and cold chain medicine transport. Moreover, it offers real-time package tracking and zero-touch delivery. Once the patients place their order, a tracking number is sent by text message to their cell phone. The processing information of the medicine is updated and sent to patients automatically. Couriers place the medicines into the customer-assigned delivery lockers, which are usually near the patient's residence, so the patient can access them using a random cipher. This process could effectively reduce viral transmission while simultaneously providing convenience for patients. At present, an advanced cooperation model is under exploration: the pharmacist-audited prescription will be sent to the manufacturer directly, and the medicines will be delivered from manufacturer's stock house. This will reduce substantial pressure on hospital drugstores and cut transportation expenses.

Future Prospects

Although in-person visits are essential when the patients are experiencing severe symptoms, IHDD can help to relieve pressure on hospitals by reducing the influx of mild cases. To make better use of IHDD during and after the current epidemic, more effort is needed. Simple and clear instructions are necessary to improve its acceptance by older people. Financial support, like adding medical insurance to payment methods, can also promote adoption by the public. The new hospital-manufacturer-patient transport model should be further evaluated. Moreover, official regulations are required in terms of standardization of the operational process and management of IHDD.

Conclusion

The pandemic of COVID-19 has clearly entered a new stage with rapid spread to countries outside China, becoming a global threat. This once-in-a-century pandemic might permanently change people's lifestyle, especially when it comes to health management. In our study, IHDD has been proven to be efficient and convenient for many types of patients during the crisis. The widespread use of this platform can help to reduce person-to-person transmission as well as the infection risk of patients with chronic diseases or disability.

Acknowledgments

This research was supported in part by the Key Laboratory of Emergency and Trauma (Hainan Medical University), the Ministry of Education (KLET-201908), and th eScience, Technology and Innovation Commission of Shenzhen Municipality (731144920168). We gratefully acknowledge the help of the Information Technology Center, People's Hospital of Baoan Shenzhen. We also thank Mr Zhiyong Zhang for assistance with data collection.

Authors' Contributions

LD, QS, and CL designed the study. QS, FC, and YL searched for relevant national and regional information. MJ and HH gathered data. ZC carried out the data analyses. LD wrote the first draft of the paper, with revisions from QS and CL. All authors contributed to revisions and approved the final version.

Conflicts of Interest

None declared.

References

1. Public Health. National Bureau of Statistics of China. 2020. URL: http://data.stats.gov.cn/english/easyquery.htm?cn=C01 [accessed 2020-04-20]
2. Yip WC, Hsiao W, Meng Q, Chen W, Sun X. Realignment of incentives for health-care providers in China. The Lancet 2010 Mar 27;375(9720):1120-1130. [doi: 10.1016/S0140-6736(10)60063-3] [Medline: 20346818]
3. Wu D, Hesketh T, Shu H, Lian W, Tang W, Tian J. Description of an online hospital platform, China. Bull World Health Organ 2019 Aug 01;97(8):578-579 [FREE Full text] [doi: 10.2471/BLT.18.226936] [Medline: 31384077]
4. Tu J, Wang C, Wu S. The internet hospital: an emerging innovation in China. The Lancet Global Health 2015 Aug;3(8):e445-e446 [FREE Full text] [doi: 10.1016/S2214-109X(15)00042-X] [Medline: 26187488]
5. Work dynamics. Ministry of Industry and Information Technology of the People's Republic of China. URL: http://www.miit.gov.cn/n1146290/n1146402/index.html [accessed 2020-04-20]
6. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
7. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr 26;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
8. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
9. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet Hospitals Help Prevent and Control the Epidemic of COVID-19 in China: Multicenter User Profiling Study. J Med Internet Res 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]
10. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine During the COVID-19 Pandemic: Experiences From Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
11. Mahase E. WHO declares pandemic because of "alarming levels" of spread, severity, and inaction. BMJ 2020 Mar 12;368:m1036. [doi: 10.1136/bmj.m1036] [Medline: 32165426]
12. Bedford J, Enria D, Giesecke J, Heymann DL, Ihekweazu C, Kobinger G, et al. COVID-19: towards controlling of a pandemic. The Lancet 2020 Mar 17;395(10229):1015-1018. [doi: 10.1016/S0140-6736(20)30673-5] [Medline: 32197103]
13. Tanne JH, Hayasaki E, Zastrow M, Pulla P, Smith P, Rada AG. Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. BMJ 2020 Mar 18;368:m1090. [doi: 10.1136/bmj.m1090] [Medline: 32188598]
14. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet Hospitals in China: Cross-Sectional Survey. J Med Internet Res 2017 Jul 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
15. Tang M, He J, Chen M, Cong L, Xu Y, Yang Y, et al. "4+7" city drug volume-based purchasing and using pilot program in China and its impact. Drug Discov Ther 2019;13(6):365-369 [FREE Full text] [doi: 10.5582/ddt.2019.01093] [Medline: 31956236]
16. Notice on accomplishing online consultation services in the epidemic prevention and control (Medical Letter 2020 No. 112). National Health Commission of the People's Republic of China. 2020 Feb 7. URL: http://www.nhc.gov.cn/yzygj/s7652m/202002/32c3e98988894fa18280e4543d2710c7.shtml [accessed 2020-04-20]
17. Specification for online consultation service for infectious disease epidemic situation. Zhejiang Digital Economy Association. 2020 Mar 15. URL: http://www.ttbz.org.cn/StandardManage/Detail/33876/ [accessed 2020-04-20]
18. Brooke J, Jackson D. Older people and COVID-19: Isolation, risk and ageism. J Clin Nurs 2020 Jul;29(13-14):2044-2046. [doi: 10.1111/jocn.15274] [Medline: 32239784]
19. Shahid Z, Kalayanamitra R, McClafferty B, Kepko D, Ramgobin D, Patel R, et al. COVID-19 and Older Adults: What We Know. J Am Geriatr Soc 2020 May;68(5):926-929 [FREE Full text] [doi: 10.1111/jgs.16472] [Medline: 32255507]

Abbreviations

CHC: community health center
COVID-19: coronavirus disease
IHDD: internet hospital plus drug delivery
mHealth: mobile health
NCDP: National Centralized Drug Purchasing
SARS: severe acute respiratory syndrome
THC: township health center

Edited by G Eysenbach; submitted 29.04.20; peer-reviewed by H Liu, F Zhang, X Xie, B Foroutan; comments to author 12.06.20; revised version received 22.06.20; accepted 23.07.20; published 06.08.20.

Please cite as:
Ding L, She Q, Chen F, Chen Z, Jiang M, Huang H, Li Y, Liao C
The Internet Hospital Plus Drug Delivery Platform for Health Management During the COVID-19 Pandemic: Observational Study
J Med Internet Res 2020;22(8):e19678
URL: http://www.jmir.org/2020/8/e19678/
doi:10.2196/19678
PMID:32716892

Original Paper

What Media Helps, What Media Hurts: A Mixed Methods Survey Study of Coping with COVID-19 Using the Media Repertoire Framework and the Appraisal Theory of Stress

Amber Pahayahay[1], BSN, MPH; Najmeh Khalili-Mahani[2,3], MEng, PhD

[1]School of Public Health and Services, University of Waterloo, Waterloo, ON, Canada

[2]PERFORM Centre, Concordia University, Montreal, QC, Canada

[3]McGill Centre for Integrative Neuroscience, Montreal Neurological Institute, McGill University, Montreal, QC, Canada

Corresponding Author:
Najmeh Khalili-Mahani, MEng, PhD
PERFORM Centre
Concordia University
PC 2.207 PERFORM Centre
7200 Rue Sherbrooke O
Montreal, QC, H4B 1R6
Canada
Phone: 1 514 848 2424 ext 5370
Email: najmeh.khalili-mahani@concordia.ca

Abstract

Background: Social and physical distancing in response to the coronavirus disease (COVID-19) pandemic has made screen-mediated information and communication technologies (media) indispensable. Whether an increase in screen use is a source of or a relief for stress remains to be seen.

Objective: In the immediate aftermath of the COVID-19 lockdowns, we investigated the relation between subjective stress and changes in the pattern of media use. Based on Lazarus's transactional model of appraisal and coping, and building on an earlier similar survey, we hypothesize that individual differences in the appraisal of media predict variations in approach or avoidance of media for coping with COVID-19 stress.

Methods: Between March 20 and April 20, 2020, a brief snowball survey entitled: "What media helps, what media hurts: coping with COVID19 through screens" was distributed via Concordia University's mailing lists and social media (PERFORM Centre, EngAGE Centre, and Media Health Lab). Using a media repertoire method, we asked questions about preferences, changes in use, and personal appraisal of media experiences (approach, avoid, and ignore) as a result of the COVID-19 pandemic and investigated interindividual differences in media use by factors such as subjective stress, age, gender, and self-reported mental health.

Results: More than 90% of the survey respondents were in Canada and the east coast of the United States. From 685 completed responses, 169 respondents were "very stressed" and 452 were "slightly worried" about the pandemic. COVID-19 stress led to increased use of Facebook (χ^2_3=11.76, P=.008), television (χ^2_3=12.40, P=.006), YouTube (χ^2_3=8.577, P=.04), and streaming services such as Netflix (χ^2_3=10.71, P=.01). Respondents who considered their mental health "not good" were twice as likely to prefer streaming services as a coping tool for self-isolation. Women and nonbinary respondents were twice as likely than men to pick social media for coping. Individuals younger than 35 years were 3 times more likely to pick computer games, and individuals older than 55 years were more likely to pick network television or print media. Gender affected the appraisal of media (less in men than others) in terms of avoid ($F_{1,637}$=5.84, P=.02) and approach scores ($F_{1,637}$=14.31, P<.001). Subjective mental health affected the ignore score (less in those who said "good" than others; $F_{1,637}$=13.88, P<.001). The appraisal score and use increase explained variations in worrying about physical and mental health stress due to increased screen time. A qualitative analysis of open-ended questions revealed that media (especially social networks) were important for coping if they provided support and connection through the dissemination of factual and positive information while avoiding the overflow of sensational and false news.

Conclusions: The relationship between appraisal of media's positive and negative facets vary with demographic differences in mental health resiliency. The media repertoire approach is an important tool in studies that focus on assessing the benefits and harms of screen overuse in different populations, especially in the context of the COVID-19 pandemic.

(J Med Internet Res 2020;22(8):e20186) doi:10.2196/20186

KEYWORDS

Netflix; social network; stress; COVID-19; information and communication technologies; survey; media; coping; infodemic; infodemiology

Introduction

Background

The necessity of social and physical distancing in response to the coronavirus disease (COVID-19) pandemic has made screen-mediated information and communication technologies (media) even more indispensable than before, raising concerns about harms or benefits of screen dependency or screen stress due to overexposure. Previously, we have examined the relation between subjective and quantitative measures of screen addiction and stress, and showed that, despite the heterogeneity of the patterns of screen use and types of stressors, a robust correlation existed between higher emotional and perceived stress, and higher likelihood of screen addiction, especially linked to social networking and entertainment-related activities [1]. However, we could not answer the question whether screen addiction caused higher stress levels or if higher stress motivated an escape into screens for coping. The unprecedented occasion of this global stressor, the COVID-19 pandemic, allows us to address this question.

There is a growing concern about the potential adverse effects of excessive screen time on emotional and physical health. To name a few harms, there is stress caused by an abundance of catastrophic news [2]; increased sedentary behavior and obesity [3,4]; sleep disorders [5]; and addiction to social media [6], computer games [7], online gambling [8], etc. Nevertheless, the debate about the directional relationship between stress and compulsive screen use is nuanced. A 2014 review by Ryan et al [6] indicated that the risks of developing an addiction to social networks may be related to use and gratification factors linked to relationship maintenance, passing time, entertainment, and companionship. In a 2018 meta-analysis of 56 independent samples (comprising >27,800 individuals), Marino et al [9] showed that the problematic use of Facebook was associated with internal motives such as coping and information-seeking, and external motives such as socialization and conformity, albeit with important moderating effects related to age and geographic location. The two main reasons for problematic Facebook use were related to the motives of reducing negative moods and meeting one's needs to cope or pass time [9]. Brailovskaia et al [10] investigated the link between daily stress and depression in samples of problematic Facebook users from Germany (N=531) and the United States (N=909). Individuals with depressive symptoms reported higher daily stress and higher Facebook use. Although they acknowledged the short-term benefits of using Facebook to cope with depression, the authors warned that this positive effect could lead to long-term maladaptation due to addiction [10].

The double-edged nature of social media harms or benefits has long been considered in adolescents. Tsitsika et al [11] evaluated a cluster sample of more than 10,000 adolescents in grade 9 or 10 from 600 classrooms from 6 European countries and found that, in younger adolescents, the heavy use of social networking was associated with lower academic performance, higher internalizing of problems, and loss of physical activity; conversely, however, in older adolescents, the same social networking use was positively correlated with social competence. In a recent qualitative study of over 100 individuals with stressful experiences, Lee et al [12] experimentally manipulated the direction of conversations in an online-support context and showed that, whereas conversations that focused on reconstruing the stress experience from a broader perspective had a helpful impact, those focusing on recounting the personal experience were likely to add emotional stress for participants. A content analysis of over 8 million online conversations (in Korean) about the Middle East respiratory syndrome (MERS) outbreak in South Korea showed that, although expressions of negative affect (anxiety and fear) were prevalent in online social media and discussion boards, informative news content was more likely to bring out expressions of positive (calm and composed) MERS emotions [13].

Mediating screens are designed to be beneficial though. The role of online social networks in facilitating information- and support-seeking, especially for mental health support, is important in understanding their relation to stress and screen addiction. In 2014, Griffith and Szabo [14] found that the most addictive forms of internet activity among a sample of 111 college-aged respondents were social networking (84%), email and chatting (69%), and watching videos (35%), and that each of those addictions satisfied a specific need of the user and helped them to improve the quality of their life. In a follow-up survey study of 1057 internet users aged 16-70 years, this team reported that the greatest source of internet dependency was information- and news-seeking, and that more than 86% of screen addicts believed that it improved their life quality [15]. In 2017, Utz and Breuer [16] reported the results of a 6-wave longitudinal study (over 3 years), with a final attrition of 1330 representative internet users in the Netherlands, and showed that in all waves the users of social networks reported higher levels of online social support than nonusers, especially for seeking advice. In a 2019 survey study of more than 1000 young Irish adults (aged 18-25 years), Petrorious et al [17] showed that more than 82% relied on online search and more than 57% on online medical support sites for seeking mental health care from reliable sources.

In our previous cross-sectional snowball survey of 650 respondents (aged 18-80 years), we showed that more than 95% considered the most important need for media technologies to be for communication and information-seeking, which was independent from perceived psychosocial stress or emotional factors such as irritability, anxiety, sadness, lack of motivation, and anger [1]. In this current study conducted in the immediate aftermath of COVID-19 shut downs in North America, we asked a subset of questions from our previous survey to specifically investigate: would higher levels of subjective COVID-19 stress predict increased media use, would subjective stress due to COVID-19 increase social and entertainment media use more than other uses, to what extent would individual differences (demographics, health, and perceived COVID-19 stress) predict variations in respondents' appraisal of media as a beneficial coping strategy, and what factors would predict individual's concerns about the health risks of increased screen time?

Theoretical Framework

Stress, coping, and media are complex multifaceted constructs, and it is important to consider the working definitions that frame this work.

Stress

First, the term *Stress* is one of the most frequently used (or misused) terms in today's health discussions, but it is not understood or even felt similarly across different cultures [18]. There are various standard stress questionnaires (such as the perceived stress scale) that help provide a quantitative index of stress, and there are psychophysiological experiments that measure the embodied experience of stress. However, the aim of this study is not to quantify different dimensions of stress, but to assess the respondent's subjective experience of stress. For the purpose of this research, we rely on Mason's [19] definition of stress, which suggests that the experience of conditions of novelty, unpredictability, threat to self, and sense of control will reproducibly trigger a neurophysiological stress response [20]. COVID-19 is a stressor because the rapid global disruptions caused by COVID-19 lockdowns are *novel* and unprecedented to the lives of many in North America and Europe (where most of our data is collected from). COVID-19 has created an unpredictable condition. Several levels of *unpredictable* outcomes are prevalent: how and when will this end, and what will be the human or financial toll? COVID-19 is perceived as *threat*ening to every aspect of life, financially, socially, and even physically (as the illness seems to be grave). The public health measures to control the spread of the virus, as well as the unknown nature of the virus' mechanisms of spread and immunity challenge every sense of control. Besides restrictions about work, social distancing, and travel, how this virus will mutate or end is outside our locus of *control*.

Coping

Second, similar to stress, the term *coping* is also imprecise [21]. As a psychophysiological response, stress is a complex phenomenon [22], and individual differences in appraisal and coping determine the behavioral approaches that alter an individual's experience of stress [23]. When we are casually talking about stress, we are often referring to a challenge that forces us to cope, and coping can be influenced by a myriad of personal, social, and environmental factors that vary with nature [24] and culture [25]. Coping is also a context dependent experience, and there are currently numerous survey studies that are trying to understand differences in coping and resilience [26].

Our study is only focusing on whether using media can help cope with COVID-19 stress, and whether patterns of media use vary with factors such as age, gender, and self-assessment of mental and physical health. After the lockdown, various behavioral and interpersonal resources that would generally be available for coping with stress (for example, exercise and fitness centers, parks and recreational areas, one-to-one or group therapy activities, social support networks, or even medical doctors) became unavailable. *Screens* are currently the only safe (from contagion) tool for coping with social isolation, interruption of work, learning and finding critical information, as well as distracting oneself from boredom and anxiety.

Therefore, we have narrowed down the question of coping to a set of factors from Marino et al's [9] findings of the internal and external motives for using social networks for coping through information-seeking, conforming, socializing, enhancing mood, and passing time.

Media Appraisal

Third, we address the complexity of the relationship between coping and stress by referring to Lazarus and Folkman's Theory of Stress Appraisal and Coping. Briefly, the Appraisal Theory postulates that when confronted with a stressor, individuals engage in a primary appraisal of its relevance, potential benefits, and potential dangers. Whether they find it beneficial or dangerous, they will then enter the second phase of appraisal to identify resources that they have or resources that they need to recruit to meet the challenges of the stressor (see the first supplemental figure in Multimedia Appendix 1). Depending on an individual's abilities, personality, or the particularity of the circumstances, the process of appraisal is mediated by cognition- or emotion-based behaviors that motivate and shape an individual's approach to, or avoidance of, different response strategies (eg, cognitive- or emotion-based). This process of appraisal is repeated recursively until an individual finds a solution (or fails) [23]. We have previously proposed a conceptual mixed methods framework for studying interactions between media (such as serious games) and stress [27]. We repeat this iterative process here by recursively asking questions about sources of actual, perceived, or anticipated stress while investigating individuals' differences in use and preferences.

Media Repertoire Approach

Similar to our previous study [1], we use a repertoire-oriented framework that emphasizes the interrelation between different available technologies and the factors that influence an individual's choice in the amount of different media or content use [28]. Research into media use often involves assessing the amount and the type of media used by the public or identifying the reasons for, and meanings of, using a specific media type within a specific context. The specific context of this study was "coping with COVID-19 disruptions," and we were interested

in comparing the prevalence of different types of media-related activities in the coping process.

The media repertoire framework includes a mixed methods analysis of large-scale surveys of use together with more qualitative research into individual preferences to bridge between the patterns of different media type use and differences in the meaning and affordance of each media for a particular person (or subgroups of people) within social contexts [29].

We defined demographics, extent of perceived stress, and beliefs about one's mental health as predictors of variations in media use and media appraisal (in the context of serving as coping tools, as well as in relation to their potential risks to mental and physical health). Hence, we encourage readers to keep these working definitions in mind for the interpretation of our findings.

Methods

Data Collection

Within days of the announcement of the lockdown in the province of Quebec (March 13, 2020), we deployed a brief multifactorial 16-item questionnaire using the SurveyMonkey platform [30]. A brief advertisement was distributed via PERFORM's mailing list, as well as various social media (Twitter and Facebook) accounts.

> *Coping with COVID-19: What media helps, what media hurts?*
>
> *Please join us in a quick anonymous survey to assess which kinds of media and information technologies matter most in these extraordinary times? Do they stress us, or help us cope better?*

The survey was available only in English, as we aimed to reach an international community that could communicate via a common language. Because it was important to deploy the survey in the early phase of the pandemic, we made the length of the survey short enough to not take more than 5 minutes to ensure high completion rate. (We attained >95% completion.)

To compensate for the brevity of the survey, which prevented us from quantitative assessment of stress and coping, we included two open-answer boxes and asked respondents to offer more details about three specific questions: how the pandemic was disrupting their life, what other coping methods than those we listed would they use, and how they envisioned a strategy for media to become a useful tool for coping.

The sample size calculation was based on a margin of error and confidence level rather than prevalence or expected effect sizes. With a 5% margin of error and a confidence level of 95%, a minimum sample size of 384 was estimated to be sufficient to reveal differences in an average response to each survey question. The survey was advertised through email lists, the PERFORM Centre website, Facebook, and Twitter, as well as through the social media of EngAGE Centre for studies in aging, and the Media-Health.ca website and social media. The distribution lists alone contained at least 10,000 people, thus, obtaining the necessary sample size could be achieved even with a conservative completion rate of 5-10%.

Dependent and Independent Variables

Table 1 summarizes the dependent and independent variables and corresponding questions that were tested in this survey.

Table 1. List of variables.

Variable	Questions	Responses
Subjective *COVID-19*[a]*stress* (IV[b], categorical)	• Which one of these statements describe how you feel about the COVID-19 pandemic?	I am: • Very stressed • Slightly worried • Not worried at all • Excited about it
Demographics (IV, categorical)	• *Age (years)* • *Gender*	• <25; 25-34; 35-54; 55-65; >65 • Man, woman, other
Self-assessed *mental/physical health* (IV, categorical)	• In general, how do you describe your mental/physical health? (categorical)	• Good • Poor • Could be better
Media repertoire (*preference*; DV[c], count)	• If you had to go in to self-isolation, choose 3 activities that would help you cope. (count)	• Netflix or similar streaming services • Exercise • Print media • Work • Computer • Video chat services • Social media • Games and puzzles • Network television • Computer games
Media repertoire (*use change*; DV, nominal)	In the past week which one of your use patterns have changed? • *Twitter* • *Facebook* • *Instagram* • *Games* • *Television* • *YouTube* • *Netflix or similar streaming services* • *Print media* • *Radio, audiobooks, etc* • *Teleconference* • *Telephone*	• Increased • Decreased • Stayed the same • Unused
Media appraisal (primary; DV, scale)	• *Approach* • *Avoid* • *Ignore*	• 0-100
Appraisal (secondary) of *mental/physical health risk* (DV, categorical)	• Are you worried that too much screen time can affect your mental/physical health negatively?	• Yes, I am worried • I am a little worried • No, I am not worried at all • I do not know or it depends

[a]COVID-19: coronavirus disease.

[b]IV: independent variable.

[c]DV: dependent variable.

Media Appraisal for Coping With the COVID-19 Pandemic

Media appraisal was assessed based on an 8-item questionnaire, asking participants to state their opinions ("Definitely true," "Somewhat true," "Not really true," "Definitely false," or "I don't know") about the following statements: (1) *I use social media to be connected while social distancing*, (2) *social media connects me to what is happening in the world*, (3) *COVID-19 news and social media posts overwhelm me*, (4) *social media spreads false information about COVID-19*, (5) *COVID-19 news gives me a sense of knowledge and control*, (6) *I play games or watch TV to distract myself from COVID-19*, (7) *there is too much media hype about COVID-19*, and (8) *I try to avoid the COVID-19 news as much as I can*. The proportions of responses to each question are illustrated in Multimedia Appendix 2.

Using principle component analysis and varimax rotation, we reduced the appraisal questionnaire to 3 factors that cumulatively

explained 57% of the variance in the sample. Factor 1 explained 21% of the variance (after rotation) loaded on items 1, 2, and 5. We refer to this factor as *approach*. The second factor explained 20% of the variance (after rotation) and loaded on items 3, 7, and 8. We refer to these factors as *avoid*. Finally, the third factor explained 16% of the variance loaded on items 4 and 6. We refer to this factor as *ignore*. To compute scores for each factor, responses to each question were recoded as follows: definitely true was recoded to +2, somewhat true was recoded to +1, somewhat not true was recoded to –1, and definitely false was recoded to –2. We then computed the variables *approach*, *avoid*, and *ignore* by calculating the score of each factor by computing the normalized average of the items' scores in that factor. Interclass correlation coefficients of items in each factor were low (Cronbach α=.6), which limits the reliability of these scores, but this scoring allows us to operationalize media appraisal.

Statistical Analyses

To compare associations between independent variables (age, gender, COVID-19 stress, and health status), we used contingency tables and chi-square tests of associations. To examine how different groups selected their coping resources, we computed the odds ratio of an activity being picked by each group category compared to the rest.

The Kruskal-Wallis and multivariate analysis of variance tests were used for group comparison of nominal dependent variables such as change in the media use and scale variable such as appraisal scores, respectively. In all cases, appropriate post hoc analyses were performed, and 95% confidence intervals were reported.

Prism8 (GraphPad Inc) and SPSS 24 (IBM Corp) were used for performing quantitative data analyses and presentation. The dropout rate (ie, respondents who accessed the survey but did not record their responses) was 49 out of 734. Because some of the survey questions were not mandatory, we present the case-wise sample size for each analysis. Details of statistical tests are presented together with results.

Qualitative Analysis of the Open-Ended Questions

We used Nvivo 12 for Mac (QSR Inc) and applied a data-driven approach to the coding of the open-ended questions by exploring the most frequently used words in the response boxes. We then explored the themes that related to disruptions caused by COVID-19 and coping strategies that mattered.

Out of the 689 respondents, 351 provided a response to the following question: *"This outbreak has caused real problems, especially to those who do not have the ability to do their work from home. Can we envision ways in which the media (news, social networks, newsletters, etc) can be used to alleviate their burden?"*

Nearly 11% (38/351) of those responses were "I don't know," "not sure," or "maybe" without any explanation. A little over 7% (26/351) responded with "No" or "Not really" without any explanation. About 5% (20/351) responded "Yes" without any explanation.

Using a word frequency analysis on 267 spell-checked and corrected entries (automatically removing transitional verbs, prepositions, pronouns, conjunctions, articles, quantifiers, and adverbs) revealed that the words *work* (217 counts), *home* (145 counts), *people* (96 counts), *social* (84 counts), *media* (67 counts), *time* (66 counts), *help* (62 counts), *school* (53 counts), *news* (52 counts), *activities* (49 counts), *information* (48 counts), *job* (47 counts), *online* (43 counts), *friends* (38 counts), *cancellation* (38 counts), and *family* (35 counts) were the most frequent ones.

These words were extracted using Nvivo 12's word-query function. Statements that included each word were studied one by one to code for the following themes: the impact and disruption caused by the pandemic, what mattered to individuals in terms of coping with the pandemic, and what kind of help the media could provide. We then performed a node matrix query on each coded concept to create a network representing the co-occurrence of nodes (ie, the number of times that any two words were co-occuring in one statement). Finally, we used an open source software Gephi (version 0.9.2 for Mac, an open source and free software for visualization and exploration of any network types) [31] to identify the emerging concepts that were more important in the open-question responses. The network was partitioned by its modularity (a measure of how a network compartmentalizes into subnetworks), and the nodes were ranked by their eigenvector centrality (EC; a measure of node importance in the network). These results were then used to create a conceptual model for addressing the main question of the study: what media helps and what media hurts?

Results

Sample Distribution in Age, Gender, Health, and COVID-19 Stress Groups

Figure 1 shows the geographical location of the sample. The majority of responders were from Canada (n=515). Descriptive statistics are presented in Table 2. COVID-19 had disrupted the normal life of more than 85% of respondents. Nearly two-thirds of the sample were women, and one-third were between the ages of 35 and 54 years. Less than one-third of the sample considered their mental or physical health to not be good.

Figure 1. Geographic location of respondents.

Table 2. Descriptive statistics.

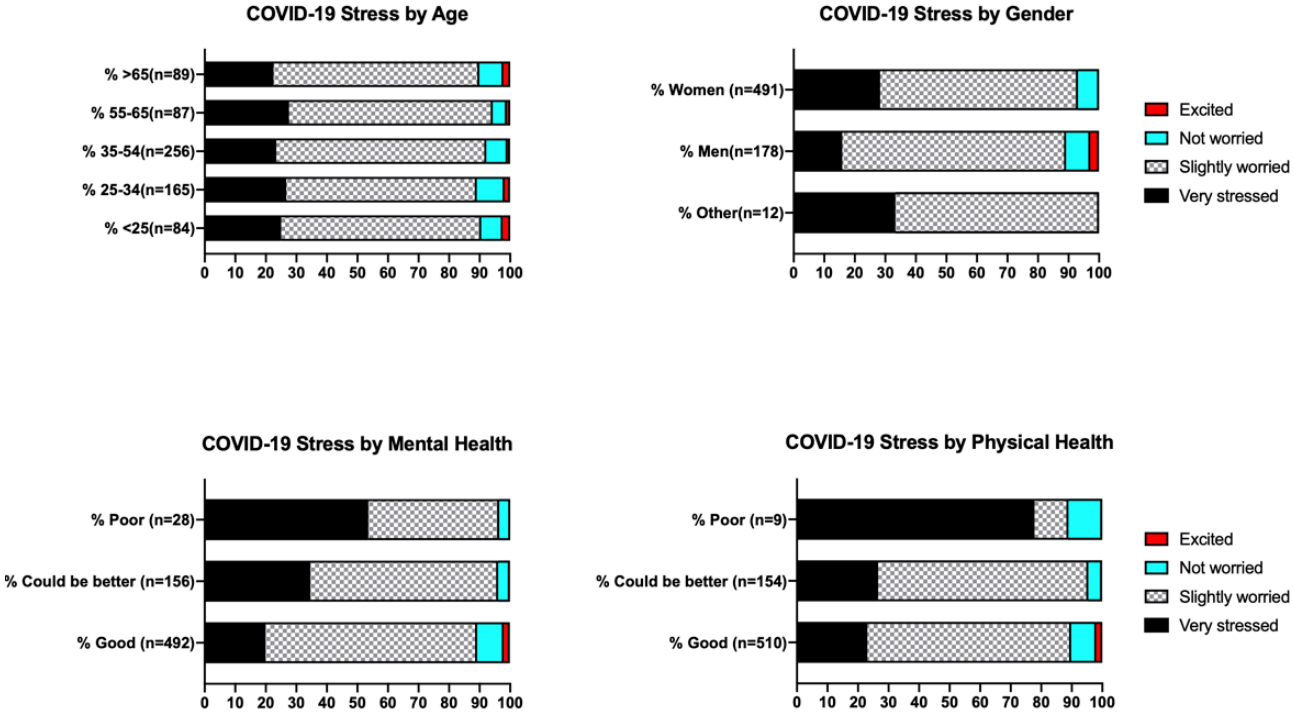
Questions and responses	Participants (N=685), n (%)
Has COVID-19[a] interrupted your normal life?	
Yes	628 (85.6)
No	57 (7.8)
Missing	49 (6.7)
Which one of these statements describe how you feel about the COVID-19 pandemic?	
Very stressed	169 (23)
Slightly worried	452 (61.6)
Not worried	50 (1.4)
Excited about it	10 (1.4)
Missing	53 (7.2)
Are you in quarantine or self-isolation?	
Yes	354 (48.2)
No	329 (44.8)
Missing	51 (6.9)
What is your age category?	
Younger than 25 years	84 (11.4)
25-34 years	165 (22.5)
35-54 years	259 (35.5)
55-65 years	88 (12)
Older than 65 years	89 (12)
Missing	49 (6.7)
What is your gender?	
Male	179 (24.4)
Female	494 (67.3)
Nonbinary	4 (0.5)
I prefer to not answer this question	8 (1.1)
Missing	49 (6.7)
Generally, how would you describe your mental health?	
Good	496 (67.6)
Poor	28 (3.8)
Could be better	156 (21.3)
Missing	54 (7.4)
Generally, how would you describe your physical health?	
Good	512 (69.8)
Poor	9 (1.2)
Could be better	166 (21.3)
Missing	57 (7.8)

[a]COVID-19: coronavirus disease.

Group Differences in Perception of COVID-19–Related Stress

Chi-square test of contingency showed no association between categories of COVID-19 stress perception and *age* (χ^2_{12}=5.04, *P*=.96), but associations with *gender* (χ^2_6=15.05, *P*=.03), self-assessed mental health (χ^2_6=30.93, *P*<.001), and self-assessed physical health (χ^2_6=20.83, *P*=.002) were significant. As can be seen in Figure 2, men were half as likely to be "very stressed" by COVID-19 (odds ratio 0.48, 95% CI 0.31-0.75), and those with good mental health were also less likely to be "very stressed" (odds ratio 0.415, 95% CI 0.29-0.60). We found a significant association between self-assessed mental health and *age* (χ^2_8=41.2, *P*<.001; a higher proportion of young respondents considered their mental health as poor or could be better) but there was a nonsignificant association with *gender* (χ^2_4=8.8, *P*=.07).

Figure 2. Perceptions of COVID-19 stress across age, gender, and self-assessed health groups. COVID-19: coronavirus disease.



Group Differences in Media-Type Preferences for Coping With Self-Isolation

The ranked counts of activities that individuals picked as the three most important for helping them cope with self-isolation and quarantine are presented in Multimedia Appendix 1. The most frequently selected option was *Netflix or similar streaming services* (402/685, ~60%). Netflix is not the only online streaming technology but the first of this genre; for simplicity, we use the name of this brand to refer to any streaming services of this kind (eg, Amazon Prime, Hulu, Home Box Office, Crave, Disney, GEM)

Interestingly, exercise was the second (358/685) and print media (264/685) the third most important activities.

However, when splitting the sample across *age* and *gender*, the within-group proportions of responses revealed different patterns (Figure 3).

In terms of *age*, individuals younger than 35 years were two times more likely than the rest of the sample to pick *Netflix or similar streaming services* (odds ratio 2.04, 95% CI 1.46-2.85) and 2.3 times more likely to pick *computer games* (odds ratio 2.28, 95% CI 1.43-3.63) for coping in a quarantined condition. By contrast, individuals older than 55 years were twice more likely to pick *print media* (odds ratio 2.02, 95% CI 1.43-2.87) and more than three times likely to pick *network television* (odds ratio 3.39, 95% CI 2.16-5.32) compared to those who were younger. To have a *work computer* was most important for those aged *35-54* years (odds ratio 1.82, 95% CI 1.32-2.5). The odds of using *social media*, *teleconferencing*, *exercise*, and *solo games or puzzles* were not significantly different between groups.

In terms of *gender*, the few (n=12) who did not specify a binary gender showed visibly different odds in terms of preferences, but in the absence of a large enough sample, we will not discuss the statistical significance of these findings. However, comparing men (n=179) and women (n=494) showed that men were twice more likely than women to pick *work computer* (odds ratio 1.45, 95% CI 1.02-2.06), three times more likely to pick *computer games* (odds ratio 3.3, 95% CI 2.05-5.27), and 1.8 times less likely to pick *social media* (odds ratio 0.55, 95% CI 0.368-0.833) for coping with self-isolation in the case of quarantine.

In terms of *mental health*, those who indicated their mental health was good were 1.5 times more likely to pick *work* (odds ratio 1.46, 95% CI 1.02-2.09) and more than twice less likely to pick *Netflix or similar streaming services* (odds ratio 0.428, 95% CI 0.30-0.62). Other differences were not significant. It is

worth mentioning that those who indicated their physical health was good were more likely to pick *exercise* (odds ratio 2.4, 95% CI 1.67-3.44), but other differences were not significant.

With the exception of one response ("reading good fiction books"), no other respondents suggested alternative coping activities.

Figure 3. Group differences in preference for activities to cope with self-isolation or quarantine.

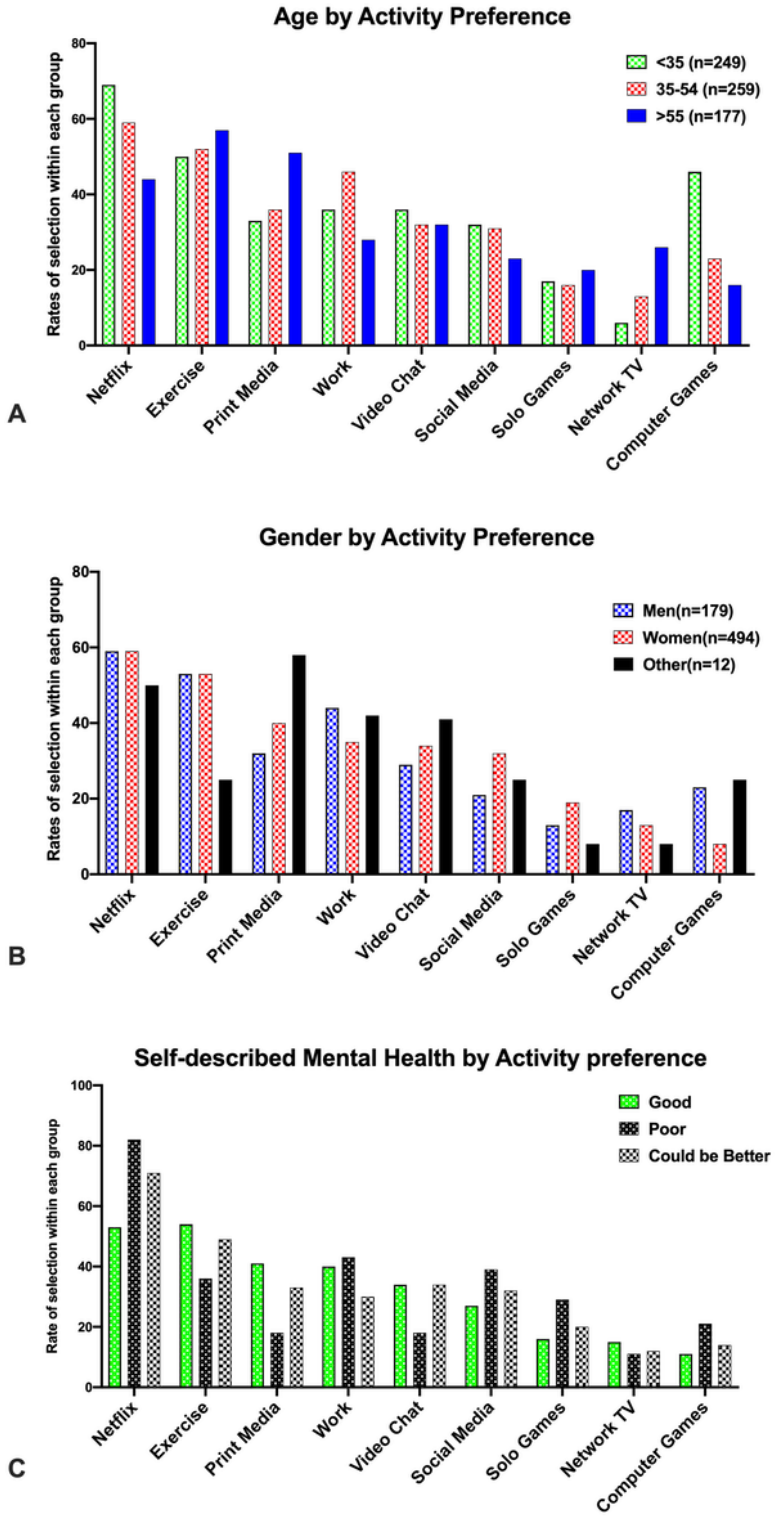


Table 3 provides the response frequency to how media use has changed as a result of the COVID-19 pandemic. The highest frequency of increased use was in *video chat*, followed by *telephone* and *Netflix* or similar streaming services. The highest frequency of unchanged use was with print media and *YouTube*. The highest frequency of unused media was *Twitter* (also with the lowest rate of increase), followed by *games* (although its use increased), *Instagram* (although its use also increased), and *audio media*.

Table 3. Frequency of response to changes in media use as a result of the coronavirus disease pandemic.

Media repertoire use	Unused, n (%)	Decreased, n (%)	Unchanged, n (%)	Increased, n (%)
Video chat (n=329)	44 (13.4)	2 (0.3)	25 (7.6)	258 (78.4)
Telephone (n=328)	36 (11)	6 (1.8)	124 (37.8)	162 (49.4)
Netflix or similar (n=667)	118 (17.7)	14 (2.1)	262 (39.3)	273 (40.9)
Facebook (n=664)	131 (19.7)	26 (3.9)	246 (37)	261 (39.3)
Television (n=664)	160 (24)	26 (3.9)	256 (38.6)	222 (33.4)
YouTube (n=664)	149 (22.4)	14 (2.1)	294 (44.3)	207 (31.2)
Print media (n=334)	61 (18.3)	17 (5.1)	162 (48.5)	94 (28.1)
Games (n=644)	312 (48)	9 (1.4)	159 (24.4)	164 (25.5)
Instagram (n=636)	278 (43.7)	27 (4.2)	172 (27)	159 (25)
Audio media (n=321)	137 (42)	21 (6.5)	97 (30.2)	66 (20.6)
Twitter (n=627)	405 (64)	9 (1.4)	116 (18.5)	97 (14.2)

The Kruskal-Wallis test (with Dunn's correction for multiple comparison of post hoc pairwise comparisons) was used for these analyses.

In terms of *age*, differences in the mean ranks of use were statistically significant for the following:

- *Twitter* (χ^2_4=19.38, *P*=.001): predominantly higher in those aged *25-34* years, with significant differences in those older than 65 years (mean rank difference=62.8, adjusted *P*=.02) and younger than *25* years (mean rank difference=62.3, adjusted *P*=.04)
- *Instagram* (χ^2_4=56.14, *P*<.001): predominantly decreasing with *age* (adjusted *P* values<.004)
- *Games* (χ^2_4=18.0, *P*=.001): predominantly and significantly higher in the two *age* groups younger than 35 years. It is notable, however, that the differences between game use of those older than *65* years were not significantly different from those younger than *25* years or between *25* and *34* years (adjusted *P* values>.89), and that the mean ranks of the game use in those older than 65 years were higher than those between the ages of 55 and 65 years.
- *Netflix* or similar streaming services (χ^2_4=21.1, *P*<.001): generally decreasing with age but statistically lower in those older than 65 years and younger than 35 years (adjusted *P* values <.004)
- *Videoconferencing* (χ^2_4=26.74, *P*<.001): with statistically significant differences between the *age* group that was older than *65* years and those younger than 25 years (adjusted *P*<.001), those aged *25-34* years (adjusted *P*=.02), and those aged *35-54* years (adjusted *P*<.001)

In terms of *gender*, differences in the mean ranks of use were statistically significant for *Facebook* (χ^2_2=7.66, *P*=.02)—higher in women (adjusted *P*=.02)—and *Instagram* (χ^2_2=19.07, *P*<.001)—higher in women (adjusted *P*<.001). However, other use differences were not statistically significant (*P* values>.2).

In terms of *mental health*, differences in the mean ranks of use were statistically significant for *Instagram* (χ^2_2=6.91, *P*=.03), *YouTube* (χ^2_2=12.14, *P*=.002), and *Netflix* (χ^2_2=11.04, *P*=.004), and higher in those who stated their mental health "could be better" versus those who considered their mental health "good" with mean rank differences of *Instagram*=41.6 (adjusted *P*=.01), *YouTube*=51.6 (adjusted *P*=.005), and *Netflix*=47.14 (adjusted *P*=.01).

Link Between Subjective COVID-19 Stress and Change in Media Use

Group differences in subjective *COVID-19 stress* ("slightly worried," "very stressed," "not worried," and "excited by it") and media *use change* (unused=0, decreased=–1, unchanged=1, increased=2) were significant for *Facebook* (χ^2_3=11.76, *P*=.008), *television* (χ^2_3=12.40, *P*=.006), *YouTube* (χ^2_3=8.577, *P*=.04), and *Netflix* (χ^2_3=10.71, *P*=.01) but not significant for any other activity (χ^2_3<7, *P* values>.1).

The post hoc Dunn's pairwise comparison (with Bonferroni correction) indicated that the mean rank of *Facebook* use for those who were "Very stressed" (95% CI 1.01-1.4) was significantly higher than those who were "Slightly worried" (95% CI 0.91-1.16), with a mean rank difference of 48.37 (adjusted *P*=.02), or "Not worried" (95% CI 0.28-1.5), with a mean rank difference of 81.55 (adjusted *P*=.03).

The mean rank of *television* use for those who were "very stressed" (95% CI 0.85-1.23) was significantly higher than the "not worried" group (95% CI 0.25-1.08), with a mean rank difference of 99.47 (adjusted *P*=.005).

The mean rank of *Netflix* use for the "Very stressed" (95% CI 1.02-1.44) was significantly higher than "Slightly worried" (95% CI 1.15-1.36), with a mean rank difference of 73.81 (adjusted *P*=.04), or "not worried" (95% CI 0.11-1.23), with a mean rank difference of 90.0 (adjusted *P*=.01).

Post hoc comparisons of *YouTube* did not yield significant results. These observations support our previous hypothesis that increased use of social media and passive entertainment is used as a coping strategy against stress.

Group Differences in Media Appraisal for Coping With COVID-19

So far, we have shown significant demographic and health-related differences in media preference and changes in media use as coping strategies. To what extent are these differences explained by differences in appraisal?

We used a multivariate general linear model with an appraisal variable (*approach*, *avoid*, and *ignore*) as dependent variables and tested a full factorial model with *age*, *gender*, and *mental health* as independent variables, with Bonferroni correction for multiple post hoc comparisons. For simplicity of interpretations, nonbinary respondents were excluded, as their number was <12; mental health self-assessments "poor" and "could be better" were coded into "not good."

The multivariate Pillai's trace test (chosen because of its robustness to violations of model assumptions) revealed significant contribution to the model from *gender by age* ($F_{12,1911}$=1.86, *P*=.04; Figure 4; mainly affecting *ignore*, $F_{4,637}$=2.70, *P*=.03), from *gender* ($F_{3,635}$=8.23, *P*<.001; affecting both *avoid*, $F_{1,637}$=5.84, *P*=.02, and *approach*, $F_{1,637}$ =14.31, *P*<.001; Figure 4), and from *mental health* ($F_{3,635}$=5.13, *P*=.002; affecting *ignore* scores only, $F_{1,637}$=13.88, *P*<.001).

Post hoc comparisons (adjusted for Bonferroni correction) showed that, compared to men, women had significantly higher scores of both *approach* (95% CI 9.44-31.405)—meaning that they found information, connection, and control in social media—and *avoid* (95% CI 2.74-26.45)—meaning that they were more overwhelmed by COVID-19 news, found the media hype too high, and tried to avoid the news as much as possible.

Figure 5 illustrates that differences in mental health were associated with significant differences in *avoid* (95% CI 1.7-17.97) and *ignore* (95% CI 11.25-36.33), meaning that they watched television or played games to distract themselves from the news and considered media a source of false information. Differences in physical health were only associated with *ignore*.

Figure 4. Age- and gender-related differences in appraisal (mean, standard error of the mean).

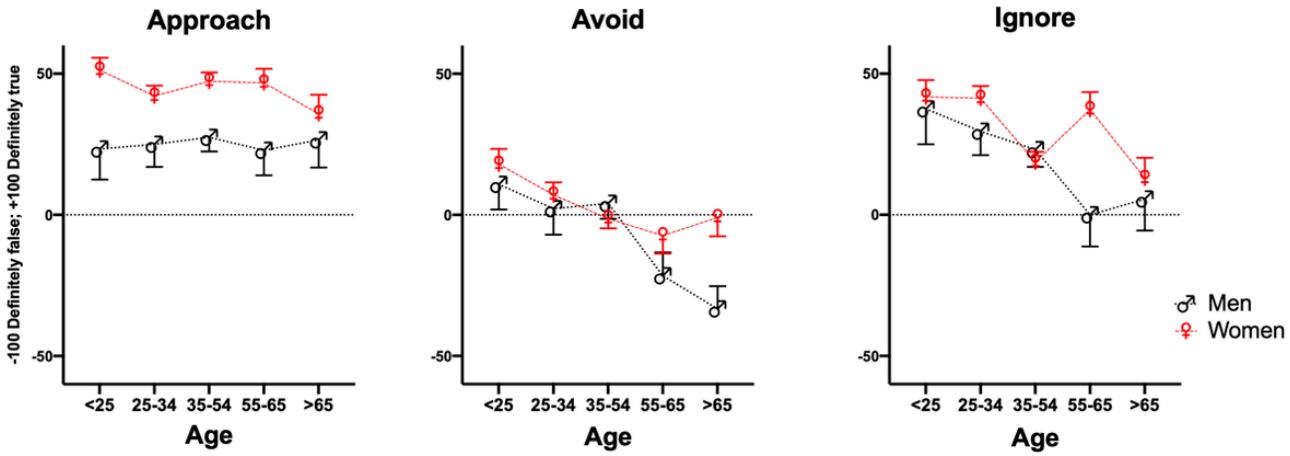
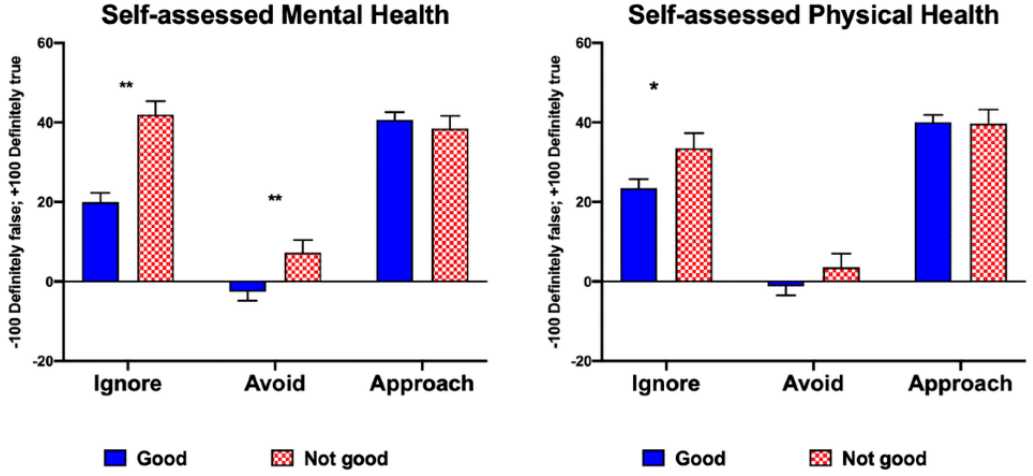


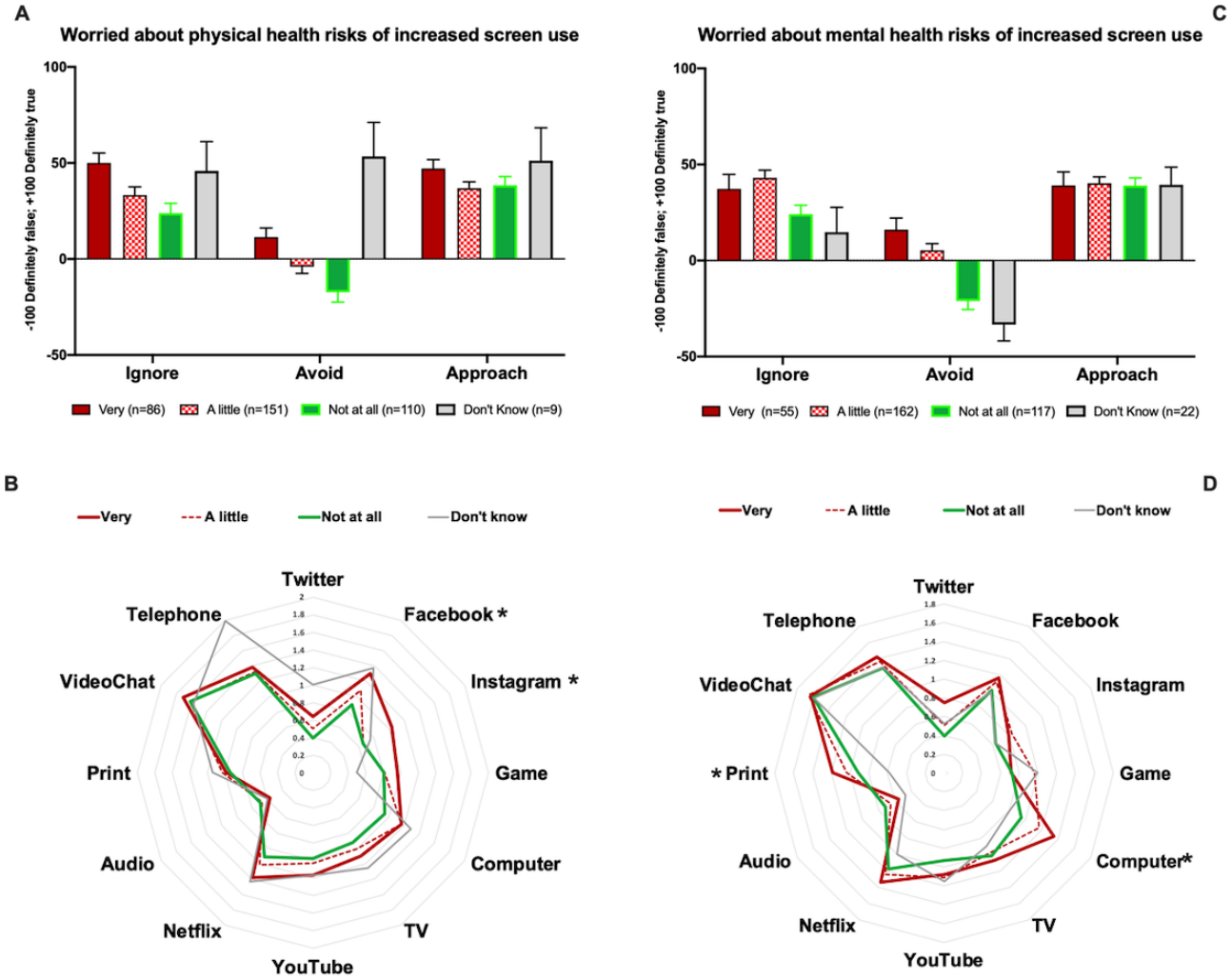
Figure 5. Physical- and mental health–related differences in appraisal (mean, standard error of the mean). Pairwise comparison of each variable independently shows significant differences related to self-assessed physical and mental health. We also found a significant likelihood that physical and mental health were related. (**P*<.05; ***P*<.005.).



Relationship Between Appraisal, Use, and Worry About Mental and Physical Health Risks of Increased Screen Time

In the third week of the sampling, when the possibility of overusing screens for work and entertainment had become higher than before, we added two questions and asked whether users were concerned about increased time on screens becoming physical or psychological stressors (Figure 6).

Figure 6. Relation between media appraisal, media use, and perceived risk of mental and physical health deterioration as a result of increased media use. * shows media types whose usage was significantly different between groups (P<.05).



In response to *"Are you worried that too much screen-time can affect your physical health negatively,"* 86 responded "Yes," 151 responded "a little," 110 responded "Not at all," and 9 responded "I don't know."

In response to *"Are you worried that too much screen-time can affect your mental health negatively,"* 41 responded "Yes," 152 responded "a little," 107 responded "Not worried at all," 21 responded "I don't know," and 33 responded "It depends."

A total 14 respondents further commented that they were concerned for anyone who might develop addiction or detrimental lifestyle habits (especially children). These were recoded into "Yes." There were 10 other respondents that mentioned the necessity of balance; using screens for *exercising*, *making art*, *seeking information*, and *online learning* was positive, but using them for *all-day television watching* was bad; *using in moderation* was good, but *addiction* was bad. We recoded these responses to "a little." There were 8 respondents that indicated that *they had control over the time they spent on screens* and that, during COVID-19, the screens provided them with an opportunity to "*educate themselves and their kids*," "*exercise*," and "*distract from boredom*." These responses were recoded to "not worried at all." There was 1 respondent that indicated that they *did not understand the question* (this response was coded to "I don't know").

Responses to the questions of concern were congruent in 53.9% (192/356), meaning that respondents expressed the same degree of concern about physical and mental health risks.

A multivariate analysis of variance with appraisal variables as dependent and *mental* or *physical health risk*s of screen time as predictive factors showed significant association between *ignore* and *mental health risk* ($F_{3,350}$=13.9, P=.009) and *physical health risk* ($F_{3,350}$=4.27, P=.006). *Avoid* was also associated with worry about *mental health risks* ($F_{3,350}$=14.5, P<.001) and *physical health risks* ($F_{3,350}$=5.55, P=.001)—in both cases, those

who were very worried ignored and avoided the media more than others.

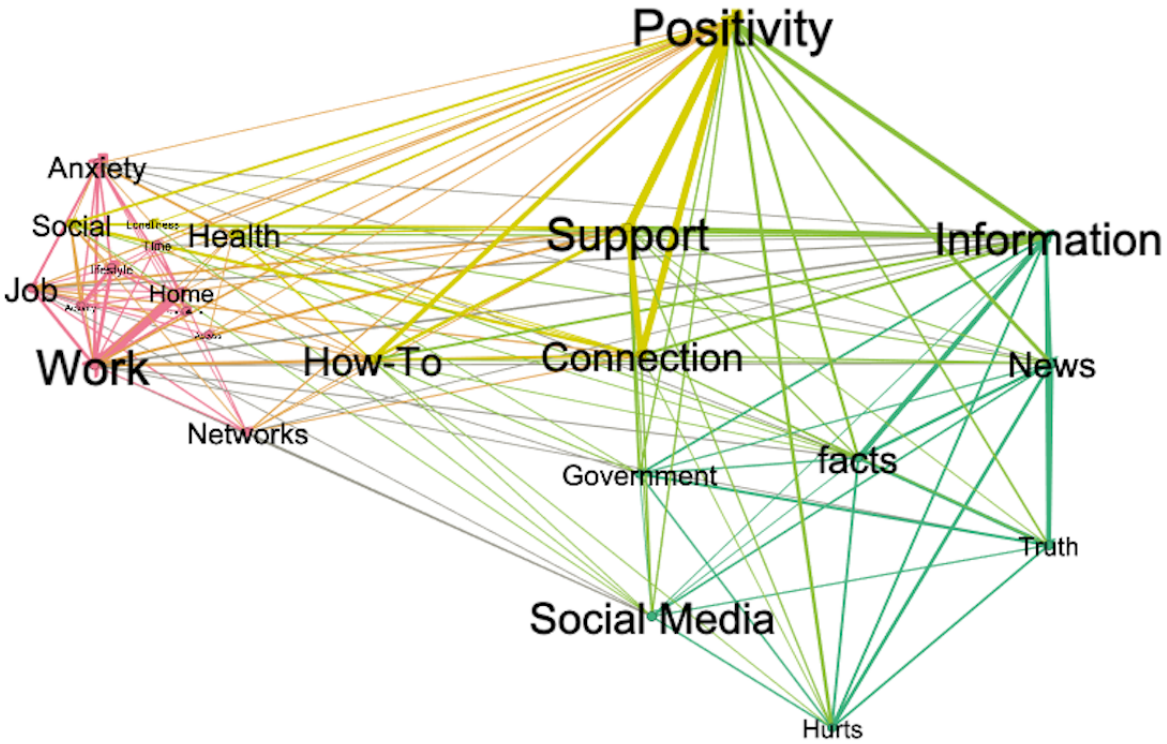
Figure 6 illustrates that concern for mental and physical health risks were associated with significant differences in some media's use. The Kruskal-Wallis test showed that those who were worried about *mental health risks* had significantly different use patterns mainly related to increased use of *work computers* (χ^2_3=7.95, *P*=.047), *print media* (χ^2_3=10.08, *P*=.02), and there was a nonsignificant use pattern for *Twitter* (χ^2_3=7.55, *P*=.06). In contrast, those who were worried about *physical health risks* had increased use of *Facebook* (χ^2_3=10.88, *P*=.01) and *Instagram* (χ^2_3=9.18, *P*=.03) but a nonsignificant trend for increase in *Netflix* (χ^2_3=6.99, *P*=.07).

Qualitative Analyses of the Open-Ended Questions

Our qualitative analyses were data driven in that we simply counted the number of words that were most frequently mentioned in response to the open-ended questions, namely, "How has COVID19 disrupted your normal life?" and "Can we envision ways in which the media (News, Social networks, newsletters, etc.) can be used to alleviate the burden?" (see the word cloud in Multimedia Appendix 3). Responses that focused on the impact of COVID-19 included frequent references to *anxiety*, *job* (*work* and *home*), and *lifestyle* (*activity*, *cancellations*, and *time*). Responses that discussed affordances of different media types included frequent mentions of the words *information* (often together with *facts*), *how-to* tips and learning opportunities, *news* (which could be hurtful if overly negative), and *support* (*social media* and specific *support networks*). In terms of what was important for users, we found frequent mention of words like *access and affordability*, *government*, *health*, *connection*, *positivity*, and *truth*.

Looking at the linkage between these nodes revealed three network communities (Q=0.301), which are represented in different colors (Figure 7). The letter size of each network node represents the importance of the node in terms of EC. The three communities were associated with three important nodes. *Positivity* was the most important one because it was linked to other important nodes such as *information*, *support*, *connection*, and *how-to*. The second most important node was *work* and was mostly represented by nodes that related to the impact of COVID-19. The third most important network was *information* and was mostly represented by nodes that referenced importance of *facts*, *news*, *truth*, *social media*, and *"hurt,"* which referred to comments about the detrimental impact of false news, sensationalized and politicized messaging, and stressful hype and catastrophizing. Examples of respondents' statements about *positivity, information, and work* are provided in Multimedia Appendix 4.

Figure 7. Results of qualitative network analysis. Colors represent network communities. The size of the letter is proportionate to eigenvector centrality (a measure of the hubness of each node). The thickness of edges reflects the weight of each edge.



Discussion

In a follow-up to our previous work studying the relation between screen use and stress [1], we conducted a brief cross-sectional survey and asked a subset of questions based on the previous study to investigate whether higher levels of subjective stress predicted an increase in use of information, social, and entertainment media as means of coping with COVID-19 disruptions. Additionally, we asked questions to assess whether respondents worried that their increased dependency on screen-based communications was perceived as a risk to their mental and physical health.

What Media Helps, What Media Hurts?

One of the aims of this survey is to address the question of what screen-mediated interventions are needed to respond to the stress caused by the COVID-19 pandemic. Our position is that, in designing and promoting any digital health interventions, we must first ask whether the digitized intervention risks becoming a stressor in and of itself, and then mitigate those risks in design [27,32].

Despite being a short survey, the mixed methods approach allowed us to explore our question from several angles: do the subjective intensity of feeling stressed by COVID-19 or the self-assessed state of mental and physical health explain variations in media use, and do we find similar results if we look at the question of media use from different perspectives, like choosing activities for quarantine, reporting on changes in use pattern, or appraisal of why to approach, avoid, or ignore media?

Indeed, we found converging results that all point to a close association between feeling stressed and reaching out to media for coping.

Passive Viewing of Self-Curated Information or Entertainment Content Helps

The first important finding is that media with passive (but selectable) viewing content were important for coping with COVID-19 stress. Increased uses of *Facebook*, *Netflix*, and *television* were significantly associated with the degree that individuals found the COVID-19 pandemic stressful—those who reported being "Very stressed" had higher use of these specific media. This supports our hypothesis that there is a causal relationship between subjective stress and higher dependency on social and entertaining media as coping strategies. Could it be that watching negative news was causing the added stress? Although we cannot reject this hypothesis, our results indicate that individuals who had higher levels of stress avoided and ignored such media more (Figures 5 and 6). In other words, the appraisal of media in relation to individual's needs motivated their approach to passive viewing but led to active avoidance and filtering of the negative content out of one's life.

This is consistent with a common finding in media studies that choice and self-determination can affect the evaluation of the situation and perceived satisfaction with media [6,14,33-35]. Studying the binge-watchers' motivations and affective states, Castro et al [36] showed that relaxation, boredom relief, and escapism were the top reasons why individuals were attracted to streaming services and that watching certain content such as comedies would have a quantifiable effect on positive and negative affective state.

It is plausible to suggest that active choice-making, even when using passive media (*streaming services*, *YouTube*, *Instagram*), was the reason why these types were more important to those who reported that their mental health was not "Good." In general, *Netflix or similar streaming services* were the most frequently selected item for coping with social isolation (followed by *exercise*, *print media*, and *work*), but those whose mental health was not "Good" were twice as likely to pick *Netflix or similar streaming services* but twice less likely to pick *work*—perhaps suggesting a need for distraction from reality or killing time. In addition, use of other self-selectable viewing services such as *YouTube* and *Instagram* also increased in those with mental health dissatisfaction (mainly the young).

Social Media Helps Women More

The role of social media, as a coping tool was ambivalent. Four out of five respondents indicated that they used social media to be connected while social distancing and remaining connected to what is happening in the world, but many of them also thought that social media news was overwhelming to them and spreading false information.

Interestingly, *social media* was generally ranked less important than *Netflix*, *exercise*, and *print media*, although it was frequently discussed in the open-ended question (see word cloud in Multimedia Appendix 3). Several studies have pointed to a link between compulsive use of social media and mental health [1,9,10,37]. Although we found that preferences for *social media* were gender dependent, differences were not related to age or self-assessments of mental health. Women were more likely than men to pick *social media* to cope with isolation but less likely to pick *work* or *computer games* (but they were not different in other categories). Women also significantly increased their use of *Facebook* and *Instagram* after the COVID-19 lockdown, although they did not significantly differ from men in use of other media.

Ryan et al [6] have shown that dependency on social media is linked to gratification factors such as relationship maintenance, passing time, entertainment, and companionship. This explains women's preference for *social media* given that women of all ages had higher *approach* scores compared to men (Figure 3). Recall that approach relates to agreeing with the statement that *social media provided an opportunity to stay connected while in isolation*, to *be informed*, and to have *knowledge and a sense of control*. Interestingly, women, especially the older ones, and those with mental health complaints (Figure 4) also had slightly higher scores of *avoid*, which referred to *finding too much hype*, *being overwhelmed*, and *trying to avoid the news about COVID-19*.

Unless Positive, News and Social Media Would Be Hurting

Taking a data-driven network analysis approach, *positivity* emerged as the most central theme, connecting between different

nodes related to the media's helpfulness in coping with COVID-19 stress. Refences to *positivity* had common nodes with *support*, *information*, and *news* (factual and nonsensational), as well as with opportunities for learning *how to* do new things or gain control by communicating with the *government* through *social media*. *Positivity* was also connected with the second most important node, *work*, which was central to the subnetwork that related to *anxiety* caused by COVID-19 and various practical interruptions such as losing *jobs*, working or not being able to work at *home*, and changing *lifestyle* and *activities*. *Work* was also strongly connected to *support*, *information*, and *networking*, which enables one to earn a living at *home*. As expected, *information* was central to the third subnetwork, with strong connections to *facts* and *truth* (which connected *information* to *positivity*), as well as to support (via *social networks*, *how-to* instructions, and *connections*.) An important connection between *positivity* was to the node *hurts*, which means that, although informing, connecting, educating, distracting, and encouraging were positive aspects of media, spreading false, fearful, and anxiety-increasing messages (due to politicization, sensationalization, or catastrophizing) were hurtful.

These observations confirm the findings of Hoog and Verboon [38], who used a momentary ecological assessment method for 63 participants who reported their affective state during 10 days of following the news and showed that exposure to bad news is a psychological stressor, albeit depending on personality factor. Marin et al [39] have also shown that exposure to negative news made women more susceptible to being physiologically responsive (in terms of cortisol release) to later experimental challenges (Trier stress test). Data in this survey suggest that self-awareness on the stressful nature of media adjusts how individuals adapt. Here, we showed that respondents who did not consider their *mental health* as "Good" or who worried that increased screen time would be a *mental health risk* had significantly higher *avoid* and *ignore* scores (Figures 5 and 6). It should also be mentioned that engagement with our open questions was higher in those who considered their mental health as good (see Multimedia Appendix 5). Thus, we note that this survey may not have reached those who find the current situation most stressful and who will be the target for media-based stress-reduction interventions.

Do Screens Cause or Reduce Stress?

To Each Stress, Their Own Screen

Informed by the fact that stress is a multifaceted adaptive experience [19,21,22,25], our earlier work that motivated this survey postulated that individual differences in stress perception and coping approaches influence the affordances of screen use [1]. In this study, we show that even general factors like age, gender, and self-assessed mental health or situation stress (due to COVID-19) reveal a heterogeneous pattern of preferences and use.

Griffiths and Szabo [14] have long emphasized that, in studying the relation between screens and stress, the context in which a particular media type is adopted is critically important. As Figures 3-6 clearly demonstrate, whether users approach, avoid, or ignore media can vary with age, gender, or mental health status. Our observation (eg, the significant difference in *avoid* scores between those younger than 25 years and those aged 35-54 years) on *age* groups is consistent with previous finding by Kuss et al [37], who have shown a generation-specific (Y vs X) link between anxiety in developing behavioral dependency on social media use. Interestingly, although we observed a significant difference in preference for games in those younger than 35 years, the differences in game use of those older than *65* years were not significantly different from those who were younger than *25* years or those aged *25-34* years. Even the mean ranks of game use in the older than 65 years category were higher than those aged 55-65 years. This is consistent with findings in a large-scale cross-sectional study by Birks et al [40], who showed that older adults play games for emotion regulation goals (rather than challenge and skill). Although variables such as *age* and *gender* are overly reductionist in explaining motivation, uses, and gratifications of different media, we have been able to demonstrate the importance of employing multifactorial, mixed methods inquiries within a media repertoire framework [28,29]. By taking a media repertoire approach within the theoretical framework of studying stress with the appraisal model [23], we have been able to illustrate the complexity of interindividual and intergenerational relationship to different types of media within the particular context of coping with COVID-19 isolation.

The Main Worry About Excessive Screen Time Is Physical Stress

Although this cross-sectional survey is able to show that some applications of media use may be helpful in psychological destressing, it cannot show whether the same application would become stressful over time. However, examining the concerns of the respondents about potential risk factors can inform whether users appraisal and behavior may change over time.

Applying the appraisal model, we hypothesized that those who have a negative appraisal of the media's affordances for coping would have a different repertoire of media preference and use. Indeed, we found that those who were "Very worried" about both *mental health risks* and *physical health risks* also had higher negative appraisal scores—more prone to avoid or ignore media. Interestingly, however, effects of the appraisal of *mental health risks* and *physical health risks* on media repertoires were not similar (Figure 6). Compared to those who were "Not worried at all," those who were "Very" or "A little worried" about the *physical health risks* of screens had increased use of *Facebook* and *Instagram* (and a trending, but nonsignificant, higher use of *Netflix*). In contrast, those who were "Very" or "A little worried" about the *mental health risks* of screens had increased use of *print media* and *work computers*. Recall (Figure 3) that *age*, *gender*, and self-assessed *mental health* all played a role in the appraisal of, and preferences for, different media. However, contrary to our expectation, appraisal of *mental health risks* were not associated with differences in use of *social media* or *Netflix*, which were increased in those who were dissatisfied with their mental health (reported it as "Poor" or "Could be Better").

This contradiction is not surprising. Although concerns about mental health stressfulness of media are debatable, either due

to cultural differences [25], wide misconceptions about the definition of stress [18], or individual reasons for use [6,14-16,36], deleterious physical health effects of excessive screen time are quantifiable [4,5,41,42]. One out of four respondents were very worried about the physical health risks of increased screen time, but only 1 out of 7 were concerned about the mental health risks. In about half of the sample, concerns about mental and physical health were concordant. However, the likelihood of responding with "I don't know" was 3 times higher with regard to concern about mental health, compared to physical health.

These observations highlight the complexity of studying the psychological underpinnings of media use, which necessitate integrative mixed methods approaches to studying the harms or benefits of screen-based interventions.

Conclusion

Summary

We used a theoretical framework based on Mason's definition of stress (experience of conditions of novelty, unpredictability, threat to self, and sense of control) [19,20] and stipulated that COVID-19 was a stressor. We applied an iterative and multifactorial method to examine the relation between use of media and individual differences in age, gender, and subjective assessments of their stress and mental health state. Using a repertoire-oriented framework, we examined interrelation between different available technologies and the factors that influence individual's choice in the amount of different media or content use.

The data for this survey was collected within the first 4 weeks of North America going into a mandatory lockdown, and its objective was to assess whether (and which) media was serving to destress or was causing more stress. To the best of our knowledge, this is the first study to have employed the media repertoire methodology to investigate the relationship between media use and coping with COVID-19.

Our mixed methods analysis revealed that higher stress was associated with higher prevalence of using passive viewing (streaming service, YouTube, and Instagram), especially in those who did not consider their mental health to be good. The relationship to social media was complex, and although many relied on it for connection, information, and control, many also tried to avoid it for being overwhelming and overhyped. Nevertheless, women were more likely to approach it, and those with mental health complaints were more likely to avoid and ignore it. Our qualitative analyses underlined the importance of positivity and information in helping individuals cope with disruptions in the work or social life, via creating networks of support that connect individuals to resources for training (or retraining), financial planning (or replanning), and social and mental health support.

Our findings are important as they underline the importance of interindividual factors in appraisal and preferences for different media types. Evaluating the affordances and stresses of using screen-based technologies are especially important for those who seek innovative screen-based solutions for helping people to deal with the new realities of this pandemic.

Limitations and Future Work

We have not used any psychometric instruments to formally assess stress of mental health and have only relied on subjective self-assessments. We have previously shown that self-assessment of stress corresponds to higher scores of stress measured by validated instruments used for measuring stress [1], and omission of such questionnaires helped us shorten the survey and thus have a higher completion rate (95%). However, this limits the clinical relevance of our finding and requires follow-up studies.

Although we tried to distribute the survey as widely as possible, as can be seen in Figure 1, more than 90% of our respondents were in North America, and with a few exceptions, there was no representation from any African countries or important Asian countries like China. In surveys such as this, accessibility to technologies such as streaming services is region dependent. Many countries impose censorship on many types of media, filtering access to services such as social media. Our survey targeted those who could read and write in English. Therefore, the findings and interpretations must not be generalized.

This survey is still open [30], and we hope that by collecting data over time, we will be able to analyze shifts in media use trends as individuals come to like or be bored with certain interventions. In any such future studies, mixed methods multifactorial assessment of the interactions between appraisal and use will remain informative.

Acknowledgments

This study is supported by PERFORM Centre (Concordia University's Office of Research) support to Media Health Lab.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Adaptation of Lazarus' transactional model of stress and coping. The source image on the left is from https://commons.wikimedia.org/wiki/File:Transactional_Model_of_Stress_and_Coping_-_Richard_Lazarus.svg. On the left, a depiction of the adaptation of the model in this study: The novelty, uncontrollability and unpredictability of the COVID-19 pandemic, which is a threat to health and financial resources makes it into a stressor. To cope, individuals who do not have

sufficient local resources will use information and communication technologies depending on their needs and preferences, as well as their coping styles. The Media usage will be re-appraised in relation to its presumed benefit to mitigate the stress (increase control, reduce threat, and reduce novelty and unpredictability via connection and information.).
[PNG File , 138 KB - jmir_v22i8e20186_app1.png]

Multimedia Appendix 2
Responses to Media Appraisal Questionnaire: More than 80% (548/685) of the respondents indicated that they used social media to be connected while social distancing and remaining connected to what is happening in the world. This was despite the fact that more than 72% (506/685) also thought that social media news where overwhelming them and were spreading false information. For 70% (480/685) of respondents, social media news and posts about COVID-19 provided a sense of control. About 50% (345/685) of respondents thought there was too much media hype about COVID-19, and about 29% (196/685) indicated that they will definitely or somewhat try to avoid the news related to COVID-19. More than 60% (417/685) of the sample relied on distraction from COVID-19 news, by playing games or watching Television.
[PNG File , 181 KB - jmir_v22i8e20186_app2.png]

Multimedia Appendix 3
Word cloud is generated from the first 1000 words extracted from all responses to all open-answer questions, asking to describe the ways in which the COVID-19 pandemic has disrupted one's life, and what role the individuals envision for media in order to make it beneficial and helpful in coping with the new reality.
[PNG File , 109 KB - jmir_v22i8e20186_app3.png]

Multimedia Appendix 4
Examples of responses to open-answer questions.
[DOCX File , 22 KB - jmir_v22i8e20186_app4.docx]

Multimedia Appendix 5
Relation between mental health and engagement with open-ended questions. Response counts are categorized based on thematic nodes extracted from qualitative analysis. Those who considered themselves in better mental health states seem to have been more elaborate in describing situations or needs related to Work and Positivity. Metal Health 1 = Good; 2, Poor; 3, Could be Better.
[PNG File , 26 KB - jmir_v22i8e20186_app5.png]

References

1. Khalili-Mahani N, Smyrnova A, Kakinami L. To Each Stress Its Own Screen: A Cross-Sectional Survey of the Patterns of Stress and Various Screen Uses in Relation to Self-Admitted Screen Addiction. J Med Internet Res 2019 Apr 02;21(4):e11485 [FREE Full text] [doi: 10.2196/11485] [Medline: 30938685]
2. Jung SJ, Winning A, Roberts AL, Nishimi K, Chen Q, Gilsanz P, et al. Posttraumatic stress disorder symptoms and television viewing patterns in the Nurses' Health Study II: a longitudinal analysis. PLoS One 2019;14(3):e0213441 [FREE Full text] [doi: 10.1371/journal.pone.0213441] [Medline: 30897111]
3. Manousaki D, Barnett TA, Mathieu M, Maximova K, Simoneau G, Harnois-Leblanc S, QUALITY Cohort Collaborative Group. Tune out and turn in: the influence of television viewing and sleep on lipid profiles in children. Int J Obes (Lond) 2020 May;44(5):1173-1184. [doi: 10.1038/s41366-020-0527-5] [Medline: 32203106]
4. Golshevsky DM, Magnussen C, Juonala M, Kao K, Harcourt BE, Sabin MA. Time spent watching television impacts on body mass index in youth with obesity, but only in those with shortest sleep duration. J Paediatr Child Health 2020 May;56(5):721-726. [doi: 10.1111/jpc.14711] [Medline: 31821651]
5. Šmotek M, Fárková E, Manková D, Kopřivová J. Evening and night exposure to screens of media devices and its association with subjectively perceived sleep: should "light hygiene" be given more attention? Sleep Health 2020 Mar 17:A. [doi: 10.1016/j.sleh.2019.11.007] [Medline: 32197951]
6. Ryan T, Chester A, Reece J, Xenos S. The uses and abuses of Facebook: a review of Facebook addiction. J Behav Addict 2014 Sep;3(3):133-148. [doi: 10.1556/jba.3.2014.016]
7. von der Heiden JM, Braun B, Müller KW, Egloff B. The association between video gaming and psychological functioning. Front Psychol 2019;10:1731. [doi: 10.3389/fpsyg.2019.01731] [Medline: 31402891]
8. Chóliz M. The challenge of online gambling: the effect of legalization on the increase in online gambling addiction. J Gambl Stud 2016 Jun;32(2):749-756. [doi: 10.1007/s10899-015-9558-6] [Medline: 26058374]
9. Marino C, Gini G, Vieno A, Spada MM. A comprehensive meta-analysis on problematic Facebook use. Comput Hum Behav 2018 Jun;83:262-277. [doi: 10.1016/j.chb.2018.02.009]

10. Brailovskaia J, Velten J, Margaf J. Relationship between daily stress, depression symptoms, and Facebook addiction disorder in Germany and in the United States. Cyberpsychol Behav Soc Netw 2019 Sep;22(9):610-614. [doi: 10.1089/cyber.2019.0165] [Medline: 31397593]
11. Tsitsika AK, Tzavela EC, Janikian M, Ólafsson K, Iordache A, Schoenmakers TM, et al. Online social networking in adolescence: patterns of use in six European countries and links with psychosocial functioning. J Adolesc Health 2014 Jul;55(1):141-147. [doi: 10.1016/j.jadohealth.2013.11.010] [Medline: 24618179]
12. Lee DS, Orvell A, Briskin J, Shrapnell T, Gelman SA, Ayduk O, et al. When chatting about negative experiences helps-and when it hurts: distinguishing adaptive versus maladaptive social support in computer-mediated communication. Emotion 2020 Apr;20(3):368-375. [doi: 10.1037/emo0000555] [Medline: 30628816]
13. Song J, Song TM, Seo D, Jin D, Kim JS. Social big data analysis of information spread and perceived infection risk during the 2015 Middle East respiratory syndrome outbreak in South Korea. Cyberpsychol Behav Soc Netw 2017 Jan;20(1):22-29. [doi: 10.1089/cyber.2016.0126] [Medline: 28051336]
14. Griffiths MD, Szabo A. Is excessive online usage a function of medium or activity? J Behav Addict 2014 Mar;3(1):74-77. [doi: 10.1556/jba.2.2013.016]
15. Pontes HM, Szabo A, Griffiths MD. The impact of Internet-based specific activities on the perceptions of Internet addiction, quality of life, and excessive usage: a cross-sectional study. Addict Behav Rep 2015 Jun;1:19-25 [FREE Full text] [doi: 10.1016/j.abrep.2015.03.002] [Medline: 29531976]
16. Utz S, Breuer J. The relationship between use of social network sites, online social support, and well-being: results from a six-wave longitudinal study. J Media Psychol 2017;29(3):115-125 [FREE Full text] [doi: 10.1027/1864-1105/a000222] [Medline: 29147141]
17. Pretorius C, Chambers D, Cowan B, Coyle D. Young people seeking help online for mental health: cross-sectional survey study. JMIR Ment Health 2019 Aug 26;6(8):e13524 [FREE Full text] [doi: 10.2196/13524] [Medline: 31452519]
18. Souza-Talarico JN, Wan N, Santos S, Fialho PPA, Chaves EC, Caramelli P, et al. Cross-country discrepancies on public understanding of stress concepts: evidence for stress-management psychoeducational programs. BMC Psychiatry 2016 Jun 03;16:181 [FREE Full text] [doi: 10.1186/s12888-016-0886-6] [Medline: 27260184]
19. MASON JW. A review of psychoendocrine research on the pituitary-adrenal cortical system. Psychosom Med 1968;30(5):576-607. [doi: 10.1097/00006842-196809000-00020]
20. Lupien S, Ouellet-Morin I, Trépanier L, Juster R, Marin M, Francois N, et al. The DeStress for Success Program: effects of a stress education program on cortisol levels and depressive symptomatology in adolescents making the transition to high school. Neuroscience 2013 Sep 26;249:74-87. [doi: 10.1016/j.neuroscience.2013.01.057] [Medline: 23403176]
21. Folkman S, Moskowitz JT. Coping: pitfalls and promise. Annu Rev Psychol 2004;55:745-774. [doi: 10.1146/annurev.psych.55.090902.141456] [Medline: 14744233]
22. McEwen BS, Akil H. Revisiting the stress concept: implications for affective disorders. J Neurosci 2020 Jan 02;40(1):12-21. [doi: 10.1523/jneurosci.0733-19.2019]
23. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York: Springer; 1984:1-445.
24. Nasca C, Menard C, Hodes G, Bigio B, Pena C, Lorsch Z, et al. Multidimensional predictors of susceptibility and resilience to social defeat stress. Biol Psychiatry 2019 Sep 15;86(6):483-491. [doi: 10.1016/j.biopsych.2019.06.030] [Medline: 31466563]
25. Tweed RG, White K, Lehman DR. Culture, stress, and coping. J Cross-Cultural Psychol 2016 Jul 26;35(6):652-668. [doi: 10.1177/0022022104270109]
26. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. Gen Psych 2020 Mar 06;33(2):e100213. [doi: 10.1136/gpsych-2020-100213]
27. Khalili-Mahani N, Assadi A, Li K, Mirgholami M, Rivard M, Benali H, et al. Reflective and reflexive stress responses of older adults to three gaming experiences in relation to their cognitive abilities: mixed methods crossover study. JMIR Ment Health 2020 Mar 26;7(3):e12388 [FREE Full text] [doi: 10.2196/12388] [Medline: 32213474]
28. Hasebrink U, Popp J. Media repertoires as a result of selective media use. A conceptual approach to the analysis of patterns of exposure. Communications 2006;31(3):369-387. [doi: 10.1515/commun.2006.023]
29. Hasebrink U, Hepp A. How to research cross-media practices? Investigating media repertoires and media ensembles. Convergence 2017 Apr 07;23(4):362-377. [doi: 10.1177/1354856517700384]
30. SurveyMonkey. URL: https://www.surveymonkey.com/r/MediaCovid19
31. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. 2009 Presented at: 3rd International AAAI Conference on Weblogs and Social Media; 2009; St Jose, USA. [doi: 10.13140/2.1.1341.1520]
32. Khalili-Mahani N, De Schutter B. Affective game planning for health applications: quantitative extension of gerontoludic design based on the appraisal theory of stress and coping. JMIR Serious Games 2019 Jun 06;7(2):e13303 [FREE Full text] [doi: 10.2196/13303] [Medline: 31172966]
33. Bowey JT, Friehs MA, Mandryk RL. Red pill or blue pill. 2019 Presented at: 14th International Conference on the Foundations of Digital Games; 2019; San Luis Obispo, CA, USA p. 1-11. [doi: 10.1145/3337722.3337734]

34. Khalili-Mahani N, De Schutter B, Mirgholami M, Holowka EM, Goodine R, DeJong S, et al. For whom the games toll: a qualitative and intergenerational evaluation of what is serious in games for older adults. Comput Game J 2020 Mar 11;9(2):221-244. [doi: 10.1007/s40869-020-00103-7]
35. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
36. Castro D, Rigby JM, Cabral D, Nisi V. The binge-watcher's journey: investigating motivations, contexts, and affective states surrounding Netflix viewing. Convergence 2019 Dec 04:135485651989085. [doi: 10.1177/1354856519890856]
37. Kuss DJ, Kanjo E, Crook-Rumsey M, Kibowski F, Wang GY, Sumich A. Problematic mobile phone use and addiction across generations: the roles of psychopathological symptoms and smartphone use. J Technol Behav Sci 2018;3(3):141-149 [FREE Full text] [doi: 10.1007/s41347-017-0041-3] [Medline: 30238057]
38. de Hoog N, Verboon P. Is the news making us unhappy? The influence of daily news exposure on emotional states. Br J Psychol 2020 May;111(2):157-173 [FREE Full text] [doi: 10.1111/bjop.12389] [Medline: 30900253]
39. Marin M, Morin-Major J, Schramek TE, Beaupré A, Perna A, Juster R, et al. There is no news like bad news: women are more remembering and stress reactive after reading real negative news than men. PLoS One 2012;7(10):e47189 [FREE Full text] [doi: 10.1371/journal.pone.0047189] [Medline: 23071755]
40. Birk MV, Friehs MA, Madnryk RL. Age-based preferences and player experience. 2017 Presented at: The Annual Symposium on Computer-Human Interaction in Play; 2017; Amsterdam p. 157-170. [doi: 10.1145/3116595.3116608]
41. Saunders TJ, Vallance JK. Screen time and health indicators among children and youth: current evidence, limitations and future directions. Appl Health Econ Health Policy 2017 Jun;15(3):323-331. [doi: 10.1007/s40258-016-0289-3] [Medline: 27798796]
42. Saunders TJ, Chaput J, Tremblay MS. Sedentary behaviour as an emerging risk factor for cardiometabolic diseases in children and youth. Can J Diabetes 2014 Feb;38(1):53-61. [doi: 10.1016/j.jcjd.2013.08.266] [Medline: 24485214]

Abbreviations

COVID-19: coronavirus disease
EC: eigenvector centrality
MERS: Middle East respiratory syndrome

Edited by G Eysenbach; submitted 15.05.20; peer-reviewed by K Aguirre, M Friehs, C Escalona-Marfil; comments to author 13.07.20; revised version received 21.07.20; accepted 22.07.20; published 06.08.20.

Please cite as:
Pahayahay A, Khalili-Mahani N
What Media Helps, What Media Hurts: A Mixed Methods Survey Study of Coping with COVID-19 Using the Media Repertoire Framework and the Appraisal Theory of Stress
J Med Internet Res 2020;22(8):e20186
URL: https://www.jmir.org/2020/8/e20186
doi:10.2196/20186
PMID:32701459

Original Paper

Mental Health Disorders and Associated Risk Factors in Quarantined Adults During the COVID-19 Outbreak in China: Cross-Sectional Study

Yan Guo[1,2,3], PhD; Chao Cheng[4], MD; Yu Zeng[1], MB; Yiran Li[1], MB; Mengting Zhu[1], BS; Weixiong Yang[4], MD; He Xu[1], MB; Xiaohua Li[1], MB; Jinhang Leng[4], MD; Aliza Monroe-Wise[5], MD, MSc; Shaomin Wu[1], MB

[1]Department of Medical Statistics, School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]Sun Yat-sen Center for Migrant Health Policy, Guangzhou, China

[3]Sun Yat-sen Center for Global Health, Guangzhou, China

[4]Department of Thoracic Surgery, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[5]Department of Global Health, University of Washington, Seattle, WA, United States

Corresponding Author:
Chao Cheng, MD
Department of Thoracic Surgery
The First Affiliated Hospital of Sun Yat-sen University
No 58 2nd Zhongshan Road
Guangzhou, 510080
China
Phone: 86 87755766
Email: drchengchao@163.com

Abstract

Background: People undergoing mass home- and community-based quarantine are vulnerable to mental health disorders during outbreaks of coronavirus disease (COVID-19), but few studies have evaluated the associated psychosocial factors.

Objective: This study aimed to estimate the prevalence of anxiety and depressive symptoms and identify associated demographic and psychosocial factors in the general Chinese population during the COVID-19 pandemic quarantine period.

Methods: Participants aged 18 years or above were recruited in a cross-sectional online survey using snowball sampling from February 26-29, 2020. The survey included questions on demographics, family relationships, chronic diseases, quarantine conditions, lifestyle, COVID-19 infection, and anxiety and depressive symptoms. Logistic regression analyses were conducted to identify factors associated with elevated anxiety or depressive symptoms.

Results: Out of 2331 participants, 762 (32.7%) experienced elevated anxiety or depressive symptoms. Nine risk factors associated with anxiety or depressive symptoms included younger age, reduced income, having cancer or other chronic diseases, having family members living with cancer, concerns related to COVID-19 infection for themselves or family members, living alone, having family conflicts, having <3 or >8 hours of sedentary time per day, and worsened sleep quality.

Conclusions: The findings highlight an urgent need for psychological support for populations at high risk for elevated anxiety or depressive symptoms during the COVID-19 pandemic.

(J Med Internet Res 2020;22(8):e20328) doi:10.2196/20328

KEYWORDS

COVID-19; anxiety or depressive symptoms; quarantine; risk and protective factors

Introduction

Pandemics and other public health crises such as severe acute respiratory syndrome (SARS), Ebola, and the current coronavirus disease (COVID-19) pandemic often result in elevated rates of mental health problems [1-3]. Anxiety, depressive symptoms, and posttraumatic stress disorder (PTSD) are the most common mental health problems during pandemics [4]. To fight against the COVID-19 pandemic, quarantine strategies have been implemented in many countries including China [5,6]. However, large-scale and long-term quarantine may have negative impacts on people's mental health. A recent

study found that the prevalence of depressive symptoms was 20.1% in the Chinese population during the first month of widely implemented quarantine due to COVID-19 [7], which is much higher than previous reports of the lifetime rate of depressive symptoms in this population (6.8%), based on a representative sample [8]. However, empirical evidence regarding psychological disorders and related risk factors among people undergoing mass home- and community-based quarantine are still lacking.

Although several studies have reported on factors associated with mental health disorders during the COVID-19 pandemic, some factors particularly characteristic of the pandemic and corresponding quarantine strategy remain understudied. In the existing literature, factors such as demographic characteristics, lifestyle, and concerns about COVID-19 infection have been examined during the pandemic [7,9,10]. Specifically, reduced income, increased sedentary time, poor sleep quality, and concerns for COVID-19 infection were reported as risk factors for elevated anxiety or depressive symptoms [7,9,10]. Yet little attention has been paid to social factors such as an individual's household composition (living alone or with others) and family relationships, all of which may reflect people's social connectedness and affect their psychological well-being especially under social distancing or quarantine policies. During the COVID-19 quarantine period, a strict movement restriction was widely implemented in China [6]. People either spent more time with family members at home or experienced isolation during quarantine. The former scenario could lead to increased family conflicts during this hardship, and the latter may result in social isolation [11]. However, the potential risk factors for mental health symptoms arising from the COVID-19 pandemic have rarely been studied during the quarantine.

In addition, previous studies have shown mixed findings about the effects of some factors such as age on anxiety or depressive symptoms during the COVID-19 pandemic. One study reported that individuals between 18 and 30 years or above 60 years had the highest rates of psychological distress during the COVID-19 outbreaks [12], while another study found only younger people presented higher levels of stress and anxiety [13]. Two other studies found no association between age and mental health problems during the COVID-19 outbreaks [10,14].

Some populations such as people living with cancer or other chronic diseases may be at increased risk for developing mental health problems during a pandemic due to their tenuous physical health, barriers to accessing medical treatment, higher risks of COVID-19 infection, and higher probability of severe illness if infected [15-19]. During the COVID-19 quarantine period, patients with cancer or other chronic diseases may experience increased challenges related to receiving routine medical care due to mobility restrictions and the potential shortages of medical workers and essential medicines [20]. One study found that the likelihood of COVID-19 infection in cancer patients was twice as high as that in the general population [17]. In addition, the World Health Organization (WHO) reported that patients with pre-existing noncommunicable diseases, including cardiovascular disease, chronic respiratory disease, diabetes, and cancer, are at increased risk of severe illness from COVID-19 [21]. Therefore, the WHO has urged the global medical community to pay additional attention to mental health in this vulnerable population, especially during the pandemic [21]. Yet, studies on mental health problems in cancer patients or other chronic diseases are scarce.

To fill these gaps in the literature, the current study aimed to estimate the prevalence of anxiety or depressive symptoms and identify associated demographic and psychosocial factors in the general Chinese population during the COVID-19 pandemic quarantine in China in February 2020. Specifically, we examined the effects of associated factors including demographics, family relationships, chronic disease status, quarantine conditions, lifestyle, and COVID-19 infection. Special attention was paid to characteristics particularly relevant to the COVID-19 pandemic quarantine such as household composition, family conflict, and chronic disease status.

Methods

Participants aged 18 years or above were recruited through a snowball sampling process via WeChat from February 26-29, 2020 during the community- or home-based quarantine in China. WeChat is the most widely used social media platform in China with over 1 billion active users [22]. We developed an online questionnaire using Questionnaire Star, the link to which could be shared via WeChat. Clicking the survey link in WeChat took participants directly to the online questionnaire. The online survey link was initially and purposely sent to 10 participants who were chosen to ensure a broad representation of age, gender, educational level, and chronic diseases status (eg, with or without chronic diseases). We asked these participants to send the survey link to friends on their WeChat contact list whom they considered suitable for this survey, and their friends were also encouraged to send the link to their own WeChat contact networks. The snowball sampling process continued until a sufficient sample size was reached. To recruit cancer patients and their family members, oncologists sent the survey link to patient groups in WeChat and encouraged them to participate in the study. Participants anonymously completed the self-administered electronic questionnaire for about 15 minutes with no financial incentive. Participants were allowed to reaccess the survey link but doing so erased their previous data. Although the probability of repeated completion by the same participants cannot be ruled out, such instances were rare as they were asked to complete it once and there was no incentive for repeat submissions.

Data on demographics (eg, age, gender, marital status, educational status, household composition, individual income), family relationships (eg, family conflict), chronic diseases (eg, cancer, hypertension, diabetes, asthma, cerebrovascular or cardiovascular diseases), quarantine conditions (eg, quarantine duration, frequency of going out during the quarantine), lifestyle (eg, physical activity, sedentary behavior, sleep quality), COVID-19 infection (eg, confirmed or suspected cases among family members, friends, colleagues, or in the community), and anxiety and depressive symptoms were collected.

Anxiety and depressive symptoms were assessed by the 14-item Hospital Anxiety and Depression Scale (HADS) that comprises two 7-item subscales; participants with an Anxiety subscale

(HADS-A) score ≥8 or a Depression subscale (HADS-D) score ≥8 were classified as having elevated anxiety or depressive symptoms [23]. We chose HADS because not only does it measure anxiety and depressive symptoms simultaneously with good reliability and validity [24], but it is also easy to understand and brief, requiring minimal time to complete. In comparison, other tools such as the Center for Epidemiologic Studies Depression Scale (CESD) and the Patient Health Questionnaire-9 (PHQ-9) measure only depressive symptoms. In addition, the 20-item CESD is too long to complete, whereas PHQ-9 has one item that is particularly sensitive and unsuitable for the general population during the stressful time of the pandemic (ie, "Thoughts that you would be better off dead, or thoughts of hurting yourself in some way") [24]. Therefore, based on the feedback from the pilot study prior to the large-scale survey deployment, we chose to use HADS.

The sample size was calculated based on the estimated rate of anxiety or depressive symptoms in the general population. A pre–COVID-19 epidemiological study reported the lifetime prevalence of anxiety and depressive symptoms in the Chinese population as 7.6% and 6.8%, respectively [8]. Considering the likelihood of increased rates of anxiety and depressive symptoms during pandemics, we hypothesized a rate of 18% for anxiety or depressive symptoms in the general population. With a significance level set at .05, an absolute error of 1.8%, and allowing for up to 10% invalid questionnaires, we estimated a minimum sample size of 2004 for the current study.

Logistic regression models were employed to identify factors associated with elevated anxiety or depressive symptoms. All variables were categorized and described by frequencies and percentages. Univariate logistic regression models were used to analyze the distribution of anxiety or depressive symptoms among different categories for each variable and select risk factors for multivariate analysis. For example, to identify whether chronic disease status was a potential risk factor for anxiety or depressive symptoms, we compared the rate of anxiety or depressive symptoms among participants with chronic diseases to that of participants without. Variables with $P<.10$ in univariate analysis were included in the multivariate logistic regression model and those with $P<.05$ were retained in the final model. R software version 3.5.1 (R foundation for Statistical Computing) was used for data analyses.

This study was approved by the institutional review board of the School of Public Health, Sun Yat-sen University, and all participants provided informed consent.

Results

Out of 2441 questionnaires collected, 2331 (95.5%) were deemed valid using a quality-control question, which stated that its purpose was to screen out invalid questionnaires and requested participants to check a specific option. Questionnaires that checked the correct option were considered valid. The mean age of participants was 34.4 (SD 11.1) years, 56.1% (n=1307) were female, 60.0% (n=1398) were married, 73.7% (n=1718) had a bachelor's degree or above, 11.5% (n=269) had a chronic disease, 41.6%(n=970) had family members with a chronic disease, 44.7% (n=1041) had been quarantined for over 3 weeks, and 54.5% (n=1271) went out no more than once a week. In total, 32.7% (n=762) of participants experienced elevated anxiety or depressive symptoms. Specifically, 25.4% (n=592) experienced anxiety, 21.3% (n=496) experienced depressive symptoms, and 13.9% (n=326) experienced both anxiety and depressive symptoms.

Rates of elevated anxiety or depressive symptoms among people without chronic diseases, with cancer, and with other chronic diseases were 31.5% (n=649), 34.8% (n=32), and 45.8% (n=81), respectively. Rates of elevated anxiety or depressive symptoms among people whose family members had no chronic diseases, cancer, and other chronic diseases were 31.7% (n=431), 45.0% (n=36), and 33.1% (n=295), respectively.

As shown in Table 1, nine variables were retained in the multivariate logistic regression model, and the odds ratio (OR) of each variable was adjusted by the remaining eight variables in the model. Risk factors associated with elevated anxiety or depressive symptoms included younger age (<40 years) (41-55 years: OR 0.666, 95% CI 0.494-0.899; ≥56 years: OR 0.507, 95% CI 0.304-0.847), reduced income during quarantine (OR 1.441, 95% CI 1.179-1.760), having cancer (OR 1.900, 95% CI 1.146-3.149) or other chronic diseases (OR 2.222, 95% CI 0.304-0.847), having family members living with cancer (OR 1.821, 95% CI 1.119-2.964), concerns for COVID-19 infection for themselves or family members (moderate concern: OR 1.603, 95% CI 1.216-2.115; severe concern: OR 2.315, 95% CI 1.705-3.142), living alone (living with family or others: OR 0.595, 95% CI 0.422-0.839), having family conflicts during the COVID-19 outbreak (OR 1.707, 95% CI 1.267-2.299), having <3 (OR 1.702, 95% CI 1.407-2.059) or >8 (OR 1.458, 95% CI 1.193-1.783) hours of sedentary time per day, and worsened sleep quality (OR 2.917, 95% CI 2.319-3.670) during the quarantine.

Table 1. Logistic regression of associated factors for elevated anxiety or depressive symptoms during the quarantine (N=2331).

Variable	Total, N	Anxiety or depressive symptoms[a], n (%)	Unadjusted results		Adjusted results	
			Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Age (years)						
18-25	530	184 (34.7)	1 (ref[b])	N/A[c]	1 (ref)	N/A
26-40	1197	406 (33.9)	0.965 (0.778-1.197)	.75	0.836 (0.661-1.058)	.14
41-55	475	138 (29.1)	0.770 (0.590-1.006)	.06	0.666 (0.494-0.899)	.01
≥56	129	34 (26.4)	0.673 (0.438-1.035)	.07	0.507 (0.304-0.847)	.01
Change in income						
Unchanged	1571	467 (29.7)	1 (ref)	N/A	1 (ref)	N/A
Worse	737	290 (39.3)	1.534 (1.277-1.842)	<.001	1.441 (1.179-1.760)	<.001
Better	23	5 (21.7)	0.657 (0.242-1.779)	.41	0.558 (0.198-1.575)	.27
Has a chronic disease						
None	2062	649 (31.5)	1 (ref)	N/A	1 (ref)	N/A
Other chronic diseases	177	81 (45.8)	1.837 (1.347-2.505)	<.001	2.222 (1.556-3.172)	<.001
Cancer	92	32 (34.8)	1.161 (0.749-1.801)	.51	1.900 (1.146-3.149)	.01
Family members with chronic diseases[d]						
None	1361	431 (31.7)	1 (ref)	N/A	1 (ref)	N/A
Other chronic diseases	890	295 (33.1)	1.070 (0.893-1.281)	.46	0.983 (0.808-1.196)	.86
Cancer	80	36 (45.0)	1.765 (1.120-2.783)	.01	1.821 (1.119-2.964)	.02
Concerns about infection for themselves or family members						
Little	400	87 (21.8)	1 (ref)	N/A	1 (ref)	N/A
Moderate	1351	430 (31.8)	1.680 (1.290-2.187)	<.001	1.603 (1.216-2.115)	.001
Severe	580	245 (42.2)	2.631 (1.971-3.513)	<.001	2.315 (1.705-3.142)	<.001
Living situation						
Alone	168	71 (42.3)	1 (ref)	N/A	1 (ref)	N/A
With family or others	2163	691 (31.9)	0.641 (0.466-0.882)	.06	0.595 (0.422-0.839)	.003
Family conflict[e]						
No	1502	409 (27.2)	1 (ref)	N/A	1 (ref)	N/A
Yes	829	353 (42.6)	1.982 (1.658-2.369)	<.001	1.702 (1.407-2.059)	<.001
Sedentary behavior (hours/day)						
3-8	1150	313 (27.2)	1 (ref)	N/A	1 (ref)	N/A
<3	261	102 (39.1)	1.715 (1.296-2.271)	<.001	1.707 (1.267-2.299)	<.001
>8	920	347 (37.7)	1.619 (1.344-1.951)	<.001	1.458 (1.193-1.783)	<.001
Sleep quality during quarantine						
Unchanged	1448	384 (26.5)	1 (ref)	N/A	1 (ref)	N/A
Worse	451	249 (55.2)	3.416 (2.743-4.253)	<.001	2.917 (2.319-3.67)	<.001
Better	432	129 (29.9)	1.180 (0.931-1.495)	.17	1.043 (0.814-1.335)	.74
Region						
Other provinces	2134	677 (31.7)	1 (ref)	N/A	N/A	N/A
Hubei Province	197	85 (43.1)	1.633 (1.214-2.197)	.001	N/A	N/A

Variable	Total, N	Anxiety or depressive symptoms[a], n (%)	Unadjusted results		Adjusted results	
			Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Infection cases[f]						
None	2285	736 (32.2)	1 (ref)	N/A	N/A	N/A
Yes	46	26 (56.5)	2.736 (1.517-4.933)	.001	N/A	N/A
Duration of quarantine						
≤3 weeks	1290	414 (32.1)	1 (ref)	N/A	N/A	N/A
>3 weeks	1041	348 (33.4)	1.063 (0.893-1.264)	.49	N/A	N/A
Frequency of going out						
Never	462	152 (19.9)	1 (ref)	N/A	N/A	N/A
≤1 time per week	809	256 (33.6)	0.944 (0.740-1.205)	.64	N/A	N/A
2-4 times per week	769	261 (34.3)	1.048 (0.820-1.339)	.71	N/A	N/A
≥5 times per week	261	93 (12.2)	0.958 (0.700-1.311)	.79	N/A	N/A

[a]Participants with an HADS-A score ≥8 or an HADS-D score ≥8 were classified as those with elevated anxiety or depressive symptoms.

[b]ref: reference.

[c]N/A: not applicable.

[d]Chronic diseases included the following options: cancer, hypertension, diabetes, cerebrovascular disease (eg, stroke, cerebral infarction, cerebral hemorrhage), ischemic heart disease, chronic hepatitis/cirrhosis, chronic bronchitis/emphysema, asthma, rheumatoid arthritis, psychological disorders, and others specified by the participants.

[e]Family conflict refers to family disputes over COVID-19–related preventive measures (eg, frequency of hand washing), concerns about mutual transmission of coronavirus among family members, and different opinions on information related to COVID-19 (eg, origins of the outbreak) were the three main triggers of family conflict.

[f]Infection cases: whether there were confirmed or suspected cases among family members, friends, colleagues, or in the community.

Discussion

China's large-scale quarantine was implemented in late January 2020, resulting in weeks of social isolation. In our sampled population, 32.7% experienced elevated anxiety or depressive symptoms during the quarantine, with 25.4% and 21.3% experiencing anxiety and depressive symptoms, respectively. This is more than triple the rates of lifetime anxiety and depressive symptoms previously reported for the Chinese population (7.6% and 6.8%, respectively) [8]. In our survey, the rates of elevated anxiety or depressive symptoms among participants with cancer or chronic diseases were found to be 34.8% and 45.8%, respectively, which are higher than the rate of 31.5% identified among participants without chronic diseases during the COVID-19 quarantine.

Consistent with the existing literature, this study also found that financial loss, infection concerns, sedentary behavior, poor sleep quality, living with cancer or other chronic diseases, or having family members with cancer were associated with elevated anxiety or depressive symptoms [11,25-27]. Previous studies reported inconsistent results regarding the impact of age on anxiety or depressive symptoms during the COVID-19 outbreak. One study revealed a greater negative psychological impact for both young adults and the elderly [12], while another study found that only younger age was associated with increased stress [13]. In our study, we found that younger people were more likely to experience elevated anxiety or depressive symptoms than those above 40 years. A possible explanation for this result is that younger people might have heavier financial burdens and more access to information about the COVID-19 epidemic through social media, both of which could lead to increased stress [28]. Some studies have indicated that increased sedentary behavior was associated with anxiety and depressive symptoms during the COVID-19 pandemic [29]. Interestingly, our study found associations between both increased (>8 hours per day) and decreased sedentary time (<3 hours per day) and mental health symptoms. This could be explained by people who were working or taking care of sick family members suffering from intensified stress due to the likelihood of exposure to COVID-19. However, this finding needs further exploration.

This study found some unique risk factors associated with elevated anxiety or depressive symptoms in the Chinese population during the COVID-19 outbreak. We found that people who lived alone had an increased risk of anxiety or depressive symptoms, which might be due to diminished social interactions during the quarantine. Conversely, we also found that family conflict related to COVID-19 might be a source of stress contributing to mental health problems. In our survey, the rate of family conflict during the COVID-19 quarantine was found to be 36%, and most family conflict was related to COVID-19, such as disagreements over how family members should protect themselves from the pandemic and different opinions about information related to COVID-19. These results

provide important empirical evidence for policy making; tailored strategies are advisable for different populations at high risk of increased psychological distress and problems. Social connectedness and support should be promoted and provided, especially to those who live alone as an effort to protect them from social isolation. Information on COVID-19 and health promotion could be delivered more effectively through multiple channels to minimize family conflict and cultivate better mental health.

There are some limitations in this study. Our sample was more educated than the general Chinese adult population and may not be representative due to limitations of the sampling method used. However, since random sampling is difficult to achieve in a situation like the COVID-19 pandemic, social media–based sampling is a preferred alternative [30]. Though we oversampled cancer patients and their family members through the WeChat contact networks of doctors, the numbers were still small. Findings should be interpreted with caution. Future studies exploring interactions between or among risk factors may be helpful in providing guidance for vulnerable populations.

The results of this study highlight an urgent need for psychological support and counseling for populations at high risk for elevated anxiety or depressive symptoms during the current pandemic or any quarantine implementation. The WHO needs to urge the global medical community to provide screening for mental health problems and psychological services to vulnerable populations such as patients with cancer or other chronic diseases.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant numbers 71573290, 81572391). The authors would like to thank the patients, their family members, and research staff in the hospitals and research sites involved.

Authors' Contributions

YG and CC have full access to all of the data of this study and take responsibility for the integrity of the data and the accuracy of the data analysis. YG and CC were responsible for study concept and design; YZ, WY, MZ, YL, HX, XL, JL, and SW contributed to data acquisition, analysis, or interpretation of data. YG and CC drafted the manuscript. YG, CC, YZ, WY, YL, MZ, and AM-W critically revised the manuscript for important intellectual content. YZ, YL, and MZ conducted the statistical analysis. YG, CC, YZ, YL, MZ, and WY offered administrative, technical, or material support. YG and CC supervised the study.

Conflicts of Interest

None declared.

References

1. Reardon S. Ebola's mental-health wounds linger in Africa. Nature 2015 Mar 05;519(7541):13-14. [doi: 10.1038/519013a] [Medline: 25739606]
2. Maunder RG. Was SARS a mental health catastrophe? Gen Hosp Psychiatry 2009;31(4):316-317 [FREE Full text] [doi: 10.1016/j.genhosppsych.2009.04.004] [Medline: 19555790]
3. Mental health and psychosocial considerations during the COVID-19 outbreak. World Health Organization. 2020 Mar. URL: https://www.who.int/publications/i/item/mental-health-and-psychosocial-considerations-during-the-covid-19-outbreak [accessed 2020-06-23]
4. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. N Engl J Med 2020 May 13:e [FREE Full text] [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
5. Coronavirus disease 2019 (COVID-19) Situation Report - 53. World Health Organization. 2020 Mar. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200313-sitrep-53-covid-19.pdf?sfvrsn=adb3f72_2 [accessed 2020-03-14]
6. Guo Y, Li Y, Monroe-Wise A, Yeung SJ, Huang Y. A dynamic residential community-based quarantine strategy: China's experience in fighting COVID-19. Infect Control Hosp Epidemiol 2020 May 23:1 [FREE Full text] [doi: 10.1017/ice.2020.172] [Medline: 32321606]
7. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]
8. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. The Lancet Psychiatry 2019 Mar;6(3):211-224. [doi: 10.1016/S2215-0366(18)30511-X] [Medline: 30792114]
9. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: A systematic review and meta-analysis. Brain Behav Immun 2020 May 08:e [FREE Full text] [doi: 10.1016/j.bbi.2020.05.026] [Medline: 32437915]

10. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):e [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
11. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The Lancet 2020 Mar 14;395(10227):912-920. [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
12. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. Gen Psych 2020 Mar 06;33(2):e100213 [FREE Full text] [doi: 10.1136/gpsych-2020-100213] [Medline: 32215365]
13. Mazza C, Ricci E, Biondi S, Colasanti M, Ferracuti S, Napoli C, et al. A Nationwide Survey of Psychological Distress among Italian People during the COVID-19 Pandemic: Immediate Psychological Responses and Associated Factors. Int J Environ Res Public Health 2020 May 02;17(9):e [FREE Full text] [doi: 10.3390/ijerph17093165] [Medline: 32370116]
14. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
15. Al-Shamsi HO, Alhazzani W, Alhuraiji A, Coomes EA, Chemaly RF, Almuhanna M, et al. A Practical Approach to the Management of Cancer Patients During the Novel Coronavirus Disease 2019 (COVID-19) Pandemic: An International Collaborative Group. Oncologist 2020 Jun;25(6):e936-e945 [FREE Full text] [doi: 10.1634/theoncologist.2020-0213] [Medline: 32243668]
16. Wang H, Zhang L. Risk of COVID-19 for patients with cancer. The Lancet Oncology 2020 Apr;21(4):e181 [FREE Full text] [doi: 10.1016/S1470-2045(20)30149-2] [Medline: 32142621]
17. Yu J, Ouyang W, Chua MLK, Xie C. SARS-CoV-2 Transmission in Patients With Cancer at a Tertiary Care Hospital in Wuhan, China. JAMA Oncol 2020 Mar 25:e [FREE Full text] [doi: 10.1001/jamaoncol.2020.0980] [Medline: 32211820]
18. Zhang L, Zhu F, Xie L, Wang C, Wang J, Chen R, et al. Clinical characteristics of COVID-19-infected cancer patients: a retrospective case study in three hospitals within Wuhan, China. Ann Oncol 2020 Jul;31(7):894-901 [FREE Full text] [doi: 10.1016/j.annonc.2020.03.296] [Medline: 32224151]
19. Pellino G, Spinelli A. How Coronavirus Disease 2019 Outbreak Is Impacting Colorectal Cancer Patients in Italy: A Long Shadow Beyond Infection. Dis Colon Rectum 2020 Jun;63(6):720-722. [doi: 10.1097/DCR.0000000000001685] [Medline: 32384401]
20. Kang C, Yang S, Yuan J, Xu L, Zhao X, Yang J. Patients with chronic illness urgently need integrated physical and psychological care during the COVID-19 outbreak. Asian J Psychiatr 2020 May 07;51:102081 [FREE Full text] [doi: 10.1016/j.ajp.2020.102081] [Medline: 32289729]
21. Clinical management of COVID-19. World Health Organization. 2020 Mar. URL: https://www.who.int/publications/i/item/clinical-management-of-covid-19 [accessed 2020-06-23]
22. WeChat. Wikipedia. URL: https://zh.wikipedia.org/wiki/%E5%BE%AE%E4%BF%A1 [accessed 2020-03-23]
23. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jul;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
24. Smarr KL, Keefer AL. Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). Arthritis Care Res (Hoboken) 2011 Dec;63 Suppl 11:S454-S466 [FREE Full text] [doi: 10.1002/acr.20556] [Medline: 22588766]
25. Du J, Dong L, Wang T, Yuan C, Fu R, Zhang L, et al. Psychological symptoms among frontline healthcare workers during COVID-19 outbreak in Wuhan. Gen Hosp Psychiatry 2020 May 03:e [FREE Full text] [doi: 10.1016/j.genhosppsych.2020.03.011] [Medline: 32381270]
26. Vancampfort D, Firth J, Schuch FB, Rosenbaum S, Mugisha J, Hallgren M, et al. Sedentary behavior and physical activity levels in people with schizophrenia, bipolar disorder and major depressive disorder: a global systematic review and meta-analysis. World Psychiatry 2017 Oct;16(3):308-315 [FREE Full text] [doi: 10.1002/wps.20458] [Medline: 28941119]
27. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Res 2020 May;287:112934 [FREE Full text] [doi: 10.1016/j.psychres.2020.112934] [Medline: 32229390]
28. Cheng C, Jun H, Liang B. Psychological health diathesis assessment system: a nationwide survey of resilient trait scale for Chinese adults. Stud Psychol Behav 2014;12:735-742 [FREE Full text]
29. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental Health and Behavior of College Students During the Early Phases of the COVID-19 Pandemic: Longitudinal Smartphone and Ecological Momentary Assessment Study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
30. Cohen N, Arieli T. Field research in conflict environments: Methodological challenges and snowball sampling. Journal of Peace Research 2011 Jul 28;48(4):423-435. [doi: 10.1177/0022343311405698]

Abbreviations

CESD: Center for Epidemiologic Studies Depression Scale
COVID-19: coronavirus disease
HADS: Hospital Anxiety and Depression Scale
HADS-A: Anxiety subscale of Hospital Anxiety and Depression Scale
HADS-D: Depression subscale of Hospital Anxiety and Depression Scale
PHQ-9: Patient Health Questionnaire-9
PTSD: posttraumatic stress disorder
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by G Eysenbach; submitted 16.05.20; peer-reviewed by C Kaylor-Hughes, M Lewia; comments to author 19.06.20; revised version received 03.07.20; accepted 26.07.20; published 06.08.20.

Please cite as:
Guo Y, Cheng C, Zeng Y, Li Y, Zhu M, Yang W, Xu H, Li X, Leng J, Monroe-Wise A, Wu S
Mental Health Disorders and Associated Risk Factors in Quarantined Adults During the COVID-19 Outbreak in China: Cross-Sectional Study
J Med Internet Res 2020;22(8):e20328
URL: http://www.jmir.org/2020/8/e20328/
doi:10.2196/20328
PMID:32716899

Original Paper

Impact of the COVID-19 Pandemic on Partner Relationships and Sexual and Reproductive Health: Cross-Sectional, Online Survey Study

Guanjian Li[1], MD; Dongdong Tang[1], MD; Bing Song[2], MD, PhD; Chao Wang[2], MD; Shen Qunshan[2], MD; Chuan Xu[2], MD; Hao Geng[1], MD; Huan Wu[2], MD, PhD; Xiaojin He[1*], MD, PhD; Yunxia Cao[1*], MD, PhD

[1]Reproductive Medicine Center, Department of Obstetrics and Gynecology, The First Affiliated Hospital of Anhui Medical University, National Health Commission Key Laboratory of Study on Abnormal Gametes and Reproductive Tract, Hefei, China

[2]Key Laboratory of Population Health Across Life Cycle, Ministry of Education of the People's Republic of China, Anhui Province Key Laboratory of Reproductive Health and Genetics, Hefei, China

[*]these authors contributed equally

Corresponding Author:
Yunxia Cao, MD, PhD
Reproductive Medicine Center, Department of Obstetrics and Gynecology
The First Affiliated Hospital of Anhui Medical University
National Health Commission Key Laboratory of Study on Abnormal Gametes and Reproductive Tract
81 Meishan Road
Hefei, 230000
China
Phone: 86 15395104659
Fax: 86 055165161058
Email: caoyunxia6@126.com

Abstract

Background: In the past few months, the coronavirus disease (COVID-19) pandemic has caused extensive economic and social damage.

Objective: The purpose of this study was to assess the impact of COVID-19–related measures on partner relationships and sexual and reproductive health in China.

Methods: From May 1 to 5, 2020, 3500 young Chinese individuals were recruited through WeChat or Weibo to participate in a survey to obtain information on sexual and reproductive health (eg, sexual desire, frequency of sexual intercourse, sexual satisfaction, etc). The questionnaire also collected demographic data (eg, age, race, education, current financial status, sexual orientation, relationship status, etc).

Results: In total, 967 participants were included in the sexual health analysis. Due to the COVID-19 pandemic and related containment measures, 22% of participants (n=212) reported a decrease in sexual desire; 41% (n=396) experienced a decrease in the sexual intercourse frequency; 30% (n=291) reported an increase in the frequency of masturbation; 20% (n=192) reported a decrease in alcohol consumption before or during sexual activities, and 31% (n=298) reported a deterioration in partner relationships during the pandemic. The logistic regression analysis indicated that the following influenced partner relationships: accommodations during the pandemic (P=.046; odds ratio [OR] 0.59; 95% CI 0.30-0.86); exclusive relationship status (yes or no) (P<.001; OR 0.44; 95 % CI 0.27-0.73); sexual desire (P=.02; OR 2.01; 95% CI 1.38-2.97); and sexual satisfaction (P<.001; OR 1.92; 95% CI 1.54-2.50). COVID-19 also caused disruptions in reproductive health services such as prenatal and postnatal care, childbirth and abortion services, contraception availability, and the management of sexually transmitted infections.

Conclusions: Our results show that many young people have wide-ranging issues affecting their sexual and reproductive health due to the COVID-19 pandemic and related containment measures. Strategies and guidelines are needed to safeguard the sexual and reproductive health of young people during this pandemic.

(J Med Internet Res 2020;22(8):e20961) doi:10.2196/20961

KEYWORDS

COVID-19; survey; novel coronavirus; sexual behavior; sexual health; reproductive health; young adults; youth; China

Introduction

In the past few months, the coronavirus disease (COVID-19) pandemic and related containment measures have caused extensive economic and social damage to many countries like China and the United States [1-3].

From February to April 2020, all schools and many businesses in China were closed. Everyone was required to implement "social distancing," and the government restricted travel, parties, and outdoor activities [4,5]. In addition, many provinces have implemented guidelines to reduce pressure on the health care system, including the suspension of nonemergency medical care and elective surgeries [6]. These disruptions have had a significant impact on the physical and mental health, as well as quality of life, of individuals [7,8].

So far, there is little report about the effects of COVID-19 on sexual and reproductive health [9,10]. Sexually active young people are facing more and more health challenges globally [11,12], and various aspects of their reproductive and sexual health may be affected by COVID-19. On the one hand, many young people are facing economic and psychological pressures caused by job loss or suspension of schooling. On the other hand, separation from sexual partners and lack of access to comprehensive health care services may be increasing the risk of experiencing negative sexual health outcomes.

Network-based sexual health risk assessments have been considered acceptable, and internet technology has become a powerful tool to promote health [13,14]. To assess the impact of COVID-19 pandemic and related containment measures on partner relationships and sexual and reproductive health, we conducted a series of preliminary analyses using data from an internet-based survey among Chinese youths and young adults.

Methods

Participant Recruitment

We conducted a cross-sectional, online survey using Questionnaire Star, an online questionnaire survey platform with 82,000,000 users in China. A 20-item survey (Multimedia Appendix 1) was constructed to assess changes in people's sexual and reproductive health during the COVID-19 pandemic. Between May 1 and 5, 2020, 3500 participants in China received links and emails from WeChat or Weibo (similar to WhatsApp and Twitter, respectively), inviting them to participate in a confidential, 20-minute, online survey about sexual and reproductive health on the Questionnaire Star platform.

Data Collection

Multiple reminders were sent via messaging software, and the invitation letter stated that the current survey was for participants who were sexually active. Approximately US $5 was offered as an incentive for participants who completed the questionnaire. Duplicate entries were prevented by restricting users with the same IP (Internet Protocol) address from accessing the survey more than once. A missed answer reminder component prompted participants about unfinished questionnaires in real time, and incomplete questionnaires were not submitted to the system. In addition, participants were unable to submit the questionnaire if their total answering time was less than 3 minutes.

Before entering the online survey system, all participants reviewed and approved the electronic consent page. This research was approved by the Ethics Review Committee of Anhui Medical University for research and publication purposes.

The questionnaire collected data on age, race, education, current financial status, SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) infection status of participants and those around them, self-rated health, accommodations before and during the pandemic, medical and surgical history of participants and their partners, pregnancy (if female), abortion history, sexual orientation, relationship status (ie, exclusive partnership or not), sexual desire, frequency of sexual intercourse, sexual satisfaction, alcohol consumption before or during sexual activities, frequency of masturbation, frequency of use of pornographic content, proportion of condom use (in sexual encounters), risky sexual behavior, and presence of sexually transmitted infections (STIs). Symptoms were assessed for during the COVID-19 pandemic. Before the questionnaire was distributed, 10 college students and 2 professors participated in a pilot study to verify and modify the contents of the questionnaire.

Participants were included in the current analysis if they were 15-35 years old, live in China, and reported penetrative sex (defined as insertion of penis into vaginal or anal orifices) at least once at any time in the past 6 months (n=1076). Of those invited, 35.7% (n=1249) of individuals completed the survey. Foreigners (n=11); homosexual or bisexual individuals (n=16); people, or people with partners, who had COVID-19 (n=0) or had contact with someone with COVID-19 (n=2); and those with systemic diseases (n=23) and other serious conditions that can cause sexual dysfunction (n=10) were excluded since their experiences may differ, and their representation in the sample was small.

Due to the complexity of physiological conditions and interference factors, people with STIs (n=8) and those who were pregnant (n=25) or had undergone a recent abortion (n=6) were also excluded from the sexual health analysis, but the COVID-19–related impact on several aspects of their reproductive health and rights were measured. Participants who were pregnant were asked "Have you experienced any difficulties in obtaining maternal care or delivery services due to COVID-19 or the plans to manage it?". Participants who reported a recent abortion were asked "Have you experienced any difficulties in obtaining abortion or post-abortion care due to COVID-19 or the plans to manage it?". Participants who reported STIs were asked "Have you experienced any difficulties in obtaining medical advice or management due to COVID-19 or the plans to manage it?". All participants who answered "yes" were invited to provide a description of their difficulties. Apart

from that, all participants were asked to fill in a response to "Have you experienced a shortage of contraceptives during the pandemic?".

Results

User Statistics

A total of 967 participants were included in the sexual health analysis. The characteristics of the study participants are shown in Table 1. The mean age was 26.6 (SD 4.86) years (range 16 to 35 years), and 55.9% (n=541) were male. All participants were Han Chinese. Of the 967 participants, almost half (n=416, 43%) reported a recent deterioration in their financial situation, and 8% (n=75) reported a poor state of health.

Table 1. Participants' demographic characteristics (N=967).

Characteristic	Total (N=967), n (%)	Male (n=541), n (%)	Female (n=426), n (%)	*F*	*P* value
Age (year)				2.26	.13
15-25	389 (40)	229 (43)	160 (36)		
25-35	578 (60)	312 (57)	266 (64)		
Education level				3.10	.21
College or below	491 (47)	263 (45)	228 (50)		
Bachelor	405 (42)	240 (44)	165 (40)		
Master or above	71 (10)	38 (11)	33 (10)		
Current financial situation				21.15	<.001
Fine	221 (23)	134 (25)	87 (20)		
Unchanged	330 (34)	151 (28)	179 (42)		
Deteriorated	416 (43)	256 (47)	160 (38)		
Self-rated health				2.76	.25
Fine	282 (29)	154 (28)	128 (30)		
General	610 (63)	351 (65)	259 (61)		
Poor	75 (8)	36 (7)	39 (9)		
Accommodation (before pandemic)				0.51	.77
Campus dormitory	416 (43)	229 (42)	187 (44)		
House with parents	348 (36)	200 (37)	148 (35)		
House without parents	203 (21)	112 (21)	91 (21)		
Accommodation (during pandemic)				1.37	.24
House with parents	706 (73)	403 (74)	303 (71)		
House without parents	261 (27)	138 (26)	123 (29)		

COVID-19–Related Impact on Sexual Health

COVID-19–related impact on sexual health is summarized in Table 2. In all, 68.8% (n=665) of students included in the analysis reported that they were currently in an exclusive relationship. There were significant differences in sexual health and outcomes between students in an exclusive relationship compared to those who were not in an exclusive relationship.

Of the 967 participants included in the analysis, 22% (n=212) reported a decrease in sexual desire, 41% (n=396) experienced a decrease in the frequency of sex, 20% (n=192) reported a recent decrease in alcohol consumption before or during sexual activities, and 10% (n=94) reported a decrease in risky sexual behavior. In addition, 31% (n=298) reported partner relationship deterioration during the pandemic (Table 2). With regard to the frequency of masturbation, 30% (n=291) of participants reported an increase in masturbation during the pandemic, while 23% (n=227) reported an increase in the use of pornography.

The logistic regression analysis indicated that accommodations during the pandemic, exclusive relationship status, sexual desire, and sexual satisfaction were closely related to partner relationships (Table 3).

Table 2. Coronavirus disease (COVID-19)–related impact on sexual health (N=967).

Items	Total (N=967)	In an exclusive relationship (n=665), n (%)	Not in an exclusive relationship[a] (n=302), n (%)	*F*	*P* value
Partner relationship				40.76	<.001
Fine	205 (21)	133 (20)	72 (24)		
General	464 (48)	285 (43)	179 (59)		
Deteriorated	298 (31)	247(37)	51 (17)		
Sexual desire				42.52	<.001
Fine	126 (13)	86 (13)	40 (13)		
General	629 (65)	395 (59)	234 (77)		
Deteriorated	212 (22)	184 (28)	28 (09)		
Sexual frequency				153.38	<.001
Increased	223 (23)	175 (26)	48 (16)		
Unchanged	348 (36)	304 (46)	44 (15)		
Decreased	396 (41)	186 (28)	210 (70)		
Sexual satisfaction				33.19	<.001
Increased	115 (12)	54 (8)	61 (20)		
Unchanged	709 (73)	498 (75)	211 (70)		
Decreased	143 (15)	113 (17)	30 (10)		
Consumed alcohol before or during sexual activities				132.01	<.001
Increased	58 (6)	44 (7)	14 (5)		
Unchanged	717 (74)	555 (83)	162 (54)		
Decreased	192 (20)	66 (10)	126 (42)		
Frequency of masturbation				21.99	<.001
Increased	291 (30)	189 (28)	102 (34)		
Unchanged	261 (27)	206 (31)	55 (18)		
Decreased	106 (11)	78 (12)	28 (9)		
None	309 (32)	192 (29)	117 (39)		
Frequency of pornography use				30.71	<.001
Increased	227 (23)	123 (19)	104 (34)		
Unchanged	330 (34)	242 (36)	88 (29)		
Decreased	115 (12)	89 (13)	26 (9)		
None	295 (31)	211 (32)	84 (28)		
Proportion of condom use				5.25	.07
Increased	97 (10)	61 (9)	36 (12)		
Unchanged	735 (76)	501 (75)	234 (77)		
Decreased	135 (14)	103 (16)	32 (11)		
Risky sexual behaviors				18.87[b]	<.001
Increased	4 (1)	3 (1)	1 (1)		
Unchanged	76 (8)	55 (8)	21 (7)		
Decreased	94 (010)	46 (7)	48 (16)		
None	793 (82)	561 (84)	232 (77)		

[a]Chi-square test was performed between the exclusive relationship group and nonexclusive relationship group.

[b]The “increased” data were merged with the “unchanged” data.

Table 3. Risk factors related to partner relationships determined by the logistic regression analysis.

Variable	Univariate analysis	Multivariate analysis		
	P value	OR[a]	*P* value	95% CI
Age (year)	.03	1.38	.41	1.13-1.80
Education level	.66	—[b]	—	—
Current financial situation	.16	—	—	—
Self-rated health	.02	1.85	.19	1.37-2.63
Accommodations (before pandemic)	.21	—	—	—
Accommodations (during pandemic)	.02	0.59	.046	0.30-0.86
In/not in an exclusive relationship	<.001	0.44	<.001	0.27-0.73
Sexual desire	.002	2.01	.02	1.38-2.97
Sexual satisfaction	<.001	1.92	<.001	1.54-2.50
Alcohol consumption before or during sexual activities	.36	—	—	—
Frequency of masturbation	.005	0.63	.21	0.46-0.88
Frequency of pornography use	.29	—	—	—
Proportion of condom use	.52	—	—	—
Risky sexual behaviors	.41	—	—	—

[a]OR: odds ratio.

[b]Not applicable.

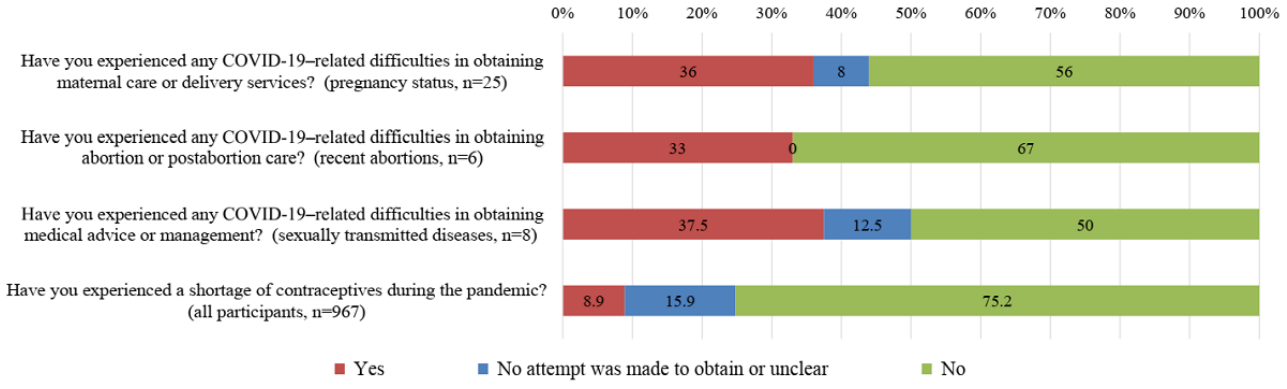
COVID-19–Related Impact on Reproductive Health

COVID-19–related impact on reproductive health is shown in Figure 1. Nine participants who were pregnant reported having difficulties accessing maternal care or delivery services due to COVID-19 or related measures. These difficulties included shortages in the number of hospital beds available for childbirth, restrictions on the number of people accompanying pregnant women for examination or hospitalization, and failure to receive timely prenatal examination.

Participants who reported a recent abortion described difficulties primarily with making appointments to see a doctor or for surgeries. Three participants who reported STIs experienced difficulties with medical management, mainly booking a doctor's appointment and accessing medicines such as antibiotics.

In the current study, the proportion of condom usage was unchanged due to COVID-19. However, 8.9% (n=86) of participants said they had experienced a shortage of contraceptives.

Figure 1. Self-reported coronavirus disease (COVID-19)–related impact among respondents.



Discussion

Principal Findings

Our study provides preliminary evidence on the direct impact of the COVID-19 pandemic and related containment measures on partner relationships and sexual and reproductive health. The results show that many young people had decreased sexual desire and frequency of sexual intercourse due to COVID-19. In addition, a relatively large number of participants reported a significant reduction in alcohol-related sexual consequences and risky sexual behavior. Increased family supervision or interference, less personal freedom overall, and poor mental health and partner relationships are likely contributors to these changes in sexual behavior.

We found that many participants reported an increase in masturbation frequency and use of pornography. Although masturbation may have helped some people achieve sexual satisfaction without the risk of SARS-CoV-2 infection, a high masturbation rate is related to a decrease in quality of life and sexual satisfaction [15]. High-frequency pornography use may also negatively impact sexual function and quality of life [16].

Our research also provides preliminary evidence of interruptions in reproductive health services due to COVID-19, such as prenatal and postnatal examination, delivery and abortion services, contraception availability, and STI management. In addition, we found that even in a country with a sound drug supply system such as China, contraceptives in some areas were out of stock or in short supply during the pandemic.

Limitations

The limitations of the current study include the use of a self-designed questionnaire and the reliance upon self-reporting in the midst of the constantly changing prevalence of COVID-19. As a result, our findings are based on cross-sectional data from local convenient samples. In addition, the impact of the COVID-19 pandemic on the sexual health of special groups, including lesbian, gay, bisexual, and transgender people, and people living with HIV, has not been reported in the current research. Further large-scale longitudinal studies are needed to understand the impact of the pandemic on sexual and reproductive health in different regions and populations.

Conclusions

The COVID-19 pandemic and related containment measures affect young people's sexual health, and targeted interventions are needed to improve health and well-being. In addition, since society is currently focusing on COVID-19 response, basic reproductive health services and supply chain operations have been disrupted. Such services should be protected from disruption and be delivered during the pandemic.

Acknowledgments

The authors acknowledge all participants and health care providers who participated in this study. This work was supported by the Non-Profit Central Research Institute Fund of the Chinese Academy of Medical Sciences (grant number 2019PT310002).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire (English version).
[DOCX File , 25 KB - jmir_v22i8e20961_app1.docx]

References

1. McKee M, Stuckler D. If the world fails to protect the economy, COVID-19 will damage health not just now but also in the future. Nat Med 2020 May;26(5):640-642. [doi: 10.1038/s41591-020-0863-y] [Medline: 32273610]
2. CDC COVID-19 Response Team. Geographic Differences in COVID-19 Cases, Deaths, and Incidence - United States, February 12-April 7, 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 17;69(15):465-471 [FREE Full text] [doi: 10.15585/mmwr.mm6915e4] [Medline: 32298250]
3. Singh RK, Rani M, Bhagavathula AS, Sah R, Rodriguez-Morales AJ, Kalita H, et al. Prediction of the COVID-19 Pandemic for the Top 15 Affected Countries: Advanced Autoregressive Integrated Moving Average (ARIMA) Model. JMIR Public Health Surveill 2020 May 13;6(2):e19115 [FREE Full text] [doi: 10.2196/19115] [Medline: 32391801]
4. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. JAMA 2020 Apr 10:e [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]
5. Huang Y, Wu Q, Wang P, Xu Y, Wang L, Zhao Y, et al. Measures Undertaken in China to Avoid COVID-19 Infection: Internet-Based, Cross-Sectional Survey Study. J Med Internet Res 2020 May 12;22(5):e18718 [FREE Full text] [doi: 10.2196/18718] [Medline: 32396516]
6. Schinköthe T, Gabri MR, Mitterer M, Gouveia P, Heinemann V, Harbeck N, et al. A Web- and App-Based Connected Care Solution for COVID-19 In- and Outpatient Care: Qualitative Study and Application Development. JMIR Public Health Surveill 2020 Jun 01;6(2):e19033 [FREE Full text] [doi: 10.2196/19033] [Medline: 32406855]
7. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May 30;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
8. Ni MY, Yang L, Leung CMC, Li N, Yao XI, Wang Y, et al. Mental Health, Risk Factors, and Social Media Use During the COVID-19 Epidemic and Cordon Sanitaire Among the Community and Health Professionals in Wuhan, China: Cross-Sectional Survey. JMIR Ment Health 2020 May 12;7(5):e19009 [FREE Full text] [doi: 10.2196/19009] [Medline: 32365044]

9. Yuksel B, Ozgor F. Effect of the COVID-19 pandemic on female sexual behavior. Int J Gynaecol Obstet 2020 Jul 11;150(1):98-102. [doi: 10.1002/ijgo.13193] [Medline: 32392400]
10. Jacob L, Smith L, Butler L, Barnett Y, Grabovac I, McDermott D, et al. Challenges in the Practice of Sexual Medicine in the Time of COVID-19 in the United Kingdom. J Sex Med 2020 Jul 14;17(7):1229-1236 [FREE Full text] [doi: 10.1016/j.jsxm.2020.05.001] [Medline: 32411271]
11. Tzilos Wernette G, Countryman K, Khatibi K, Riley E, Stephenson R. Love My Body: Pilot Study to Understand Reproductive Health Vulnerabilities in Adolescent Girls. J Med Internet Res 2020 Mar 30;22(3):e16336 [FREE Full text] [doi: 10.2196/16336] [Medline: 32224484]
12. Nielsen A, Bågenholm A, De Costa A. Development of a Mobile Phone App to Promote Safe Sex Practice Among Youth in Stockholm, Sweden: Qualitative Study. JMIR Form Res 2020 Jan 28;4(1):e12917 [FREE Full text] [doi: 10.2196/12917] [Medline: 32012038]
13. Tebb KP, Leng Trieu S, Rico R, Renteria R, Rodriguez F, Puffer M. A Mobile Health Contraception Decision Support Intervention for Latina Adolescents: Implementation Evaluation for Use in School-Based Health Centers. JMIR Mhealth Uhealth 2019 Mar 14;7(3):e11163 [FREE Full text] [doi: 10.2196/11163] [Medline: 30869649]
14. Lykens J, Pilloton M, Silva C, Schlamm E, Wilburn K, Pence E. Google for Sexual Relationships: Mixed-Methods Study on Digital Flirting and Online Dating Among Adolescent Youth and Young Adults. JMIR Public Health Surveill 2019 May 16;5(2):e10695 [FREE Full text] [doi: 10.2196/10695] [Medline: 31099335]
15. Brody S, Costa RM. Satisfaction (sexual, life, relationship, and mental health) is associated directly with penile-vaginal intercourse, but inversely with other sexual behavior frequencies. J Sex Med 2009 Jul;6(7):1947-1954. [doi: 10.1111/j.1743-6109.2009.01303.x] [Medline: 19453891]
16. Bőthe B, Tóth-Király I, Potenza MN, Orosz G, Demetrovics Z. High-Frequency Pornography Use May Not Always Be Problematic. J Sex Med 2020 Apr;17(4):793-811 [FREE Full text] [doi: 10.1016/j.jsxm.2020.01.007] [Medline: 32033863]

Abbreviations

COVID-19: coronavirus disease
IP: Internet Protocol
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 02.06.20; peer-reviewed by S Yang, B Ma; comments to author 30.06.20; revised version received 01.07.20; accepted 22.07.20; published 06.08.20.

Please cite as:
Li G, Tang D, Song B, Wang C, Qunshan S, Xu C, Geng H, Wu H, He X, Cao Y
Impact of the COVID-19 Pandemic on Partner Relationships and Sexual and Reproductive Health: Cross-Sectional, Online Survey Study
J Med Internet Res 2020;22(8):e20961
URL: https://www.jmir.org/2020/8/e20961
doi:10.2196/20961
PMID:32716895

Viewpoint

Telehealth for Noncritical Patients With Chronic Diseases During the COVID-19 Pandemic

Na Liu[1], PhD; Robin Huang[2], MPhil; Tanya Baldacchino[3], BNurs; Archana Sud[4,5], MBBS, MD, FRACP; Kamal Sud[6,7], FRACP; Mohamed Khadra[8], BMed, MEd, PhD, FAICD, FRACS; Jinman Kim[2,3], PhD

[1]The University of Sydney Business School, Sydney, Australia

[2]School of Computer Science, The University of Sydney, NSW, Australia

[3]Telehealth and Technology Centre, Nepean Blue Mountains Local Health District, NSW, Australia

[4]Department of Infectious Diseases, Nepean Hospital, NSW, Australia

[5]Department of Medicine (Infectious Diseases and Immunology), Nepean Clinical School, University of Sydney, NSW, Australia

[6]Nepean Clinical School, Faculty of Medicine and Health, The University of Sydney, NSW, Australia

[7]Department of Renal Medicine, Nepean Hospital, Kingswood, NSW, Australia

[8]Strategy and Innovation, Nepean Blue Mountains Local Health District, NSW, Australia

Corresponding Author:
Na Liu, PhD
The University of Sydney Business School
Abercrombie Building (H70)
Sydney, 2006
Australia
Phone: 61 2 86274277
Email: liu.na@sydney.edu.au

Abstract

During the recent coronavirus disease (COVID-19) pandemic, telehealth has received greater attention due to its role in reducing hospital visits from patients with COVID-19 or other conditions, while supporting home isolation in patients with mild symptoms. The needs of patients with chronic diseases tend to be overlooked during the pandemic. With reduced opportunities for routine clinic visits, these patients are adopting various telehealth services such as video consultation and remote monitoring. We advocate for more innovative designs to be considered to enhance patients' feelings of "copresence"—a sense of connection with another interactant via digital technology—with their health care providers during this time. The copresence-enhanced design has been shown to reduce patients' anxiety and increase their confidence in managing their chronic disease condition. It has the potential to reduce the patient's need to reach out to their health care provider during a time when health care resources are being stretched.

(J Med Internet Res 2020;22(8):e19493) doi:10.2196/19493

KEYWORDS

telehealth; chronic diseases; COVID-19; coronavirus; pandemic; remote monitoring; copresence

Introduction

Given the current coronavirus disease (COVID-19) outbreak, telehealth is being adopted widely in innovative ways, so that patients can access care while maintaining social distancing without risking exposure to and spread of COVID-19. Telehealth also promotes the safety of essential clinical staff. Additionally, using telehealth can avoid unnecessary hospital visits, thereby providing some relief to the currently overstretched health resources [1]. With modern telehealth services, patients can arrange a video consultation (VC) with their general practitioner (GP) or specialists instead of traveling to the clinics [2]. Remote patient monitoring (RPM) and patient engagements are also major components of telehealth, which allow patients to share their health-related data with their health care workers in real time [3] and be involved in shared decision making with improved health literacy (Muscat et al, unpublished data) [4]. One of core enablers of telehealth is the provision of timely intervention to patients based on their real-time health indicators (eg, from RPM) before their next scheduled formal visit or via VC service.

As health care resources may be limited during the pandemic, the allocation of resources will focus on managing acutely sick patients with COVID-19 rather than non–COVID-19 patients [5]. During the pandemic, patients with chronic conditions, such

as obesity, diabetes, cardiovascular disease, and kidney disease, are at a higher risk of developing serious complications due to COVID-19 and dying [6]; there are also concerns that missing usual ongoing care can lead to adverse health outcomes.

Another concern is that patients with chronic conditions may be fearful of accessing their usual health care services in order to minimize their risk of infection and complications that may arise from infection [7]. Consequently, patients with chronic conditions in self-isolation are not attending clinic visits that not only provide health care services but also serve as a source of motivation for maintaining adherence to medications and behavior changes, such as diet and exercise, required to treat the chronic condition. With reduced clinic visits, patients' communication channels with clinicians and nursing staff are cut off, thereby compromising the gains that were obtained with such interactions before the COVID-19 pandemic.

Telehealth for patients with chronic diseases during the pandemic is increasingly important, and we suggest that the innovative adoption of digital technologies can continue to provide valuable patient-clinician communication, not only for clinical care but also for maintaining adherence and behavior changes in patients. We suggest that care providers consider the inclusion of the "copresence" concept, which can be used to enhance the perception of presence and thus the relationship between patients and clinicians in the telehealth era. Copresence refers to a sense of connection with another interactant [8]. It is the perception by a communicator that another person in a mediated or online environment is real, immediate, or present [9]. When a person has a high copresence with others in a digitally mediated communication setting, he or she tends to feel more satisfied with the medium and is more likely to use digital tools. The feeling of copresence does not have to be built through rich or advanced technologies; it can be achieved by building up social cues for the interactions. Copresence was widely studied in the field of human-computer interactions, and its application has been used in the context of online shopping experience [10] and virtual team collaborations [11]. Copresence, when incorporated, can provide patients with the feeling that the clinicians are virtually with them, while not operationally increasing the workload of the clinical staff. We provide three copresence strategies that can be readily applied to ease the burden on care delivery and challenges to patients with chronic diseases during the pandemic.

Use of Telehealth for Patients With Chronic Diseases During the Pandemic

During the pandemic, telehealth can be used to provide more than conventional VC and RPM by building an efficient copresence-enhanced approach to reduce patients' anxiety while their face-to-face meeting with GPs are reduced. Additionally, it could help relieve stress among health care workers by reducing the need to constantly monitor patients' health-related data. During the pandemic, where health care resources are centered around the diagnosis and treatment of COVID-19 patients, it leaves patients with chronic diseases with a potentially longer feedback loop due to the lack of remote clinical support.

A Copresence-Enhanced Approach

In our previous work, we introduced a concise design to enhance patients' perception of copresence through the sharing of information and emotions with their clinicians during remote monitoring [12]. The design has been built into a remote monitoring system for patients undergoing hemodialysis at home. The system allows patients to rate their emotions at the end of a dialysis session as part of the self-health reporting exercise, so the clinicians have a general understanding of the patient's feelings at the end of their dialysis session. Dialysis data of patients expressing a low mood are reviewed as a priority. Patients can also include additional comments for each submission, and clinicians can send feedback (with or without comments) by simply clicking the confirmed function in the system to let patients know their dialysis data, including their self-reported emotions, has been reviewed. The design was shown to reduce patients' anxiety caused by isolation during remote monitoring. The clinicians and nursing staff also reported satisfaction with the ease sending positive feedback to patients who were doing well with just one click.

During the COVID-19 pandemic, because of the social distancing and isolation rules, this remote monitoring system has become even more useful, as patients on home hemodialysis are able to continue managing their dialysis treatments themselves, without a visit to the dialysis clinic. The data from 65 active patients during past 5 months (February to June 2020) has led to 3166 recorded sessions. There is no increase in negative emotions expressed by patients after their dialysis treatments. In total, 31% of the sessions were recorded with additional comments. Patients continue to report feeling less isolated and lower anxiety levels when they receive feedback from their health care providers via the RPM system.

Copresence Strategies

Alleviating Patients' Anxiety of Not Being Able to Visit Hospitals

During the pandemic, opportunities for patients with chronic diseases to meet their clinicians are limited despite increased health concerns. With telehealth solutions featuring a copresence-enhanced design, patients can find an easy-to-access channel to share their feelings and thoughts with their health care providers [12]. Allowing patients to express themselves is an important part of clinician-patient communication, and it can be effective in reducing patients' anxiety and providing comfort [13]. Even without real-time or synchronous communication, patients can still benefit from copresence, and this can result in reduced anxiety caused by disconnection with their health care team. Patients' confidence is also enhanced when they receive simple, encouraging messages [14]. They can be more mindful about their health and better adhere to recommended regimens and behaviors without constant checks and reminders.

Offsetting the Demand for Health Care Teams

The current way of adopting telehealth requires an additional investment in time for health care staff (eg, increased standby time of clinical staff to respond to patients' needs outside of normal working hours) [15]. Unfortunately, the additional resources needed to monitor patients are not necessarily

available, especially during the pandemic. With copresence-enhanced functions, such as a one-click acknowledgment from the health care providers and prioritized response to patients' records based on their emotions and health-related data, staff can spend less time on nonurgent responses, thus reducing the stress associated with managing patients' expectations of a real-time connection with them through telehealth [16].

Providing Continuality When the Pandemic Ends

Patients with chronic diseases play a central role in the self-management of their medical condition [16] by making changes to their diet and physical activities to cope with their condition over a long period. Telehealth applications for patients with chronic conditions should be built to provide patients with more assistance and support for a positive behavioral change [17]. The copresence-enhanced design might give patients more confidence and assurance through subtle social cues delivered from the telehealth system [12] and has the potential to provide a continuality in nudging the shift from a hospital-centric chronic care model to a patient-centric one when the pandemic ends.

Conclusions

During the recent COVID-19 pandemic, more attention has been drawn to telehealth to emphasize its role in reducing hospital visits from both COVID-19 and non–COVID-19 patients and supporting home isolation in patients with mild symptoms. The needs of patients with chronic diseases tend to be overlooked during the pandemic. With reduced opportunities for routine clinic visits, patients with chronic diseases are adopting various telehealth services such as VC and RPM. We advocate for more innovative designs to be considered to enhance patients' feelings of copresence with their health care providers during this time. The copresence-enhanced design has been shown to reduce patients' anxiety and increase confidence in managing their chronic disease condition. It also has the potential to reduce patients' needs to engage their health care providers during a time when health care resources are stretched in most countries severely affected by the pandemic.

Conflicts of Interest

None declared.

References

1. Noel HC, Vogel DC, Erdos JJ, Cornwall D, Levin F. Home telehealth reduces healthcare costs. Telemed J E Health 2004 Jun;10(2):170-183. [doi: 10.1089/tmj.2004.10.170] [Medline: 15319047]
2. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Jun 03;69(686):e586-e594. [doi: 10.3399/bjgp19x704141]
3. Field MJ, Grigsby J. Telemedicine and remote patient monitoring. JAMA 2002 Jul 24;288(4):423-425. [doi: 10.1001/jama.288.4.423] [Medline: 12132953]
4. Timmers T, Janssen L, Kool RB, Kremer JA. Educating Patients by Providing Timely Information Using Smartphone and Tablet Apps: Systematic Review. J Med Internet Res 2020 Apr 13;22(4):e17342 [FREE Full text] [doi: 10.2196/17342] [Medline: 32281936]
5. Hartmann-Boyce J, Mahtani K. Supporting people with long-term conditions (LTCs) during national emergencies. CEBM. 2020. URL: https://www.cebm.net/covid-19/supporting-people-with-long-term-conditions-ltcs-during-national-emergencies/ [accessed 2020-04-20]
6. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet 2020 Mar;395(10229):1054-1062. [doi: 10.1016/s0140-6736(20)30566-3]
7. Lee J, Duan A. Australians with chronic illnesses urged to continue regular health check-ups. SBS Serbian. 2020 Apr. URL: https://www.sbs.com.au/language/english/audio/australians-with-chronic-illnesses-urged-to-continue-regular-health-check-ups_2 [accessed 2020-04-15]
8. Nowak K. Defining and differentiating copresence, social presence and presence as transportation. 2001 Presented at: Presence 2001 Conference—4th Annual International Workshop on Presence; May 21-23, 2001; Philadelphia, PA p. 1-23 URL: https://astro.temple.edu/~lombard/ISPR/Proceedings/2001/Nowak1.pdf [doi: 10.5040/9780571284931.00000003]
9. Callaway Russo T. Operationalizing mediated presence: Initial steps toward a measure of the construct. 2001 Presented at: Presence 2001 Conference—4th Annual International Workshop on Presence; May 21-23, 2001; Philadelphia, PA URL: https://astro.temple.edu/~lombard/ISPR/Proceedings/2001/Russo.pdf
10. Kim H, Suh K, Lee U. Effects of collaborative online shopping on shopping experience through social and relational perspectives. Information & Management 2013 Jun;50(4):169-180. [doi: 10.1016/j.im.2013.02.003]
11. Kahai SS, Carroll E, Jestice R. Team collaboration in virtual worlds. SIGMIS Database 2007 Oct 28;38(4):61-68. [doi: 10.1145/1314234.1314246]

12. Liu N, Kim J, Jung Y, Arisy A, Nicdao MA, Mikaheal M, et al. Remote Monitoring Systems for Chronic Patients on Home Hemodialysis: Field Test of a Copresence-Enhanced Design. JMIR Hum Factors 2017 Aug 29;4(3):e21 [FREE Full text] [doi: 10.2196/humanfactors.7078] [Medline: 28851680]
13. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
14. DiMatteo MR, Haskard-Zolnierek KB, Martin LR. Improving patient adherence: a three-factor model to guide practice. Health Psychology Review 2012 Mar;6(1):74-91. [doi: 10.1080/17437199.2010.537592]
15. MacNeill V, Sanders C, Fitzpatrick R, Hendy J, Barlow J, Knapp M, et al. Experiences of front-line health professionals in the delivery of telehealth: a qualitative study. Br J Gen Pract 2014 Jun 30;64(624):e401-e407. [doi: 10.3399/bjgp14x680485]
16. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
17. Suter P, Suter WN, Johnston D. Theory-based telehealth and patient empowerment. Popul Health Manag 2011 Apr;14(2):87-92. [doi: 10.1089/pop.2010.0013] [Medline: 21241182]

Abbreviations

COVID-19: coronavirus disease
GP: general practitioner
RPM: remote patient monitoring
VC: video consultation

Edited by G Eysenbach, G Fagherazzi; submitted 22.04.20; peer-reviewed by A Al Rajeh, M Allman-Farinelli; comments to author 25.06.20; revised version received 16.07.20; accepted 26.07.20; published 07.08.20.

Please cite as:
Liu N, Huang R, Baldacchino T, Sud A, Sud K, Khadra M, Kim J
Telehealth for Noncritical Patients With Chronic Diseases During the COVID-19 Pandemic
J Med Internet Res 2020;22(8):e19493
URL: http://www.jmir.org/2020/8/e19493/
doi:10.2196/19493
PMID:32721925

Review

Approaches Based on Artificial Intelligence and the Internet of Intelligent Things to Prevent the Spread of COVID-19: Scoping Review

Aya Sedky Adly[1], MSc, PhD; Afnan Sedky Adly[2,3], MSc; Mahmoud Sedky Adly[4,5], DDS, MSc, HDD

[1]Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt

[2]Faculty of Physical Therapy, Cardiovascular-Respiratory Disorders and Geriatrics, Laser Applications in Physical Medicine, Cairo University, Cairo, Egypt

[3]Faculty of Physical Therapy, Internal Medicine, Beni-Suef University, Beni-Suef, Egypt

[4]Faculty of Oral and Dental Medicine, Cairo University, Cairo, Egypt

[5]Royal College of Surgeons of Edinburgh, Scotland, United Kingdom

Corresponding Author:
Mahmoud Sedky Adly, DDS, MSc, HDD
Faculty of Oral and Dental Medicine
Cairo University
Cairo University Road
Cairo,
Egypt
Phone: 20 1145559778
Email: dr.mahmoud.sedky@gmail.com

Abstract

Background: Artificial intelligence (AI) and the Internet of Intelligent Things (IIoT) are promising technologies to prevent the concerningly rapid spread of coronavirus disease (COVID-19) and to maximize safety during the pandemic. With the exponential increase in the number of COVID-19 patients, it is highly possible that physicians and health care workers will not be able to treat all cases. Thus, computer scientists can contribute to the fight against COVID-19 by introducing more intelligent solutions to achieve rapid control of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes the disease.

Objective: The objectives of this review were to analyze the current literature, discuss the applicability of reported ideas for using AI to prevent and control COVID-19, and build a comprehensive view of how current systems may be useful in particular areas. This may be of great help to many health care administrators, computer scientists, and policy makers worldwide.

Methods: We conducted an electronic search of articles in the MEDLINE, Google Scholar, Embase, and Web of Knowledge databases to formulate a comprehensive review that summarizes different categories of the most recently reported AI-based approaches to prevent and control the spread of COVID-19.

Results: Our search identified the 10 most recent AI approaches that were suggested to provide the best solutions for maximizing safety and preventing the spread of COVID-19. These approaches included detection of suspected cases, large-scale screening, monitoring, interactions with experimental therapies, pneumonia screening, use of the IIoT for data and information gathering and integration, resource allocation, predictions, modeling and simulation, and robotics for medical quarantine.

Conclusions: We found few or almost no studies regarding the use of AI to examine COVID-19 interactions with experimental therapies, the use of AI for resource allocation to COVID-19 patients, or the use of AI and the IIoT for COVID-19 data and information gathering/integration. Moreover, the adoption of other approaches, including use of AI for COVID-19 prediction, use of AI for COVID-19 modeling and simulation, and use of AI robotics for medical quarantine, should be further emphasized by researchers because these important approaches lack sufficient numbers of studies. Therefore, we recommend that computer scientists focus on these approaches, which are still not being adequately addressed.

(J Med Internet Res 2020;22(8):e19104) doi:10.2196/19104

KEYWORDS

SARS-CoV-2; COVID-19; novel coronavirus; artificial intelligence; internet of things; telemedicine; machine learning; modeling; simulation; robotics

Introduction

With the emergence of the novel coronavirus disease (COVID-19) pandemic, many regions and countries have been facing different risks at different times. Due to the high infectivity and spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes COVID-19, combined with its increased hospitalization rate, COVID-19 is a serious public health threat worldwide [1,2]. Artificial intelligence (AI) systems such as cognitive computing, deep learning, convolutional neural networks, and machine learning can play a critical role in detection, large-scale screening, monitoring, reduction of caregiver workload, and prediction of possible interactions with the new therapies for this virus [3-6]. In this review, we provide an analysis of AI applications that may help limit the spread of or even help eradicate this virus if properly adapted and applied by countries.

Methods

We conducted an electronic article search using the MEDLINE, Google Scholar, Embase, and Web of Knowledge databases. The search included only English-language publications, with a focus on evidence-based research articles. The search focused on both randomized and nonrandomized controlled trials, longitudinal studies, cross-sectional studies, and retrospective studies. A list of keywords were used with different combinations, as follows: *SARS-CoV-2*, *COVID-19*, *novel coronavirus*, *artificial intelligence*, *internet of things*, *telemedicine*, *machine learning*, *modeling*, *simulation*, and *robotics*.

Results

Figure 1 shows a summary of AI-based approaches to prevent the spread of COVID-19.

AI for Detection of Suspected COVID-19 Cases

A pandemic creates unique challenges to the delivery of health care, and these challenges must be faced by a limited number of health care workers. AI can help address these problems; for example, a smartphone app could be developed that collects signs, symptoms, previous locations of the patient, travel history, and updated areas of the outbreak, then processes and filters this information using algorithms so that only suspected cases are examined by physicians (Figure 2) [7].

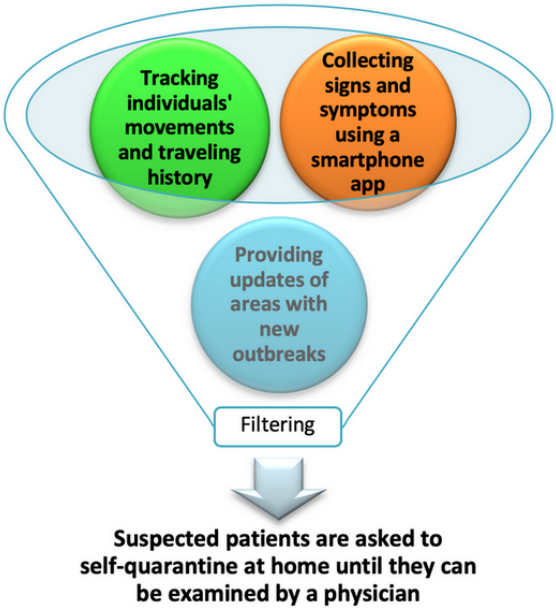
On January 22, 2020, the Center for Systems Science and Engineering at Johns Hopkins University launched a publicly shared web-based interactive dashboard [8]. The aim of this dashboard was to accurately visualize and track reported cases of COVID-19 in real time. This revolutionary idea has contributed to the rapid identification of new outbreaks of the disease. The information on this dashboard is updated twice daily, which provides an added advantage. This has opened a new spectrum of AI to predict and recommend quarantine in areas where a threshold number of cases is reached. It can also aid early diagnosis of patients if they report travelling to these areas. All these data are displayed freely through Google Sheets and the ArcGIS Living Atlas [9].

Another important application of AI is remote monitoring of home-quarantined patients and their families via smartphones or smart bracelets. An automatic alarm will sound and provide a warning message if the patient or family member breaks quarantine.

All these contributions can significantly decrease the burden on health care workers and enable them to work more efficiently in a safe environment [7].

Figure 1. Summary of AI approaches to address the COVID-19 pandemic. AI: artificial intelligence; CAD: computer-aided diagnosis; COVID-19: novel coronavirus disease.

Figure 2. Potential artificial intelligence–based detection of suspected patients with coronavirus disease using a smartphone app.



AI for Large-Scale COVID-19 Screening

Noncontact systems that use AI to measure signs and symptoms of COVID-19 are extremely important for large-scale screening in a significantly short period (Table 1). Rapid screening using these systems not only enables remote detection of suspected cases but can also be performed by significantly fewer personnel, which can reduce the work stain on areas such as airports and supermarkets and further augment social distancing [7].

Noncontact systems that can be used for large-scale screening include infrared thermal imaging technology and camera-based motion detection software to detect and analyze abnormal respiratory patterns [10].

One of the diagnostic signs of COVID-19 is fever [10]. Infrared thermal cameras enable real-time visualization of any transient or constant changes in the energy radiating from individuals, which enables estimation of surface temperature [11]. With the aid of AI detection algorithms, suspected individuals with COVID-19 can be automatically identified and tracked using infrared thermal cameras with minimal need for human monitoring.

Respiratory pattern is another diagnostic sign of individuals with COVID-19 [12,13]. The pattern of respiration in COVID-19 is quite different from that in the common cold or influenza [14,15]. Tachypnea, or abnormally rapid breathing, may indicate infection with COVID-19. Using a depth camera to conduct noncontact respiratory measurements of individuals and analyzing these measurements using AI detection algorithms was found to be a promising method for additional confirmation of COVID-19 cases [12,13].

Table 1. Comparison of the cost and efficiency of AI systems versus conventional human labor for large-scale screening of COVID-19.

Parameter	AI[a] systems for COVID-19[b] screening	Conventional human labor for COVID-19 screening
Cost	May be initially high but significantly decrease with time, leading to lower cumulative cost.	May be high or low according to country but will generally have higher cumulative cost, including the added cost of applying preventive measures for direct contact.
Sensitivity (probability of detection)	Very high if multiple confirmatory methods are used.	Affected by distractors.
Specificity (excluding negative conditions)	False positive results may occur due to other conditions having similar signs.	When screening is performed by expert health care workers, specificity will be high.
Duration of screening	Very short.	Relatively long and may require additional employees to shorten the duration.
Number of working hours	24 hours, 7 days a week in addition to the working hours of quarantined physicians and health care workers who are suspected to have COVID-19, who will be able to work from home.	No more than 48 hours of work per week and no more than an average of 8 hours of nighttime work per 24 hours of total work [16].
Possibility of COVID-19 infection among examined subjects and examiners	Very little or no contact between persons, which significantly decreases the possibility of transmission of COVID-19 infection.	High risk of transmission of COVID-19 infection even if preventive measures are followed, as human error may occur.

[a]AI: artificial intelligence.

[b]COVID-19: coronavirus disease.

AI for COVID-19 Monitoring

Large amounts of data from patients with COVID-19 are routinely manually collected and interpreted in hospitals with or without standalone monitoring devices. The processes of collecting patients' monitor data in hospital wards are different in many hospitals, as different approaches are used worldwide. In some hospitals, data are only collected manually and then recorded in spreadsheets, which are discarded after the patient is discharged [17].

AI systems can collect monitor data from patients using personal digital assistants (PDAs), tablets, and similar equipment; the patients' data can be stored in electronic health records (EHRs), where they can be easily shared and rapidly transferred when needed [18].

With different technologies and AI, the collection, analysis, and interpretation of patients' monitor data can be fully or partially automatized, which diminishes both infection risk and the burdens imposed on medical staff to constantly gather, store, analyze, and interpret these data [17].

Several intelligent and expert systems have been developed for medical monitoring using wireless sensor networks. These methods have been found to be efficient when signals are continuous; however, the applicability of these expert systems to incomplete data has not yet been established [19].

Thoracic computed tomography (CT) without contrast is considered to be an effective method to detect, quantify, and monitor symptoms of patients with COVID-19. Deep learning algorithms could be developed to contribute to the analysis, interpretation, and tracking of large numbers of thoracic CT examinations [20].

Developing a video imaging–based high-speed medical monitoring system that uses motion tracking monitoring algorithms for patients with COVID-19 can also provide a large amount of factual information regarding vital signs (temperature, heart rate, respiratory rate, blood pressure, oxygen saturation) as well as status, condition severity, existing comorbidities, and patient discharge. Modal parameters can be subsequently extracted to analyze the level of severity or damage from COVID-19. Motion tracking monitoring has been investigated in several studies with very promising results [21,22].

It has been reported that patients with COVID-19 can be discharged from the quarantine ward if they meet the following criteria: being afebrile for a minimum duration of three days, respiratory symptoms resolved, improvements in radiological signs for pulmonary infiltrates, at least two consecutive negative COVID-19 nucleic acid tests with sampling intervals of at least 24 hours [15].

AI-based automated analysis tools for tracking the discharge criteria for patients with COVID-19 should be developed to determine and differentiate treated patients from those who still need to be isolated. In addition, AI-based automated CT image analysis tools for tracking, quantification, and detection of COVID-19 can assist in differentiating patients with COVID-19 from individuals who do not have the disease. Moreover, artificial neural networks can be trained to infer qualitative characteristics that are based on intensity as well as network inferences, which can be correlated according to the patient's condition.

An exploratory study showed that AI can reliably and efficiently contribute to accurate detection and tracking of the progression or resolution of COVID-19. In this work, it was also reported that the rapid developments in AI-based automated tools for CT image analysis can achieve a high degree of accuracy in detection of COVID-19–positive patients in addition to precise assessment of the severity of the disease [23,24].

AI for Analysis of Experimental Therapies for COVID-19

Computer-aided drug design is highly efficient for rapidly identifying drug repurposing candidates and should be further investigated.

One example is the mechanism-based inhibitor N3, which was identified using computer-aided drug design. N3 was found to particularly inhibit the main protease of SARS-CoV-2 because it can fit within a substrate-binding pocket [25].

A crucial application for computer-aided drug repurposing is treating novel diseases such as COVID-19 by identifying drugs that were developed to treat other diseases. Drug repurposing can be realized by conducting systematic drug-drug interaction analyses and studying drug-target interactions, which can be performed using AI-based algorithms [26-28].

AI has been recognized to have a transformative influence on drug development. In accordance with a recent report, machine learning and big data can have great effects on health care systems and may have positive outcomes in the pharmaceutical market, as it has been predicted by industry experts that developing drugs through AI methods will provide significantly improved feedback. However, AI-based drug development may be slower to launch in the long term. If applied correctly, these methods can be highly competitive in the pharmaceutical industry [29,30].

Natural language processing is another branch of AI that can be applied in COVID-19 drug development. Its methods can be beneficial for extracting meaning from text using machine learning approaches and searching for external biomedical content in drug discovery [31,32]. Another AI approach is machine or computer vision, which is the use of algorithms to enable computers to comprehend the content of images. These images can be used to understand cell anatomy and identify novel treatments for COVID-19 [33].

The use of AI is prevalent in drug discovery, and many pharmaceutical corporations have established in-house partnerships or initiatives with AI companies. Some organizations are currently using AI approaches to find novel uses for late-stage drug candidates or repurpose existing drugs [34,35]. However, data of sufficient quality are needed to train systems for COVID-19 AI implementation. Data accessibility is an additional challenge because the systems will be trained via supervised learning, which requires substantial amounts of data on COVID-19 to accurately perform complex tasks.

Accessibility of data can also create costs if these data must be accessed by corporate partners or technology providers. Additionally, the data should be of high quality, as poor or corrupt data can affect the results. At present, although industry standards for data have been established for many uses, they currently may not apply in the case of COVID-19. Moreover, a substantial amount of effort is required to integrate data on COVID-19 into corporation systems to use it for AI.

AI for Screening COVID-19 Pneumonia

COVID-19 is spreading very rapidly worldwide; meanwhile, it can be difficult to diagnose the disease. These limitations can be partially accredited to the limited number of physicians who are capable of interpreting data and using identification methods compared to the great increase in the number of cases [36].

It has been reported that radiologists can distinguish COVID-19 pneumonia from other types of pneumonia in chest CT images with high specificity. Other reports have demonstrated that reverse transcriptase–polymerase chain reaction (RT-PCR) analysis has a low sensitivity of about 60% to 71% for detection of COVID-19 [37]. This may be due to decreased viral load in the test specimen or to laboratory error. The false negatives produced by RT-PCR analysis can hinder quarantine efforts and require tests to be repeated. On the other hand, the sensitivity of chest CT has been established to be approximately 56% to 98% for COVID-19 detection at initial presentation, and it may help rectify false negative results that are obtained by RT-PCR analysis during the early stages of development of the disease. Additionally, chest CT images can reveal disease progression [33].

Compared to non–COVID-19 pneumonia, CT images of COVID-19 pneumonia are more likely to show vascular thickening, fine reticular opacity, ground-glass opacity, peripheral distribution, and reverse halo sign in addition to bilateral peripheral involvement of multiple lobes. The CT signs may improve gradually approximately 14 days after the onset of symptoms. On the other hand, CT images of COVID-19 pneumonia are unlikely to show both peripheral and central distribution, pleural effusion, pleural thickening, lymphadenopathy, or air bronchogram [38]. Thus, future directions in chest CT may involve developing an AI-based classifier that can further augment the performance of radiologists when combined with clinical information.

Exploration of automated pneumonia analysis via deep learning is a very important issue for examination for many different reasons. The most important reason is that the chest radiographs of COVID-19 patients must be reviewed by highly trained specialists, which creates large amounts of work for those specialists. Further, it is very difficult to read these images because pneumonia is normally revealed over one or more areas of increased opacity [39]. This increase may be due to a reduction of the ratio of gas to soft tissue (lung parenchyma, stroma, and blood) in the lungs. Thus, when an amplified attenuation area (opacification) is reviewed on a CT or chest radiograph, it is crucial to define where the opacifications occur. Diagnosis is very complicated because other conditions can alter the appearance of a radiograph, such as bleeding, surgical changes, post-radiation changes, pulmonary edema, lung cancer, and volume loss due to collapse or atelectasis; aspects such as inspiration depth and patient positioning may also affect the radiograph. Due to these issues, interpreting chest images of COVID-19 patients by the human eye alone is extremely challenging [38].

Although the fields of machine learning and pneumonia research are individually well developed, very limited work is currently available regarding the application of machine learning for diagnosis of pneumonia, especially for patients with COVID-19. Research on deep learning algorithms has recently shown rapid progress. In general, pneumonia diagnosis for medical fields is

a broadly known discipline. The combination of those two areas would be novel and promising [40]. Previous research has been performed to apply image processing techniques for identification of cases with pneumonia. Other algorithms have been developed to crop and extract the lung regions from images. The Otsu thresholding method has been used to segregate infected cloudy pneumonia lung regions from healthy regions. Another detection method has been proposed based on cellular neural networks; the simulated results showed remarkable performance in differentiating the lung region area and normal area based on changes in segmentation and grayscale color [41].

It was also reported that machine learning has the important advantage of increased capability to evaluate the effects of interventions when studying subpopulations in clinical pneumonia research [40]. Another significant aspect of machine learning for identifying COVID-19 infection is image classification. This has resulted in the development of a range of networks for image classification. These networks usually employ backpropagation algorithms. Very large neural networks are usually needed for deep learning architectures. Some commonly used deep learning architectures include GoogLeNet, ResNet, and AlexNet [42].

Computational models for deep learning can discover intricate structures in very large data sets by using backpropagation algorithms to indicate how a machine would modify the internal parameters it uses to compute representations of each layer from representations of the previous layer [43]. Training deep learning networks that can recognize symptoms of COVID-19 pneumonia would be very beneficial. These networks could facilitate prescreening and automated diagnosis.

Research has provided insights into testing difficulties associated with very large radiograph image data sets. Most research in this area focuses on small and controlled data sets. For these data sets, irregular features may not be an issue. However, as the size of the data set increases, the number of irregular images also increases. It is particularly important to overcome this potential hurdle for COVID-19 diagnosis because a model that is used by medical industries must be capable of considering all radiographic forms. However, the model will still be required to function at high levels regardless of whether these image types are imputed [42].

Applying computer-aided detection to radiographs of patients with pneumonia may provide a supplement or alternative to human reading [44]. To develop an efficient AI algorithm that uses chest radiographs to predict clinical outcomes, the clinical settings and disease burden profiles of COVID-19 prevalence should be considered.

Software should be trained to use selected chest radiographs of patients with COVID-19 to automatically segment fields of the lung in other radiographs. After the classifier is trained on randomly selected chest radiographs, it can be applied to the remaining chest radiographs.

Chest radiographs commonly demonstrate many categories of abnormalities. Some disease processes may show a combination of abnormal shapes where the disease process has altered the patient's normal anatomy, such as cardiomegaly and focal abnormalities. These focal abnormalities may appear as isolated density changes, such as pulmonary nodules and texture abnormalities. Texture abnormalities are characterized by diffuse changes [45]. In these situations, texture analysis may be useful for assigning a probability to each location in the lung fields [39].

In general, further developments and validation should be performed in multicenter studies, which would be significant for future research on AI and computer-aided diagnosis in COVID-19 radiology.

AI and the Internet of Intelligent Things for COVID-19 Data and Information Gathering and Integration

Patients with COVID-19 can provide novel types of data that are relevant to achieving medical goals. Self-tracking devices, mobile health apps, and social media can provide patients with information about COVID-19 and can enable them to monitor their health or even achieve a certain goal. Incorporating AI into devices offers tools and methods for designing and analyzing therapies that can be made accessible to patients and clinicians while supporting consistent integration into patients' lives and clinicians' practices [17,46,47]. This field not only offers promising challenges for patients with COVID-19 but can also improve systematic data collection to determine treatment effectiveness.

In the context of the Internet of Intelligent Things (IoIT), a "thing" is a system or entity composed of subsystems, and every subsystem is an indispensable component of the system [48]. Thus, if we divide COVID-19 medical and diagnostic devices in health centers or hospitals into things and modulate these things to be sufficiently intelligent to operate on their own, we can establish a behavior for every subsystem. The challenge is to understand each thing separately. Therefore, to understand the behavior of things used to help control COVID-19, it is necessary to understand the main subcomponents of each thing; then, we must understand the behavior of each subsystem to understand the behavior of the thing as a whole and how to enable it to connect and communicate with other things via the internet.

A crucial challenge for the advancement of digital health is efficient and effective integration of incomplete or heterogeneous information about COVID-19 from different sources and different types, such as interoperability solutions, insufficient availability, and existence of current information silos. All this contradicting information is hindering the development of effective applications. Different information types must be integrated, such as clinical information (including EHRs), information extracted from the biomedical literature by text mining, and high-throughput information on how drugs or chemicals interact in different circumstances [36].

Additionally, big data may be useful to optimize resource use when making informed decisions based on the availability of data related to COVID-19 cases. AI, machine learning, and the internet of things (IoT) could contribute significantly to this process [49-51].

Big data analysis performed via AI can be interpreted as a means of training computers to mimic thinking patterns and even simulate human behaviors. It is assumed that the accuracy of the achieved results will increase with increasing availability of processing power and data because it is desirable to perform machine learning using real-time data [52,53].

Intelligent computing systems can support decision-making even when the problems are complex. A great deal of success has been achieved in integrating expert systems into intelligent systems. However, expert systems may face difficulties in acquiring and processing COVID-19 knowledge. Thus, to recognize the involved patterns and the knowledge gained from different fields, it is vital to combine data mining with intelligent computing systems to determine the information gathered and the patterns involved, which may include clustering algorithms, neural network algorithms, regression algorithms, and Bayesian algorithms [54].

Numerous other challenges may also emerge, including privacy breaches, ethical concerns, and lack of information security. Thus, the ability to share, analyze, and gather information about COVID-19 in real time with different devices may add to the difficulty of maintaining patient privacy.

The mining of very large COVID-19 data sets may present difficulties in terms of computation and storage. For example, combining various types of information in heterogeneous COVID-19 data sets with global information systems can be complicated. Additionally, numerous COVID-19 experts are needed to formulate the data mining process. Finally, the accuracy of data mining results depends on the level of diversity of the gathered COVID-19 data set.

Data mining can also provide many benefits. For example, powerful high speed processes can be established to examine the enormous amounts of information related to COVID-19 and can provide a knowledge base for a specific area of COVID-19 information. Additionally, the diagnosis and prediction of COVID-19 can be automated, and data mining can enhance decision-making processes.

AI for Resource Allocation to Patients With COVID-19

As more countries are affected by COVID-19 worldwide, it is increasingly necessary to prioritize allocation of resources to patients with the disease [55].

If we attempt to calculate conservative estimates of the resources that will be needed to fight COVID-19, we find that the resources needed are beyond the available capacities of health care facilities. To achieve an accurate estimation of these resources, estimations should include human resources (such as medical staff, including therapists and nurses) and other facilities (including numbers of hospitals, emergency departments, intensive care units, adult beds, neonatal beds, pediatric beds, ventilators, oxygen concentrators, oxygen cylinders, oxygen plants, liquid oxygen, medications, personal protective equipment, critical medical supplies, and pulse oximeters). It is very important to assess all these resources before establishing any action plan for resource allocation.

Given these numbers, if the curve of infected individuals is not reduced and flattened over a short period of time, the COVID-19 pandemic will likely result in a shortage of medical resources, particularly ventilators and hospital beds. The number of medical workers who will be affected should also be taken into consideration, as these workers will commonly be quarantined.

In addition, even after a vaccine is developed, time will be needed to produce, distribute, and administer it; shortages of the vaccine will probably arise as well.

Supply limitations restrain the speed of producing more resources, such as ventilators. Currently, manufacturers cannot state with certainty how many ventilators they can produce during a year.

Policy makers are facing decisions about which patients will be provided with available resources while taking into consideration that patients who do not receive these resources will very likely die.

These decisions cannot be made based on age alone. Professionals have indicated that prioritization should be based on which patients are more likely to survive to save the maximum possible number of lives [56].

To address the COVID-19 resource allocation problem, it is necessary to understand complex data structures related to the prioritization criteria and resources in question.

Applying big data analysis and data mining algorithms, including unsupervised and supervised machine learning, to optimize the prioritization process would be very helpful to reduce harm as much as possible in emergency situations when resources are scarce. With the use of a trained model, machine learning can also minimize human effort while improving or maintaining accurate prioritization.

Moreover, intelligent estimation of resource needs (especially oxygen needs) is crucial to be as accurate as possible in addition to deciding which source of oxygen would be better for a patient by considering the total gross oxygen flow that would be needed through anticipating the load of each patient based on severity and number of patients.

Developing an expert system or framework for reasonable and fair allocation of COVID-19 therapeutics and resources is very complex and requires coordination of many variables and estimations that must be guided by comprehensive and accurate information based on IoT technology, which can provide information about distributions, numbers, sizes, capacities, and risks related to both resources and the affected populations worldwide. The realization of such frameworks or expert systems will mainly depend on sharing and collecting data; this can be greatly facilitated by the IoT.

AI for COVID-19 Prediction

Transmission Rates

In the initial stages of the COVID-19 outbreak, it is critical to analyze and understand the dynamics of transmission of the virus. Changes in the estimations of transmission over time can offer insights in epidemiological situations and demonstrate the effectiveness of outbreak control measures [57].

These analyses can also provide predictions of future growth, assist risk estimation for countries, and guide strategy for alternative interventions [58]. These analyses present many challenges, especially in real time. Moreover, with COVID-19, the appearance of symptoms may be delayed due to the period of incubation; also, confirmation of cases may be delayed in accordance with detection and testing capacity.

AI approaches can account for these delays as well as for uncertainty by explicitly incorporating delays that result from the natural history of virus infection or reporting processes. In addition, individual data sources may be incomplete, be biased, or only capture some aspects of the dynamics of the outbreak. However, evidence synthesis approaches that fit with multiple data sources can enable more robust estimations of the transmission dynamics that can be gathered from noisy data [58,59].

Many factors can affect the transmission dynamics of COVID-19 throughout a country, such as outflow population size from a certain place to affected provinces or cities, geographic locations, interventions, social and economic activities, health care facilities, and environmental heterogeneity. The process of clustering temporal dynamics will provide various insights into the patterns of COVID-19 propagation. In addition, modified auto-encoders have been used to predict the accumulative number of new confirmed cases of COVID-19. By hypothesizing the initial amounts of the epidemical outbreak, modified autoencoders can be used with known architectures and parameters to predict the sizes of future outbreaks and simulate the impact of interventions on the severity and size of epidemics [43,57,60].

Data-driven AI-based methods offer real-time forecasting techniques for estimating and tracking the severity of epidemics, assessing their trajectory, predicting their length, and supporting decision-making by health care workers and governments [61,62].

Mutations

The prediction of genetic mutations in the SARS-CoV-2 genome has attracted much attention. Rapid progress has been realized to predict these mutations and analyze their effects. Tracing the mutations of SARS-CoV-2 can provide comprehensive understanding of the evolution dynamics of the virus.

In some studies, antigenic cartographies have been developed for quantifying and visualizing site mutations and antigenic differences [63]. Neural networks were applied in another study to predict point mutations that may appear on structure alignments of primary RNA sequences [64]. Network models were also outlined to demonstrate the dynamics and evolutionary patterns of a virus [65].

Many RNN-based neural networks have been developed for predicting time series tasks [66]. K-means clustering can also be used to find clusters of mutations of SARS-CoV-2, which can provide insights into the nature of mutations and how they can be addressed.

A model was proposed to forecast the properties of viruses that are not characterized antigenically using phylogenetic trees. Modeling sequential data dynamically is important. Recent research has provided ways to embed biological sequences into lower-dimensional vector spaces [67].

Severity and Mortality Rates

The assessment of COVID-19 severity by clinical presentation can no longer meet urgent clinical needs. Thus, introducing a deep learning–based model by quantitating clinical features to predict the severity and mortality rates of COVID-19 will be of significant value. Deep learning–based quantitative CT measurements of the extent of lesions and clinical features on initial admission can assist in predicting COVID-19 severity; this will enable physicians to triage patients and design treatment protocols and follow-up evaluations in advance.

Convolutional neural networks were introduced as a potential solution to problems faced in automatic organ segmentation [57,68].

In a recent study, a new model to forecast the prognosis of COVID-19 was established. It has been reported that parsimonious models, which contain five features (age, lactate dehydrogenase, C-reactive protein, $CD4^+$ T-cell counts, and mass of infection), are an ideal measure for predicting COVID-19 severity. This is a common regression method with high-dimensional data (Cox proportion hazard regression model) that has been extended to and broadly used in logistic regression models for outcome forecasting and survival analysis. This approach may be superior to conventional methods when choosing predictors and may allow researchers to combine selected particular features into single signatures [69,70].

Another study demonstrated that machine learning algorithms are superior to traditional statistical modeling approaches for predicting mortality in patients with pneumonia. However, it was found that none of the samples or models assessed showed overall precise predictions of patient mortality, and all the samples revealed wide variations in performance based on the measures used [71].

In a recent study, researchers suggested an algorithm that could anticipate the mortality rate of patients with COVID-19 with accuracy that reached 90%. In that study, machine learning methods were used to establish a predictive model for early recognition of critically sick patients based on clinical and epidemiological data obtained from patients infected with COVID-19. The working mechanism for this machine learning model was based on quantitative sorting of the clinical features according to their criticality. The revealed features were then sorted, and an interpretable clinical route was obtained [23,72].

AI for COVID-19 Modeling and Simulation

Mathematical modeling of viruses and infections may help simplify the process of understanding virus dynamics. Many authors have used ordinary differential equations in virology and epidemiology to model and simulate different scenarios [73-75].

Viruses are believed to be among the most numerous and divergent biological systems [76]. However, despite their diversity, many shared events and common processes are found

in most or possibly all viruses, such as viral replication cycles, which are necessary for productive infection.

Disruption of one or more of these steps may impair or prevent propagation of the virus. Similarly, destabilizing the infective virions before they can attack host cells may be an effective way to prevent propagation. From a therapeutic viewpoint, the process of modeling and simulating the molecular-level dynamics of SARS-CoV-2 in detail at every stage in addition to the virions themselves is important and desirable in such situations and can greatly aid understanding and management of infection with the virus. This knowledge would also be required to address emerging drug-resistant viral strains and future outbreaks of novel pathogenic species similar to COVID-19. In theory, developing therapeutics that can target single or multiple steps in the viral replication cycle or critical processes that have limited capacity for viable mutation can reduce the opportunity for SARS-CoV-2 to develop resistance to administered drugs. Likewise, if simulations aid understanding of the dynamic and structural basis for COVID-19 drug resistance, antiviral drugs can be modified to account for mutations [77].

As in numerous areas in biology, to obtain a comprehensive understanding of virus operations, multidisciplinary approaches are required. Supported by structural biology advancements, AI and computational methods have emerged as very powerful tools that can complement experimental techniques with the use of mathematical modeling and simulations. In several cases, AI and computational approaches can help bridge information gaps among experiments through reporting in different temporal and spatial domains in addition to their considerable predictive powers.

AI Robotics for Medical Quarantine and Isolated Patients With COVID-19

COVID-19 is a highly transmissible disease that poses a real threat to health care workers. Transmission of this disease to health care workers is highly likely, especially during pandemics when hospitals are overloaded with infected patients.

AI can offer safe and efficient solutions, such as robots that health care professionals can operate while teleconferencing with patients. Teleoperated robots can accomplish common nursing tasks in hazardous areas, such as delivering meals or medications, collecting specimens, and transporting waste, with high accuracy and efficiency [78]. An obvious advantage of using these robots is that a single operator can control multiple robots while rapidly switching between quarantine areas. Other advantages include the ability to communicate with patients via a virtual telepresence system 24 hours per day, 7 days per week. Moreover, a robot called TRINA (Tele-Robotic Intelligent Nursing Assistant) was used to perform error-prone nursing jobs and showed promising results [79].

Toward Preventative Medicine Using AI and Telemedicine

Several studies have specifically demonstrated the significance of using telemedicine in public health emergencies and disasters. Telemedicine programs take time to develop; however, health systems that have already developed telemedical innovations can leverage and modify them to rapidly manage COVID-19 outbreaks [80].

Forward triage is considered to be a central strategy of health care surge control; it depends mainly on sorting patients prior to their arrival at the emergency department. On-demand or direct-to-consumer telemedicine is a forward triage method that enables effective screening of patients. This screening protects patients, health care workers, and the community from exposure; additionally, it is both patient-centered and conducive to self-quarantine. Telemedicine allows patients and physicians to communicate at any time by using smartphones and webcam-enabled computers [18,81].

At present, the main barrier to large-scale telemedicine screening for COVID-19 is testing coordination. As the availability of testing sites increases, development and integration of local systems into telemedicine workflows is needed to test appropriate patients while decreasing exposure using tents, in-car testing, or dedicated office space. To keep pace with the evolving recommendations regarding COVID-19, health systems are employing bots or automated logic flows that can refer only moderate- or high-risk patients to nurse triage lines and can also allow patients to request video visits with on-demand providers [7]. It is important that practices not routinely refer patients to urgent care medical centers or emergency departments, as this will create exposure risk for health care providers and overload these centers with patients.

Before the outbreak of COVID-19, several emergency departments adjusted their provider-in-triage models for rapid initial testing and evaluation to allow remote providers to perform intake [82]. In emergency situations, web conferencing software with secured open lines from the triage room to a provider can be rapidly implemented [83]. Employing a single remote clinician to cover several sites can address workforce challenges; however, this measure is difficult to implement if the software lacks a queuing function. To avoid exposing staff, telehealth visits can be conducted using paired tablets or commercial systems that enable communication with providers through dedicated connections. However, this system does not fully eliminate exposure of health care providers to patients who require certain manual procedures.

Electronic monitoring programs enable physicians and nurses to remotely monitor patients' status in several hospitals. Through mobile integrated medical programs or community paramedicine, patients can be managed from their homes, with medical support provided virtually. In Houston, the ETHAN (Emergency Telehealth and Navigation) project uses telemedical oversight to augment in-person care provided by nearly 1000 responders, decreasing the requirement for transportation to emergency departments [84,85].

Telemedicine can offer rapid access to specialists who are not instantly available in person. Barriers to implementing these programs are largely related to credentialing, payment, and specialist staffing [86]. COVID-19 has raised concerns regarding workforce capacity. Telemedicine can enable quarantined physicians to remotely manage and treat patients, freeing time for other physicians to provide in-person care.

In addition, remote training sessions and online training modules can be made available to patients or clinicians who need assistance or in-time training. Program implementation, regulatory and payment structures, credentialing across hospitals, and state licensing will all require time; however, health systems that have already invested in telemedicine are well positioned to ensure that the patients with COVID-19 obtain the care they require. In this instance, telemedicine may be a perfect virtual solution.

Discussion

AI can potentially provide novel and reliable paradigms for health care services. Due to the nearly unlimited abilities of AI that are gained from its numerous algorithms and approaches, it can help address the virulent spread of the SARS-CoV-2 virus worldwide. Proper application of AI through the use of both existing and novel machine learning approaches may be pivotal to eliminating COVID-19. Furthermore, there is a need for major investment in this field to enable rapid response to the dangers of this disease; this may be a major factor in saving lives worldwide.

Conflicts of Interest

None declared.

References

1. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
2. Mahase E. Coronavirus covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. BMJ 2020 Feb 18;368:m641. [doi: 10.1136/bmj.m641] [Medline: 32071063]
3. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nat Biomed Eng 2017 Jan 30;1(2):0024. [doi: 10.1038/s41551-016-0024]
4. Yeung S, Downing NL, Fei-Fei L, Milstein A. Bedside Computer Vision — Moving Artificial Intelligence from Driver Assistance to Patient Safety. N Engl J Med 2018 Apr 05;378(14):1271-1273. [doi: 10.1056/nejmp1716891]
5. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct;2(10):719-731. [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
6. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. Lancet Digit Health 2020 Apr;2(4):e166-e167. [doi: 10.1016/s2589-7500(20)30054-6]
7. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/nejmp2003539]
8. COVID-19 Dashboard by the Center for Systems Science and Engineering at Johns Hopkins University. Johns Hopkins University of Medicine Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-06-29]
9. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
10. Management of ill travellers at points of entry – international airports, seaports and ground crossings – in the context of COVID -19 outbreak. World Health Organization. 2020 Mar 19. URL: https://www.who.int/publications/i/item/10665-331512 [accessed 2020-06-29]
11. Lee I, Wang C, Lin M, Kung C, Lan K, Lee C. Effective strategies to prevent coronavirus disease-2019 (COVID-19) outbreak in hospital. J Hosp Infect 2020 May;105(1):102-103 [FREE Full text] [doi: 10.1016/j.jhin.2020.02.022] [Medline: 32142886]
12. Mestha L, Shilla E, Bernal E, Pennington G, Madhu H. Breathing pattern identification for respiratory function assessment. Google Patents. 2019. URL: https://patents.google.com/patent/US20150094606 [accessed 2020-06-29]
13. Huang C, Xu X, Cai Y, Ge Q, Zeng G, Li X, et al. Mining the Characteristics of COVID-19 Patients in China: Analysis of Social Media Posts. J Med Internet Res 2020 May 17;22(5):e19087 [FREE Full text] [doi: 10.2196/19087] [Medline: 32401210]
14. Paules CI, Marston HD, Fauci AS. Coronavirus Infections-More Than Just the Common Cold. JAMA 2020 Jan 23:707-708. [doi: 10.1001/jama.2020.0757] [Medline: 31971553]
15. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, et al. Time Course of Lung Changes at Chest CT during Recovery from Coronavirus Disease 2019 (COVID-19). Radiology 2020 Jun;295(3):715-721. [doi: 10.1148/radiol.2020200370]
16. Dembe AE. Ethical Issues Relating to the Health Effects of Long Working Hours. J Bus Ethics 2008 Mar 14;84(S2):195-208. [doi: 10.1007/s10551-008-9700-9]
17. Zheng X, Sun S, Mukkamala RR, Vatrapu R, Ordieres-Meré J. Accelerating Health Data Sharing: A Solution Based on the Internet of Things and Distributed Ledger Technologies. J Med Internet Res 2019 Jun 06;21(6):e13583 [FREE Full text] [doi: 10.2196/13583] [Medline: 31172963]

18. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and Facilitators That Influence Telemedicine-Based, Real-Time, Online Consultation at Patients' Homes: Systematic Literature Review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
19. Salehi H, Das S, Biswas S, Burgueño R. Data mining methodology employing artificial intelligence and a probabilistic approach for energy-efficient structural health monitoring with noisy and delayed signals. Expert Systems with Applications 2019 Nov;135:259-272. [doi: 10.1016/j.eswa.2019.05.051]
20. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. Radiology 2020 Feb 19:200432 [FREE Full text] [doi: 10.1148/radiol.2020200432] [Medline: 32073353]
21. Adly AS, Adly AS, Adly MS, Serry ZMH. Laser acupuncture versus reflexology therapy in elderly with rheumatoid arthritis. Lasers Med Sci 2017 Jul;32(5):1097-1103. [doi: 10.1007/s10103-017-2213-y] [Medline: 28451815]
22. Adly A, Adly A, Adly M, Mostafa Y, Sayed A. Effectiveness of laser acupuncture as adjunctive therapy for limited range of motion in rheumatoid arthritis. Laser Phys 2018 Dec 18;29(2):025601. [doi: 10.1088/1555-6611/aaf327]
23. Srinivasa Rao ASR, Vazquez JA. Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey when cities and towns are under quarantine. Infect Control Hosp Epidemiol 2020 Mar 03:1-5 [FREE Full text] [doi: 10.1017/ice.2020.61] [Medline: 32122430]
24. Robson B. Computers and viral diseases. Preliminary bioinformatics studies on the design of a synthetic vaccine and a preventative peptidomimetic antagonist against the SARS-CoV-2 (2019-nCoV, COVID-19) coronavirus. Comput Biol Med 2020 Apr;119:103670 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.103670] [Medline: 32209231]
25. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, et al. Structure of M from SARS-CoV-2 and discovery of its inhibitors. Nature 2020 Jun;582(7811):289-293. [doi: 10.1038/s41586-020-2223-y] [Medline: 32272481]
26. Grizzle AJ, Horn J, Collins C, Schneider J, Malone DC, Stottlemyer B, et al. Identifying Common Methods Used by Drug Interaction Experts for Finding Evidence About Potential Drug-Drug Interactions: Web-Based Survey. J Med Internet Res 2019 Jan 04;21(1):e11182 [FREE Full text] [doi: 10.2196/11182] [Medline: 30609981]
27. Wang C, Lin P, Cheng C, Tai S, Kao Yang Y, Chiang J. Detecting Potential Adverse Drug Reactions Using a Deep Neural Network Model. J Med Internet Res 2019 Feb 06;21(2):e11016 [FREE Full text] [doi: 10.2196/11016] [Medline: 30724742]
28. Baysari MT, Zheng WY, Van Dort B, Reid-Anderson H, Gronski M, Kenny E. A Late Attempt to Involve End Users in the Design of Medication-Related Alerts: Survey Study. J Med Internet Res 2020 Mar 13;22(3):e14855 [FREE Full text] [doi: 10.2196/14855] [Medline: 32167479]
29. Ho D. Addressing COVID‐19 Drug Development with Artificial Intelligence. Advanced Intelligent Systems 2020 Apr 27;2(5):2000070. [doi: 10.1002/aisy.202000070]
30. Abdullah R, Fakieh B. Health Care Employees' Perceptions of the Use of Artificial Intelligence Applications: Survey Study. J Med Internet Res 2020 May 14;22(5):e17620 [FREE Full text] [doi: 10.2196/17620] [Medline: 32406857]
31. Funk B, Sadeh-Sharvit S, Fitzsimmons-Craft EE, Trockel MT, Monterubio GE, Goel NJ, et al. A Framework for Applying Natural Language Processing in Digital Health Interventions. J Med Internet Res 2020 Feb 19;22(2):e13855 [FREE Full text] [doi: 10.2196/13855] [Medline: 32130118]
32. Wang J, Deng H, Liu B, Hu A, Liang J, Fan L, et al. Systematic Evaluation of Research Progress on Natural Language Processing in Medicine Over the Past 20 Years: Bibliometric Study on PubMed. J Med Internet Res 2020 Jan 23;22(1):e16816 [FREE Full text] [doi: 10.2196/16816] [Medline: 32012074]
33. Dai W, Zhang H, Yu J, Xu H, Chen H, Luo S, et al. CT Imaging and Differential Diagnosis of COVID-19. Can Assoc Radiol J 2020 May;71(2):195-200 [FREE Full text] [doi: 10.1177/0846537120913033] [Medline: 32129670]
34. Zhao M, Yang CC. Drug Repositioning to Accelerate Drug Development Using Social Media Data: Computational Study on Parkinson Disease. J Med Internet Res 2018 Oct 11;20(10):e271 [FREE Full text] [doi: 10.2196/jmir.9646] [Medline: 30309833]
35. Ke Y, Peng T, Yeh T, Huang W, Chang S, Wu S, et al. Artificial intelligence approach fighting COVID-19 with repurposing drugs. Biomed J 2020 May 15:in press [FREE Full text] [doi: 10.1016/j.bj.2020.05.001] [Medline: 32426387]
36. Fagherazzi G, Goetzinger C, Rashid M, Aguayo G, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. J Med Internet Res 2020 Jun 16;22(6):e19284. [doi: 10.2196/19284] [Medline: 32501804]
37. Zhao S, Lin Q, Ran J, Musa S, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int J Infect Dis 2020 Mar;92:214-217 [FREE Full text] [doi: 10.1016/j.ijid.2020.01.050] [Medline: 32007643]
38. Wang W, Tang J, Wei F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. J Med Virol 2020 Apr;92(4):441-447 [FREE Full text] [doi: 10.1002/jmv.25689] [Medline: 31994742]
39. Mahomed N, van Ginneken B, Philipsen RHHM, Melendez J, Moore DP, Moodley H, et al. Computer-aided diagnosis for World Health Organization-defined chest radiograph primary-endpoint pneumonia in children. Pediatr Radiol 2020 Apr;50(4):482-491. [doi: 10.1007/s00247-019-04593-0] [Medline: 31930429]
40. Wiemken T, Kelley R, Mattingly W, Ramirez J. Clinical Research in Pneumonia: Role of Artificial Intelligence. JRI 2019;3(1):1-4. [doi: 10.18297/jri/vol3/iss1/1]

41. Abdullah A, Posdzi N, Nishio Y. Preliminary study of pneumonia symptoms detection method using Cellular Neural Network. 2011 Presented at: International Conference on Electrical, Control and Computer Engineering; June 21-22, 2011; Pahang, Malaysia p. 497-500. [doi: 10.1109/inecce.2011.5953933]
42. O'Quinn W, Haddad R, Moore D. Pneumonia Radiograph Diagnosis Utilizing Deep Learning Network. 2019 Presented at: 2019 IEEE 2nd International Conference on Electronic Information and Communication Technology; January 20-22, 2019; Harbin, China p. 763-767. [doi: 10.1109/iceict.2019.8846438]
43. Yuan X, Huang B, Wang Y, Yang C, Gui W. Deep Learning-Based Feature Representation and Its Application for Soft Sensor Modeling With Variable-Wise Weighted SAE. IEEE Trans Ind Inf 2018 Jul;14(7):3235-3243. [doi: 10.1109/tii.2018.2809730]
44. Mouton A, Pitcher R, Douglas T. Computer-aided detection of pulmonary pathology in pediatric chest radiographs. Med Image Comput Comput Assist Interv 2010;13(Pt 3):619-625. [doi: 10.1007/978-3-642-15711-0_77] [Medline: 20879452]
45. Qin C, Yao D, Shi Y, Song Z. Computer-aided detection in chest radiography based on artificial intelligence: a survey. Biomed Eng Online 2018 Aug 22;17(1):113 [FREE Full text] [doi: 10.1186/s12938-018-0544-y] [Medline: 30134902]
46. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study. J Med Internet Res 2020 May 28;22(5):e19421 [FREE Full text] [doi: 10.2196/19421] [Medline: 32452804]
47. Nathani B, Vijayvergia R. The Internet of Intelligent things: An overview. 2017 Presented at: 2017 International Conference on Intelligent Communication and Computational Techniques; December 22 to 23, 2017; Jaipur, India. [doi: 10.1109/intelcct.2017.8324031]
48. Mehta V. A Novel Approach to Realize Internet of Intelligent Things. In: Aggarwal V, Bhatnagar V, Mishra D, editors. Big Data Analytics: Advances in Intelligent Systems and Computing. Singapore: Springer; Oct 04, 2017:413-419.
49. Adly A. Technology trade-offs for IIoT systems and applications from a developing country perspective: case of Egypt. In: Mahmood Z, editor. The Internet of Things in the Industrial Sector. Cham: Springer; 2019:299-319.
50. Adly AS. Integrating Vehicular Technologies Within the IoT Environment: A Case of Egypt. In: Mahmood Z, editor. Connected Vehicles in the Internet of Things. Cham: Springer; 2020.
51. Komenda M, Bulhart V, Karolyi M, Jarkovský J, Mužík J, Májek O, et al. Complex Reporting of the COVID-19 Epidemic in the Czech Republic: Use of an Interactive Web-Based App in Practice. J Med Internet Res 2020 May 27;22(5):e19367 [FREE Full text] [doi: 10.2196/19367] [Medline: 32412422]
52. Allam Z, Dhunny ZA. On big data, artificial intelligence and smart cities. Cities 2019 Jun;89:80-91. [doi: 10.1016/j.cities.2019.01.032]
53. Triantafyllidis AK, Tsanas A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. J Med Internet Res 2019 Apr 05;21(4):e12286 [FREE Full text] [doi: 10.2196/12286] [Medline: 30950797]
54. Swamidason ITJ. Survey of data mining algorithm's for intelligent computing system. JTCSST 2019 Sep 8;01(01):14-23. [doi: 10.36548/jtcsst.2019.1.002]
55. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. N Engl J Med 2020 May 21;382(21):2049-2055. [doi: 10.1056/NEJMsb2005114] [Medline: 32202722]
56. Montero-Odasso M, Hogan DB, Lam R, Madden K, MacKnight C, Molnar F, et al. Age Alone is not Adequate to Determine Health-care Resource Allocation During the COVID-19 Pandemic. Can Geriatr J 2020 Mar;23(1):152-154 [FREE Full text] [doi: 10.5770/cgj.23.452] [Medline: 32550953]
57. Wu J, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nat Med 2020 Apr;26(4):506-510 [FREE Full text] [doi: 10.1038/s41591-020-0822-7] [Medline: 32284616]
58. Kucharski A, Russell T, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020 May;20(5):553-558. [doi: 10.1016/S1473-3099(20)30144-4]
59. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial Intelligence and the Implementation Challenge. J Med Internet Res 2019 Jul 10;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
60. Wolff J, Pauling J, Keck A, Baumbach J. The Economic Impact of Artificial Intelligence in Health Care: Systematic Review. J Med Internet Res 2020 Feb 20;22(2):e16866 [FREE Full text] [doi: 10.2196/16866] [Medline: 32130134]
61. Adly AS, Haggag MH, Mostafa MS. Low Intensity Laser Irradiation Influence Proliferation of Mesenchymal Stem Cells: Comparison of Experimental Data to Intelligent Agent-Based Model Predictions. In: Liu L, Zhu Q, Cheng L, Wang Y, Zhao S, editors. Applied Methods and Techniques for Mechatronic Systems. Berlin: Springer; 2014:293-306.
62. Adly AS, Haggag MH, Mostafa MSM. Laser Promotes Proliferation of Stem Cells: A Comprehensive Case Study Consolidated by Intelligent Agent–Based Model Predictions. In: Issac B, Israr N, editors. Case Studies in Intelligent Computing. Boca Raton, FL: Auerbach Publications; 2014:58-87.
63. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus ADME, et al. Mapping the antigenic and genetic evolution of influenza virus. Science 2004 Jul 16;305(5682):371-376 [FREE Full text] [doi: 10.1126/science.1097211] [Medline: 15218094]

64. Salama MA, Hassanien AE, Mostafa A. The prediction of virus mutation using neural networks and rough set techniques. EURASIP J Bioinform Syst Biol 2016 Dec;2016(1):10 [FREE Full text] [doi: 10.1186/s13637-016-0042-0] [Medline: 27257410]
65. Du X, Wang Z, Wu A, Song L, Cao Y, Hang H, et al. Networks of genomic co-occurrence capture characteristics of human influenza A (H3N2) evolution. Genome Res 2008 Jan;18(1):178-187 [FREE Full text] [doi: 10.1101/gr.6969007] [Medline: 18032723]
66. Yin R, Luusua E, Dabrowski J, Zhang Y, Kwoh C. Tempel: time-series mutation prediction of influenza A viruses via attention-based recurrent neural networks. Bioinformatics 2020 May 01;36(9):2697-2704. [doi: 10.1093/bioinformatics/btaa050] [Medline: 31999330]
67. Neher RA, Bedford T, Daniels RS, Russell CA, Shraiman BI. Prediction, dynamics, and visualization of antigenic phenotypes of seasonal influenza viruses. Proc Natl Acad Sci USA 2016 Mar 22;113(12):E1701-E1709 [FREE Full text] [doi: 10.1073/pnas.1525578113] [Medline: 26951657]
68. Pang T, Guo S, Zhang X, Zhao L. Automatic Lung Segmentation Based on Texture and Deep Features of HRCT Images with Interstitial Lung Disease. Biomed Res Int 2019;2019:2045432 [FREE Full text] [doi: 10.1155/2019/2045432] [Medline: 31871932]
69. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
70. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb;395(10223):507-513. [doi: 10.1016/s0140-6736(20)30211-7]
71. Wiemken TL, Furmanek SP, Mattingly WA, Guinn BE, Cavallazzi R, Fernandez-Botran R, et al. Predicting 30-day mortality in hospitalized patients with community-acquired pneumonia using statistical and machine learning approaches. JRI 2017;1(3):10. [doi: 10.18297/jri/vol1/iss3/10/]
72. Santosh KC. AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data. J Med Syst 2020 Mar 18;44(5):93 [FREE Full text] [doi: 10.1007/s10916-020-01562-1] [Medline: 32189081]
73. Adly AS, Kandil OAD, Ibrahim MS, Adly MS, Adly AS, Adly AS. Computational and Theoretical Concepts for Regulating Stem Cells Using Viral and Physical Methods. In: Ao SI, Rieger B, Amouzegar M, editors. Machine Learning and Systems Engineering. Lecture Notes in Electrical Engineering, vol 68. Dordrecht: Springer; 2010:533-546.
74. Adly A, Aboutabl A, Ibrahim M. Modeling of gene therapy for regenerative cells using intelligent agents. Adv Exp Med Biol 2011;696:317-325. [doi: 10.1007/978-1-4419-7046-6_32] [Medline: 21431572]
75. Mahrouf M, Lotfi E, Maziane M, Hattaf K, Yousfi N. Stability Analysis for Stochastic Differential Equations in Virology. BJMCS 2017 Jan 10;20(1):1-12. [doi: 10.9734/bjmcs/2017/30345]
76. Hulo C, de Castro E, Masson P, Bougueleret L, Bairoch A, Xenarios I, et al. ViralZone: a knowledge resource to understand virus diversity. Nucleic Acids Res 2011 Jan;39(Database issue):D576-D582 [FREE Full text] [doi: 10.1093/nar/gkq901] [Medline: 20947564]
77. Jefferys E, Sansom M. Computational virology: molecular simulations of virus dynamics and interactions. Physical Virology: Springer; 2019:201-233.
78. Panzirsch M, Weber B, Rubio L, Coloma S, Ferre M, Artigas J. Tele-healthcare with humanoid robots: A user study on the evaluation of force feedback effects. 2017 Presented at: 2017 IEEE World Haptics Conference; June 6-9, 2017; Munich, Germany p. 245-250. [doi: 10.1109/whc.2017.7989909]
79. Li Z, Moran P, Dong Q, Shaw R, Hauser K. Development of a tele-nursing mobile manipulator for remote care-giving in quarantine areas. 2017 Presented at: 2017 IEEE International Conference on Robotics and Automation; May 29-June 3, 2017; Singapore p. 3581-3586. [doi: 10.1109/icra.2017.7989411]
80. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine During the COVID-19 Pandemic: Experiences From Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
81. deBronkart D, Eysenbach G. Gimme My Damn Data (and Let Patients Help!): The #GimmeMyDamnData Manifesto. J Med Internet Res 2019 Nov 22;21(11):e17045 [FREE Full text] [doi: 10.2196/17045] [Medline: 31755873]
82. Joshi AU, Randolph FT, Chang AM, Slovis BH, Rising KL, Sabonjian M, et al. Impact of Emergency Department Tele-intake on Left Without Being Seen and Throughput Metrics. Acad Emerg Med 2020 Feb 26;27(2):139-147. [doi: 10.1111/acem.13890] [Medline: 31733003]
83. Lewis P, Catanzano T, Davis L, Jordan S. Web-based Conferencing: What Radiology Educators Need to Know. Acad Radiol 2020 Mar;27(3):447-454. [doi: 10.1016/j.acra.2019.05.017] [Medline: 31300359]
84. Goodman-Casanova JM, Dura-Perez E, Guzman-Parra J, Cuesta-Vargas A, Mayoral-Cleries F. Telehealth Home Support During COVID-19 Confinement for Community-Dwelling Older Adults With Mild Cognitive Impairment or Mild Dementia: Survey Study. J Med Internet Res 2020 May 22;22(5):e19434 [FREE Full text] [doi: 10.2196/19434] [Medline: 32401215]
85. Adly MS, Adly AS, Adly AS. Assessment of early orthodontic treatment on functional shifts by telemonitoring mandibular movements using a smart phone. J Telemed Telecare 2018 Oct 29;26(3):150-160. [doi: 10.1177/1357633x18800876]

86. Tanne JH, Hayasaki E, Zastrow M, Pulla P, Smith P, Rada AG. Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. BMJ 2020 Mar 18;368:m1090. [doi: 10.1136/bmj.m1090] [Medline: 32188598]

Abbreviations

AI: artificial intelligence
COVID-19: coronavirus disease
CT: computed tomography
EHR: electronic health record
ETHAN: Emergency Telehealth and Navigation
IoIT: Internet of Intelligent Things
IoT: Internet of Things
PDA: personal digital assistant
RT-PCR: reverse transcriptase–polymerase chain reaction
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
TRINA: Tele-Robotic Intelligent Nursing Assistant

Edited by G Eysenbach, G Fagherazzi; submitted 03.04.20; peer-reviewed by S Mohamed, P Giabbanelli, F Alvarez-Lopez; comments to author 18.06.20; revised version received 24.06.20; accepted 25.06.20; published 10.08.20.

Please cite as:
Adly AS, Adly AS, Adly MS
Approaches Based on Artificial Intelligence and the Internet of Intelligent Things to Prevent the Spread of COVID-19: Scoping Review
J Med Internet Res 2020;22(8):e19104
URL: https://www.jmir.org/2020/8/e19104
doi:10.2196/19104
PMID:32584780

Original Paper

Assessment of the Impact of Media Coverage on COVID-19–Related Google Trends Data: Infodemiology Study

Bernardo Sousa-Pinto[1,2], MD, PhD; Aram Anto[3], MSc; Wienia Czarlewski[3,4], MD, PhD; Josep M Anto[5,6,7], MD, PhD; João Almeida Fonseca[1,2], MD, PhD; Jean Bousquet[8,9,10], MD, PhD

[1]Department of Community Medicine, Information and Health Decision Sciences, Faculty of Medicine, University of Porto, Porto, Portugal
[2]Center for Health Technology and Services Research, University of Porto, Porto, Portugal
[3]MASK-air, Montpellier, France
[4]Medical Consulting Czarlewski, Levallois, France
[5]Centre for Research in Environmental Epidemiology, Barcelona Institute for Global Health, Barcelona, Spain
[6]Universitat Pompeu Fabra, Barcelona, Spain
[7]CIBER Epidemiología y Salud Pública, Barcelona, Spain
[8]Charité, Universitätsmedizin Berlin, Humboldt-Universität zu Berlin, Berlin, Germany
[9]Comprehensive Allergy Center, Department of Dermatology and Allergy, Berlin Institute of Health, Berlin, Germany
[10]MACVIA-France, Montpellier, France

Corresponding Author:
Bernardo Sousa-Pinto, MD, PhD
Department of Community Medicine, Information and Health Decision Sciences
Faculty of Medicine
University of Porto
Rua Plácido Costa s/n
Porto, 4400
Portugal
Phone: 351 912362153
Email: bernardosousapinto@protonmail.com

Abstract

Background: The influence of media coverage on web-based searches may hinder the role of Google Trends (GT) in monitoring coronavirus disease (COVID-19).

Objective: The aim of this study was to assess whether COVID-19–related GT data, particularly those related to ageusia and anosmia, were primarily related to media coverage or to epidemic trends.

Methods: We retrieved GT query data for searches on *coronavirus*, *cough*, *anosmia*, and *ageusia* and plotted them over a period of 5 years. In addition, we analyzed the trends of those queries for 17 countries throughout the year 2020 with a particular focus on the rises and peaks of the searches. For *anosmia* and *ageusia*, we assessed whether the respective GT data correlated with COVID-19 cases and deaths both throughout 2020 and specifically before March 16, 2020 (ie, the date when the media started reporting that these symptoms can be associated with COVID-19).

Results: Over the last five years, peaks for *coronavirus* searches in GT were only observed during the winter of 2020. Rises and peaks in *coronavirus* searches appeared at similar times in the 17 different assessed countries irrespective of their epidemic situations. In 15 of these countries, rises in *anosmia* and *ageusia* searches occurred in the same week or 1 week after they were identified in the media as symptoms of COVID-19. When data prior to March 16, 2020 were analyzed, *anosmia* and *ageusia* GT data were found to have variable correlations with COVID-19 cases and deaths in the different countries.

Conclusions: Our results indicate that COVID-19–related GT data are more closely related to media coverage than to epidemic trends.

(J Med Internet Res 2020;22(8):e19611) doi:10.2196/19611

KEYWORDS

COVID-19; infodemiology; infodemic; Google Trends; media coverage; media; coronavirus; symptom; monitoring; trend; pandemic

Introduction

Infodemiology is defined as "the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy" [1,2]. This field comprises both "supply-based" and "demand-based" infodemiology, with the latter assessing individuals' health-seeking behavior (eg, through online searches) [2]. Over the years, infodemiological studies have become increasingly popular, focusing on different fields such as chronic diseases, risk behaviors, and infectious diseases [3,4]. Regarding the latter, the use of search query data to predict or monitor infectious outbreaks can be traced to back to the 2002 severe acute respiratory syndrome (SARS) epidemic [5]. Subsequent studies have been conducted on influenza and other infectious diseases. Google Trends (GT) is one of the most commonly used data sources, albeit with mixed results. In fact, despite the initial optimism regarding the use of GT for influenza prediction (Google Flu Trends) [6] and despite the strong correlation of data with influenza-related emergency department visits [7], the unsatisfactory performance of Google Flu Trends led to its discontinuation [8].

In the context of the coronavirus disease (COVID-19) pandemic, there has been interest in GT (or other data on web-based activity), particularly concerning the potential role of these data in defining the proper timing and location for practicing appropriate risk communication strategies to affected populations [9]. In Europe, significant correlations were observed between COVID-19 cases and deaths and online interest on this topic [10]. In addition, GT data were found to predict COVID-19 incidence in Iran [11]. In contrast, as the number of COVID-19 cases increased, interest in telehealth and telemedicine among the US population did not correlate with the proportion of hospitals providing telehealth services [12].

Using GT to obtain information regarding COVID-19, presents two difficulties. One is that information demand may be disproportionate to the epidemiologic on account of media coverage (as described in other contexts [5]), and the other is the low specificity of the main COVID-19 symptoms. However, regarding the latter, while cough, fever and dyspnea can also occur in several other diseases, some more specific manifestations of COVID-19 have been described. Two symptoms that appear to be more specific are anosmia and ageusia [13]. This was not widely known to the general public before the publication of an interview with Hendrik Streeck in the German newspaper *Frankfurter Allgemeine Zeitung* on March 16, 2020 [14], which was then cited by media worldwide. The identification of these more specific symptoms raised interest in whether GT data for these manifestations could better correlate with COVID-19 incidence and deaths than data for less specific symptoms. While strong correlations between searches for smell-related information and the number of COVID-19 cases and deaths have been described in several countries [15], the role of media coverage in motivating smell-related searches cannot be disregarded.

Therefore, we aimed to assess whether searches for the terms *anosmia* and *ageusia* were primarily related to media releases or to COVID-19 epidemic trends.

Methods

This is a GT-based infodemiology study that complies with the methodological framework described by Mavragani and Ochoa [16].

Keyword Selection

In this study, we retrieved GT data on the keywords *coronavirus* (as a virus and search term), *cough* (as a topic), *anosmia* (as a disease), and *ageusia* (as a topic).

With the exception of *coronavirus*, no other nontopic or nondisease search terms were used. In fact, we tested the search terms *loss of smell*, *hyposmia*, *olfaction*, *dysgeusia*, and *loss of taste* [8] using translations of the terms into native languages of the studied countries (using double quotation marks when searching for keywords containing more than one word); however, the data retrieved with these queries were not consistent or of sufficient quality.

Region and Period Selection

We obtained country-level GT data for all analyses except for the worldwide analysis of the last five years. We retrieved GT data for the following time periods:

- A time frame of the last five years (up to the week of April 5 to 11, 2020): This time frame allowed us to assess worldwide search spikes of selected keywords over a long-term period.
- A time frame comprising the year 2020 (ie, the period ranging from the week of January 5 to 11 to the week of April 5 to 11, 2020): This time frame allowed us to identify the search trends for selected keywords throughout the year 2020 in 17 Western countries (where search data for *anosmia* were sufficient to perform an analysis). These GT data were plotted (without performing formal correlations) alongside data on COVID-19 cases in different countries. Note that for this time frame, we retrieved data starting on January 5 (and not on the date that the first COVID-19 case was registered in each country), not only to allow between-country comparison but also because in the Western World, news coverage on SARS-CoV-2 infection started before the first confirmed cases were identified, and also because it is possible that there were COVID-19 cases in the Western World prior to the first identified cases (which may have been reflected in symptom web searches).
- A time frame ranging from the date of the first confirmed COVID-19 case in each country until March 15: This time frame allowed a closer analysis of search trends before the media started reporting that anosmia and ageusia can be

symptoms of COVID-19. To assess the impact of this media coverage, we analyzed 8 different countries and correlated web searches with the respective data on COVID-19 cases for that period. Correlations with the daily number of deaths were also performed (in this case, using a time frame ranging from the date of the first death in each country until March 15, 2020).

Search Categories

Categories and subcategories were not selected when searching for keywords.

Data Analysis

After plotting worldwide GT data on the selected keywords for the last five years, we retrieved GT data for the year 2020 and assessed the trends of those queries in the 17 countries where searches for *anosmia* were sufficient to perform an analysis.

To further assess the impact of media coverage on COVID-19–related GT data, and to assess whether the GT data correlated with COVID-19 cases, we focused on 8 countries in different stages of the COVID-19 pandemic: France, Germany, Italy, Portugal, Spain, the United Kingdom, Brazil, and the United States. For each country, we plotted the weekly GT data for selected keywords together with weekly data on new COVID-19 cases (numbers retrieved from official sources).

Subsequently, we performed an analysis restricted to the time period prior to March 16, 2020, the date that the media started reporting that anosmia and ageusia can be symptoms of COVID-19. In fact, from that date onward, GT data could largely reflect interest in media coverage rather than searches for symptoms that patients were experiencing. Therefore, for each country, between the date of the first confirmed COVID-19 case and March 15, 2020, we assessed the correlation (by means of the Pearson correlation coefficient, r) between the daily average of GT for *anosmia* and *ageusia* (herein reported as *anosmia/ageusia*) and daily data on new COVID-19 cases. Similar analyses were performed for new COVID-19 deaths (in the time frame from the date of the first COVID-19 death to March 15, 2020).

To facilitate plot reading, we plotted normalized weekly data on COVID-19 cases and deaths. That is, we plotted the total number of new COVID-19 cases and deaths as percentages of the respective maximum weekly values observed during the defined time period.

Results

On a worldwide scale, the GT data for different countries showed peaks appearing at similar times, with higher peaks for *coronavirus* than for other searches (Figure 1). Throughout five years, these peaks were only found in the winter of 2020. Queries for *coronavirus* as a virus and as a search term peaked at the same time. The second highest peaks were for *cough*, with two peaks coinciding with those of *coronavirus*. *Anosmia* and *ageusia* had slightly delayed identical peaks that were not identified when the terms *cough* or *coronavirus* were searched.

We analyzed search trends in 17 countries where the *anosmia* peak was clearly identifiable (in 2 of these countries, *ageusia* did not show any peak, and in 2 others, no peak was observed for *cough*). In particular, we started by observing the week when searches for each topic started to rise (Table 1). The first rise in *coronavirus* searches started in late January 2020, while the second peak of *coronavirus* searches appeared between February 16 and 22 in one country (Italy) and between February 23 and 29 in the remaining countries. *Cough* queries started in the same week in 3 countries, 1 week later in 4 countries, 2 weeks later in 2 countries, and more than 2 weeks later in 6 countries.

Figure 1. Global GT data on *coronavirus*, *cough*, *anosmia*, and *ageusia*. Data are presented as a percentage of the maximum value and on a weekly basis for periods of 5 years and 1 year up to the week of April 5 to 11, 2020. GT: Google Trends.

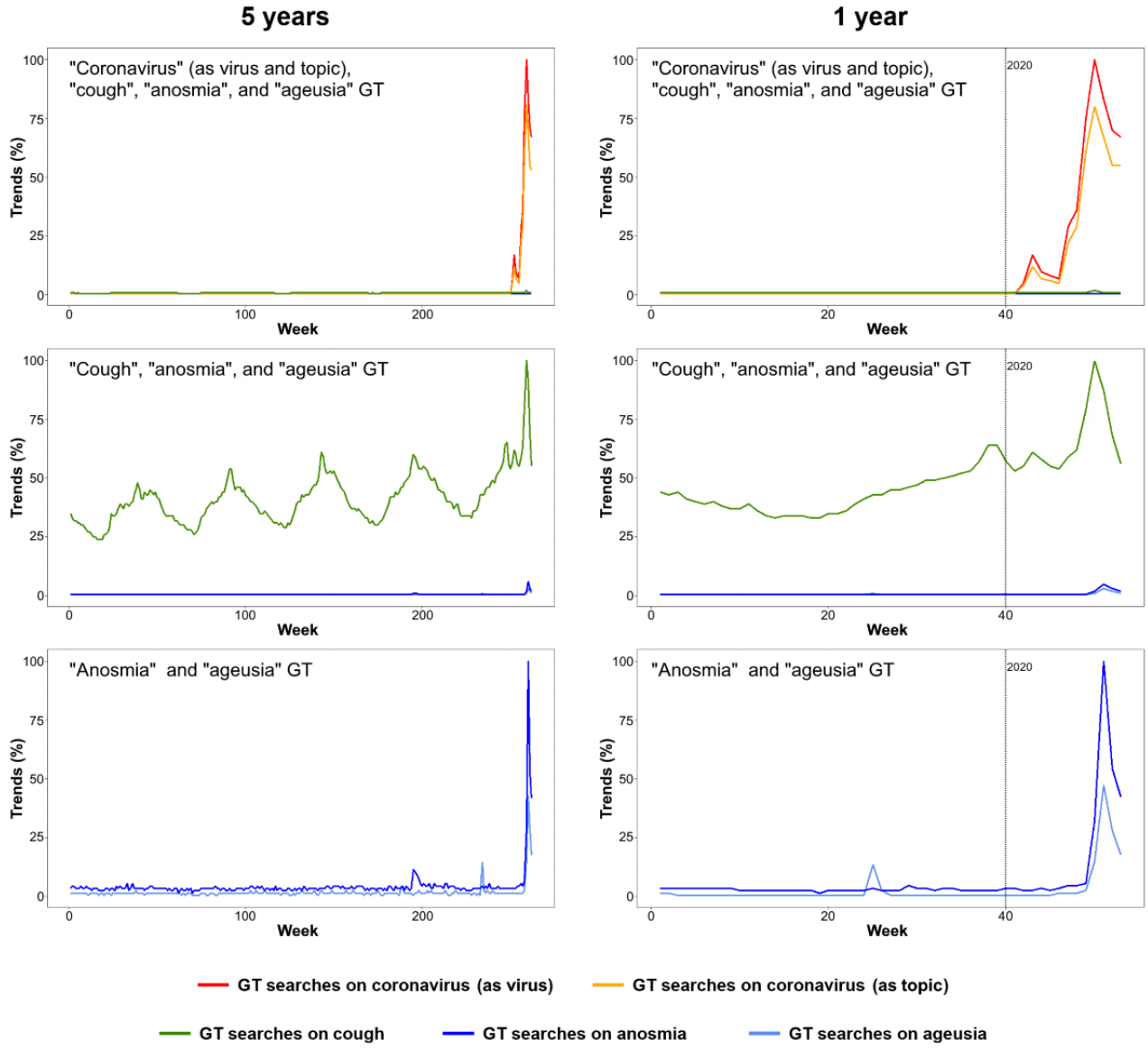


Table 1. Weeks of onset of Google Trends peaks for search terms related to COVID-19 in 2020 in 17 countries.

Country	GT[a] peak onset for *coronavirus* (as a virus)[b]		GT peak onset for *anosmia* (as a disease)	GT peak onset for *ageusia* (as a topic)	GT peak onset for *cough* (as a topic)
	Peak 1	Peak 2			
Argentina	January 19 to 25	February 23 to 29	March 22 to 28[c]	N/A[d]	March 8 to 14
Australia	January 19 to 25	February 23 to 29	March 22 to 28[e]	March 15 to 21[e]	February 16 to 22
Belgium	January 19 to 25	February 23 to 29	March 15 to 21[e]	March 15 to 21[e]	February 16 to 22
Brazil	January 19 to 25	February 23 to 29	March 15 to 21[e]	March 15 to 21[e]	February 16 to 22
Canada	January 19 to 25	February 23 to 29	March 15 to 21[e]	March 22 to 28[e]	March 8 to 14
Chile	January 19 to 25	February 23 to 29	March 22 to 28[c]	March 22 to 28[e]	February 23 to 29
France	January 19 to 25	February 23 to 29	March 15 to 21[e]	March 15 to 21[e]	March 1 to 7
Germany	January 19 to 25	February 23 to 29	March 15 to 21[e]	March 15 to 21[e]	February 23 to 29
Italy	January 19 to 25	February 16 to 22	March 1 to 7[e]	March 8 to 14[e]	February 23 to 29
Portugal	January 19 to 25	February 23 to 29	March 15 to 21[e]	March 15 to 21[e]	March 8 to 14
Russia	January 19 to 25	February 23 to 29	March 22 to 28[e]	N/A	N/A
Spain	January 19 to 25	February 23 to 29	March 15 to 21[e]	March 8 to 14[e]	March 8 to 14
Sweden	January 19 to 25	February 23 to 29	March 15 to 21[e]	March 22 to 28[e]	N/A
Switzerland	January 19 to 25	February 23 to 29	March 8 to 14[e]	March 15 to 21[e]	March 1 to 7
The Netherlands	January 19 to 25	February 23 to 29	March 15 to 21[e]	March 22 to 28[e]	February 23 to 29
United Kingdom	January 19 to 25	February 23 to 29	March 15 to 21[e]	March 15 to 21[e]	March 8 to 14
United States	January 19 to 25	February 23 to 29	March 22 to 28[e]	March 15 to 21[e]	March 8 to 14

[a]GT: Google Trends.

[b]Two GT peaks consistently appeared for *coronavirus*. Peak 1 is a minor peak that appeared by late January 2020, and Peak 2 is the largest Google Trends peak.

[c]GT data peaked in the week of April 5 to 11.

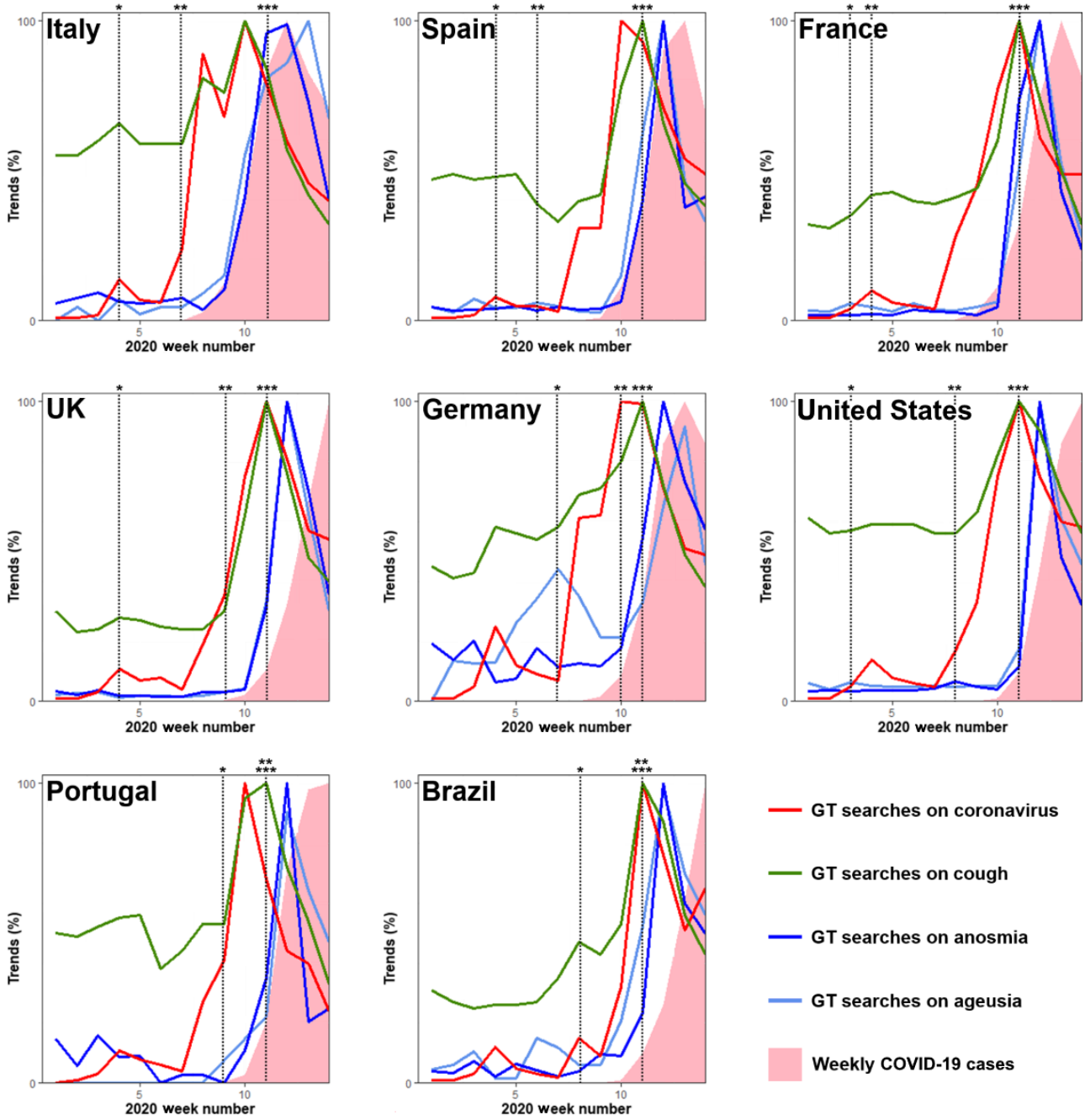
[d]Not applicable.

[e]GT data peaked in the week of March 15 to 21.

We observed that the onset of *anosmia* queries occurred from March 15 to 21 in 10 countries (corresponding to the week of the Hendrik Streeck interview in *Frankfurter Allgemeine Zeitung*) and from March 22 to 28 in 5 other countries (in Italy and Switzerland, the queries started before the week of March 15 to 21). The weeks of onset for *ageusia* and *anosmia* queries were the same in 7/15 countries (47%). The GT peaks for *anosmia* and *ageusia* were first observed from March 22 to 28, 2020, for all countries except Argentina and Chile (one week after the Streeck interview).

Subsequently, we analyzed 8 countries by plotting the average GT data for *anosmia* and *ageusia* with the number of COVID-19 cases. We observed that the GT peak coincided with the maximum weekly number of new COVID-19 cases in Italy but not in the other countries (Figure 2). For all countries (except Italy and Germany), the GT peaks were followed by sharp decreases.

Figure 2. GT data for *coronavirus*, *cough*, *anosmia*, and *ageusia* and relative frequency of new COVID-19 infections. Data are presented as a percentage of the maximum value on a weekly basis, from the week of January 5 to 11, 2020, to the week of April 5 to 11, 2020. *First confirmed COVID-19 case. **First confirmed death due to COVID-19. ***Hendrik Streeck interview to Frankfurter Allgemeine Zeitung reporting that anosmia and ageusia can be COVID-19 symptoms. COVID-19: coronavirus disease. GT: Google Trends.



Analyzing the data from the date of first confirmed case of COVID-19 until March 16, we observed that in countries with higher COVID-19 infection or death rates, there were moderate to good correlations between Google Trends for *anosmia*/*ageusia* and new COVID-19 cases or deaths (Table 2, Table 3, and Figure 3). By contrast, poor correlations were observed in countries with lower COVID-19 rates by March 15. The only exception was the United Kingdom, in which we observed strong correlations between Google Trends searches on *anosmia*/*ageusia* and new COVID-19 cases (r=0.739) and deaths (r=0.668) despite the low COVID-19 infection and death rates (0.3 deaths per million inhabitants).

These results are supported by between-countries comparisons (Figure 4). Prior to March 16, Italy was the country with the largest volume of searches for *anosmia*/*ageusia*; however, it was surpassed by France, the United Kingdom, and Spain following extensive media coverage of those symptoms.

Table 2. Frequency of new COVID-19 cases and deaths in the countries examined in the study. The analysis time frame for new COVID-19 cases was from the date of the first confirmed COVID-19 case in the respective country until March 15, 2020. The analysis time frame for COVID-19 deaths was from the date of the first confirmed COVID-19 death in the respective country until March 15, 2020.

Country	COVID-19[a] cases per million inhabitants as of March 15, 2020	COVID-19 deaths per million inhabitants as of March 15, 2020
Italy	411.1	30.0
Spain	169.6	6.24
France	80.8	1.89
United Kingdom	20.9	0.32
Germany	58.2	0.14
United States	10.6	0.19
Portugal	23.8	0
Brazil	0.95	0

[a]COVID-19: coronavirus disease.

Table 3. Pearson correlation coefficients between Google Trends data on anosmia/ageusia and the frequency of new COVID-19 cases and deaths in Table 2.

Country	Correlations with average GT[a] searches for *anosmia* (as a disease)/*ageusia* (as a topic)				Correlations with GT searches for *anosmia* (as a disease)				Correlations with GT searches for *ageusia* (as a topic)			
	COVID-19[b] cases, *r*	*P* value	COVID-19 deaths, *r*	*P* value	COVID-19 cases, *r*	*P* value	COVID-19 deaths, *r*	*P* value	COVID-19 cases, *r*	*P* value	COVID-19 deaths, *r*	*P* value
Italy	0.796	<.001	0.776	<.001	0.810	<.001	0.855	<.001	0.646	<.001	0.621	.001
Spain	0.568	<.001	0.755	<.001	0.460	.001	0.531	.002	0.506	<.001	0.632	<.001
France	0.552	<.001	0.761	<.001	0.434	.001	0.575	.008	0.438	.001	0.647	.002
United Kingdom	0.739	<.001	0.668	.025	0.663	<.001	0.745	.008	0.457	.002	0.500	.118
Germany	–0.005	.975	N/A[c]	N/A	0.104	.478	N/A	N/A	–0.099	.500	N/A	N/A
United States	–0.081	.559	–0.141	.602	0.015	.916	–0.545	.029	–0.115	.404	0.331	.210
Portugal	–0.312	.277	N/A	N/A	–0.229	.431	N/A	N/A	–0.182	.534	N/A	N/A
Brazil	–0.014	.953	N/A	N/A	–0.031	.899	N/A	N/A	0.003	.990	N/A	N/A

[a]GT: Google Trends.

[b]COVID-19: coronavirus disease.

[c]Not applicable.

Figure 3. Average GT data for *anosmia* and *ageusia* and relative frequency of new COVID-19 infections and deaths. Data are presented as Loess-smoothed percentages of the maximum value (smoothed trends) on a daily basis from February 1 to March 15, 2020 (before the media publicized that anosmia and ageusia can be symptoms of COVID-19). Lines were smoothed to display trends more clearly. COVID-19: coronavirus disease. GT: Google Trends.

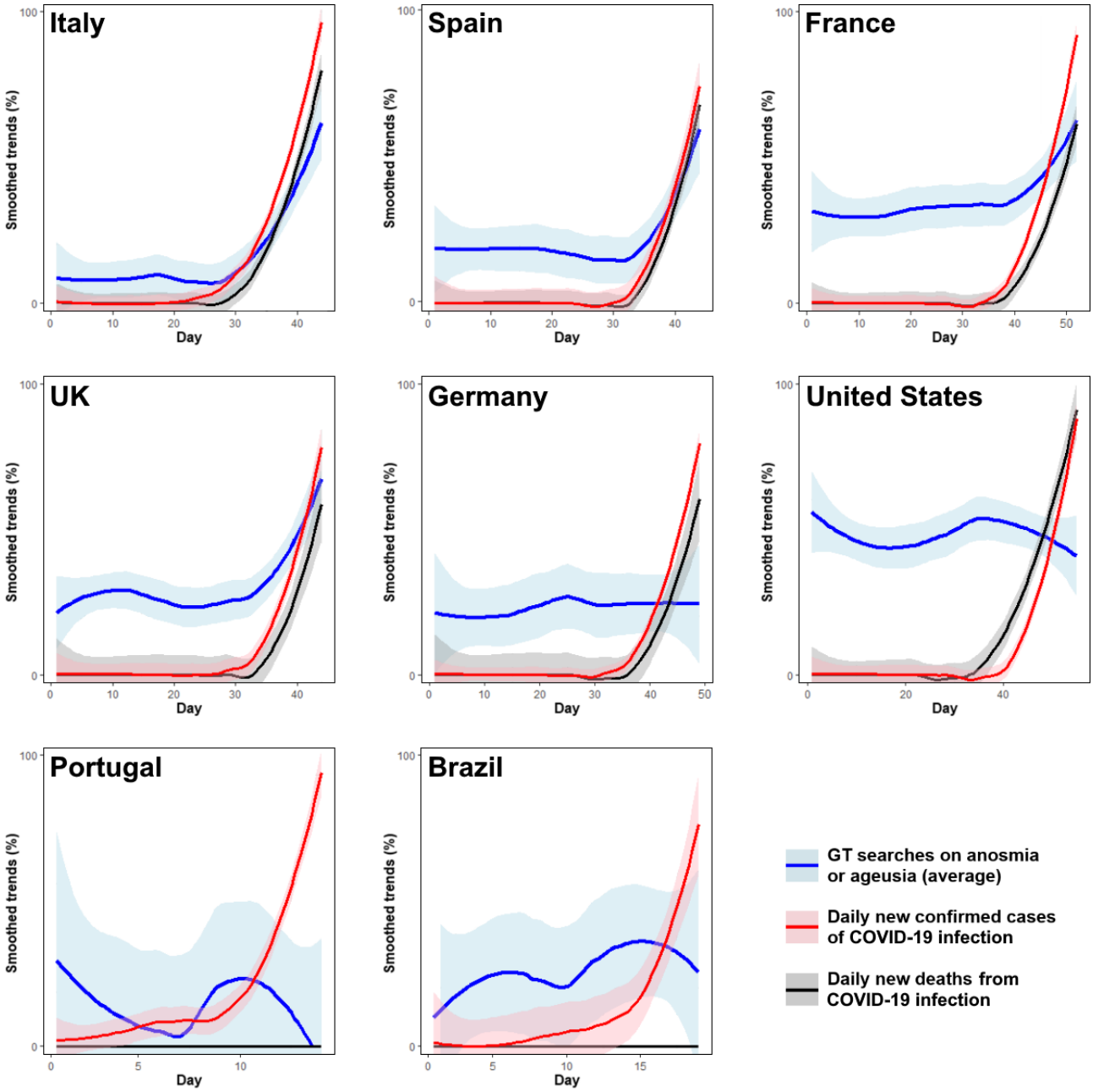
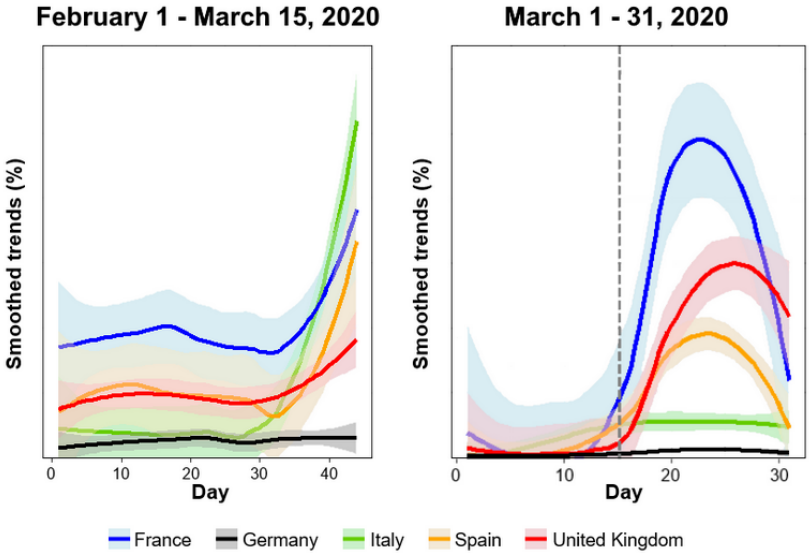


Figure 4. Average GT data for *anosmia* and *ageusia* before the media publicized that these terms can be symptoms of COVID-19 (February 1 to March 15, 2020) and in the 2 weeks before and after this media release (marked with a grey dashed line) (March 1 to 31, 2020). Data are presented on a daily basis as Loess-smoothed percentages of the maximum value and are adjusted for the population.



Discussion

Principal Findings

The results of this study suggest that COVID-19–related GT queries do not necessarily follow the evolution of the epidemic and, in particular for *anosmia* and *ageusia*, are more closely related to media coverage.

Using a stepwise approach based on 1- and 5-year perspectives, we showed that search peaks not only for *coronavirus* but also for *anosmia*/*ageusia* appeared for the first time in 2020; also, there may be a relationship between the two peaks. This is different from the *cough* search term, for which searches were detected for all years but which also showed a peak coincident with the *coronavirus* peak.

We then assessed countries with an identifiable *anosmia* peak in 2020 in the northern and southern hemispheres. Surprisingly, in all countries, peaks for *coronavirus*, *cough*, and *anosmia*/*ageusia* all occurred simultaneously, irrespective of the pandemic stage. A simple interpretation is that this is unlikely to be associated with COVID-19 incidence. However, the time of onset differed for *coronavirus* or *cough* versus *anosmia* or *ageusia*; the latter coincided with the timing with which media news covered information on these symptoms (Table 4).

Table 4. Media coverage on the identification of anosmia and ageusia as COVID-19 symptoms.

Language	Date	Title of index media news	Source
German	March 16, 2020	*Virologe Hendrik Streeck : "Wir haben neue Symptome entdeckt"*	*Frankfurt Allgemeine Zeitung* [14]
Italian	March 17, 2020	*Coronavirus, tra i sintomi frequenti la perdita totale di gusto e olfatto*	*Corriere de la Serra* [17]
English (United Kingdom)	March 17, 2020	Coronavirus symptoms shock: Scientists discover NEW symptoms including lack of taste	*Daily Express* [18]
English (United States)	March 13, 2020	Coronavirus is most contagious before and during the first week of symptoms	*Science News* [19]
French	March 17, 2020	*Coronavirus : toux, fièvre, fatigue... quels sont les symptômes du Covid-19 ?*	*Le Parisien* [20]
Spanish	March 18, 2020	*El coronavirus neutraliza los sentidos del olfato y el gusto*	ABC [21]
Portuguese (Brazil)	March 18, 2020	*Virologista alemão revela novos sintomas do coronavírus*	*Sputnik News* [22]

We subsequently studied the peaks for *coronavirus*, *cough*, and *anosmia*/*ageusia*. The peak for *anosmia*/*ageusia* is delayed compared to that for *cough*, which is a major symptom of COVID-19. The peaks were usually short (1 week), confirming that most of the queries were driven by media coverage. Prior studies have also pointed out that GT data are highly influenced by media [23,24]; due to media coverage, aberrant ragweed pollen peaks were observed during the grass pollen season [25]. In fact, one important limitation of demand-based infodemiological studies is the difficulty of distinguishing the effects of a true biological epidemic from what generates interest or apprehension in internet users [2,5]. In that sense, complementing search data with click data has been suggested as a partial solution to overcome this limitation [2].

The correlation between *anosmia*/*ageusia* and deaths or new cases of COVID-19 varied substantially among countries. Depending on the country, there was a high correlation or no correlation at all. Prior to March 16, in countries with higher COVID-19 infection or death rates, there were moderate to good correlations between queries on *anosmia*/*ageusia* and new COVID-19 cases or deaths. This suggests that in the absence of substantial changes in media coverage and in the presence of a sufficiently high COVID-19 incidence, GT data mostly reflect searches for symptoms patients are experiencing. Thus, the strong correlations found by Walker et al [15] may reflect the facts that they analyzed GT data for *anosmia*/*ageusia* only up to March 25, 2020 (ie, up to the week before searches for *anosmia*/*ageusia* started to decrease); that their analyses on the associations between COVID-19 cases/deaths and premediatic coverage of *anosmia* GT data were restricted to three countries (the United Kingdom, Spain, and Italy); and that this premediatic coverage was considered by the authors to have occurred up to March 20, 2020 (ie, searches between March 16 and 19 were misclassified because they had already occurred under the potential influence of media coverage).

Limitations

Our study has some potentially relevant limitations. We used data at national levels, which may have not captured within-country heterogeneity on COVID-19 incidence or GT data; different results may have been obtained if the data were assessed at a more granulated level. Another relevant limitation concerns the fact that by March 16, 2020, the incidence of COVID-19 was still low in most Western countries; with the exception of Italy and Spain, the remaining Western countries had fewer than 100 confirmed COVID-19 cases per million inhabitants. The possibility of assessing a larger number of countries with higher numbers of COVID-19 cases would have allowed us to more confidently assess *anosmia* and *ageusia* search patterns (and their association with COVID-19 epidemiology) before and after media coverage on those symptoms.

Another important GT limitation concerns the representativeness of internet users [26]. Internet use is lowest among older persons, who constitute the age group with the highest COVID-19 morbidity. Finally, GT provides relative rather than absolute numbers, which may limit across-country comparisons. However, as expected, similar correlation coefficients were obtained when comparing GT data with relative or absolute numbers of COVID-19 cases/deaths.

Conclusions

At least in the initial stages of the SARS-CoV-2 pandemic, COVID-19–related web searches may more closely reflect media coverage (and subsequent users' interest or apprehension) than epidemiological trends. The use of Google Trends has increased dramatically in the last decade; whereas in the past, the focus had been on surveillance and monitoring, the focus of research has now shifted to forecasting changes [27]. It appears to be important to link GT with other sources of data to overcome the limitations of using search information alone.

Conflicts of Interest

None declared.

References

1. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
2. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
3. Bernardo T, Rajic A, Young I, Robiadek K, Pham M, Funk J. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013 Jul 18;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]
4. Mavragani A. Infodemiology and Infoveillancecoping Review. J Med Internet Res; 2020 ;22(4):e16206.
5. Eysenbach G. SARS and population health technology. J Med Internet Res 2003 Jun 30;5(2):e14 [FREE Full text] [doi: 10.2196/jmir.5.2.e14] [Medline: 12857670]
6. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
7. Klembczyk JJ, Jalalpour M, Levin S, Washington RE, Pines JM, Rothman RE. Google Flu Trends Spatial Variability Validated Against Emergency Department Influenza-Related Visits. J Med Internet Res 2016;18(6):e175. [Medline: 27354313]
8. Butler D. When Google got flu wrong. Nature 2013;494(7436):155-156.
9. Husnayain A, Fuad A, Su E. Applications of Google Search Trends for risk communication in infectious disease management: A case study of the COVID-19 outbreak in Taiwan. Int J Infect Dis 2020 Mar 12;95:221-223 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.021] [Medline: 32173572]
10. Mavragani A. Tracking COVID-19 in Europe: Infodemiology Approach. JMIR Public Health Surveill 2020 Apr 20;6(2):e18941 [FREE Full text] [doi: 10.2196/18941] [Medline: 32250957]

11. Ayyoubzadeh S, Ayyoubzadeh S, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: 10.2196/18828] [Medline: 32234709]
12. Hong Y, Lawrence J, Williams DJ, Mainous IA. Population-Level Interest and Telehealth Capacity of US Hospitals in Response to COVID-19: Cross-Sectional Analysis of Google Search and National Hospital Survey Data. JMIR Public Health Surveill 2020 Apr 07;6(2):e18961 [FREE Full text] [doi: 10.2196/18961] [Medline: 32250963]
13. Lechien JR, Chiesa-Estomba CM, De Siati DR, Horoi M, Le Bon SD, Rodriguez A, et al. Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. Eur Arch Otorhinolaryngol 2020 Apr 06:1-11 [FREE Full text] [doi: 10.1007/s00405-020-05965-1] [Medline: 32253535]
14. Schmitt P. Frankfurter Allgemeine Zeitung. 2020 Mar 16. „Wir haben neue Symptome entdeckt“ URL: https://www.faz.net/aktuell/gesellschaft/gesundheit/coronavirus/neue-corona-symptome-entdeckt-virologe-hendrik-streeck-zum-virus-16681450/hendrik-streeck-leitet-das-16681500.html [accessed 2020-06-12]
15. Walker A, Hopkins C, Surda P. The use of google trends to investigate the loss of smell related searches during COVID-19 outbreak. Int Forum Allergy Rhinol 2020 Apr 11:Preprint [FREE Full text] [doi: 10.1002/alr.22580] [Medline: 32279437]
16. Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
17. Cuppini L, Orlando S. Corriere de la Serra. 2020 Mar 17. Coronavirus, tra i sintomi frequenti la perdita totale di gusto e olfatto URL: https://www.corriere.it/salute/malattie_infettive/20_marzo_17/coronavirus-sintomi-frequenti-perdita-gusto-olfatto-6d3b0932-6836-11ea-9725-c592292e4a85.shtml?refresh_ce-cp [accessed 2020-06-12]
18. Mowat L. Daily Express. 2020 Mar 17. Coronavirus symptoms shock: Scientists discover NEW symptoms including lack of taste URL: https://www.express.co.uk/news/world/1256433/coronavirus-symptoms-latest-uk-covid-19-coronavirus-taste-smell [accessed 2020-06-12]
19. Saey TH. Science News. 2020 Mar 13. Coronavirus is most contagious before and during the first week of symptoms URL: https://www.sciencenews.org/article/coronavirus-most-contagious-before-during-first-week-symptoms [accessed 2020-06-12]
20. Gallet L. Le Parisien. 2020 Mar 17. Coronavirus : toux, fièvre, fatigue... quels sont les symptômes du Covid-19 ? URL: https://www.leparisien.fr/societe/coronavirus-quels-sont-les-symptomes-du-covid-19-17-03-2020-8282199.php [accessed 2020-06-12]
21. Sánchez R. ABC. 2020 Mar 18. El coronavirus neutraliza los sentidos del olfato y el gusto URL: https://www.abc.es/sociedad/abci-coronavirus-neutraliza-sentidos-olfato-y-gusto-202003180214_noticia.html [accessed 2020-03-12]
22. Sputnik News. 2020 Mar 18. Virologista alemão revela novos sintomas do coronavírus URL: https://br.sputniknews.com/ciencia_tecnologia/2020031815343670-virologista-alemao-revela-novos-sintomas-do-coronavirus/ [accessed 2020-06-12]
23. Hamaoui-Laguel L, Vautard R, Liu L, Solmon F, Viovy N, Khvorostyanov D, et al. Effects of climate change and seed dispersal on airborne ragweed pollen loads in Europe. Nature Clim Change 2015 May 25;5(8):766-771. [doi: 10.1038/nclimate2652]
24. Pandey A, Abdullah K, Drazner MH. Impact of Vice President Cheney on public interest in left ventricular assist devices and heart transplantation. Am J Cardiol 2014 May 01;113(9):1529-1531. [doi: 10.1016/j.amjcard.2014.02.007] [Medline: 24630787]
25. Bousquet J, Agache I, Berger U, Bergmann K, Besancenot J, Bousquet P, et al. Differences in Reporting the Ragweed Pollen Season Using Google Trends across 15 Countries. Int Arch Allergy Immunol 2018 May 9;176(3-4):181-188 [FREE Full text] [doi: 10.1159/000488391] [Medline: 29742519]
26. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
27. Jun S, Yoo HS, Choi S. Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. Technol Forecast Soc Change 2018 May;130:69-87. [doi: 10.1016/j.techfore.2017.11.009]

Abbreviations

COVID-19: coronavirus disease
GT: Google Trends
SARS: severe acute respiratory syndrome
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 24.04.20; peer-reviewed by H Mehdizadeh, A Mavragani; comments to author 23.05.20; revised version received 29.05.20; accepted 11.06.20; published 10.08.20.

Please cite as:
Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J
Assessment of the Impact of Media Coverage on COVID-19–Related Google Trends Data: Infodemiology Study
J Med Internet Res 2020;22(8):e19611
URL: https://www.jmir.org/2020/8/e19611
doi:10.2196/19611
PMID:32530816

Original Paper

Global Research on Coronaviruses: An R Package

Thierry Warin[1], PhD

HEC Montreal, Montreal, QC, Canada

Corresponding Author:
Thierry Warin, PhD
HEC Montreal
3000, chemin de la Côte-Sainte-Catherine
Montreal, QC, H3T 2A7
Canada
Phone: 1 514 340 6185
Email: thierry.warin@hec.ca

Abstract

Background: In these trying times, we developed an R package about bibliographic references on coronaviruses. Working with reproducible research principles based on open science, disseminating scientific information, providing easy access to scientific production on this particular issue, and offering a rapid integration in researchers' workflows may help save time in this race against the virus, notably in terms of public health.

Objective: The goal is to simplify the workflow of interested researchers, with multidisciplinary research in mind. With more than 60,500 medical bibliographic references at the time of publication, this package is among the largest about coronaviruses.

Methods: This package could be of interest to epidemiologists, researchers in scientometrics, biostatisticians, as well as data scientists broadly defined. This package collects references from PubMed and organizes the data in a data frame. We then built functions to sort through this collection of references. Researchers can also integrate the data into their pipeline and implement them in R within their code libraries.

Results: We provide a short use case in this paper based on a bibliometric analysis of the references made available by this package. Classification techniques can also be used to go through the large volume of references and allow researchers to save time on this part of their research. Network analysis can be used to filter the data set. Text mining techniques can also help researchers calculate similarity indices and help them focus on the parts of the literature that are relevant for their research.

Conclusions: This package aims at accelerating research on coronaviruses. Epidemiologists can integrate this package into their workflow. It is also possible to add a machine learning layer on top of this package to model the latest advances in research about coronaviruses, as we update this package daily. It is also the only one of this size, to the best of our knowledge, to be built in the R language.

(J Med Internet Res 2020;22(8):e19615) doi:10.2196/19615

KEYWORDS

COVID-19; SARS-CoV-2; coronavirus; R package; bibliometric; virus; infectious disease; reference; informatics

Introduction

The coronavirus disease (COVID-19) outbreak finds its roots in Wuhan with potential first observations in late 2019 [1]. On December 30, 2019, clusters of cases of pneumonia of unknown origin were reported to the China National Health Commission. On January 7, 2020, a novel coronavirus (2019-nCoV) was isolated. Two previous outbreaks have taken place since the year 2000 involving coronaviruses: (1) severe acute respiratory syndrome–related coronavirus (SARS-CoV) and (2) Middle East respiratory syndrome–related coronavirus (MERS-CoV) [2].

In the context of the global propagation of the virus, numerous initiatives have occurred mobilizing our current global technological infrastructure: (1) the internet and server farms; (2) the convergence in coding languages; and (3) the use of data, structured and unstructured.

First, the internet is used for exchanges between universities, research laboratories, or political leaders to cite a few examples. Second, the convergence in coding languages, in particular functional languages such as R (R Foundation for Statistical Computing), Python (Python Software Foundation), or Julia, has helped facilitate the communications between researchers. Reproducible research has accelerated in these past few years,

with principles such as open science, open data, and open code. The use of data has been amplified by the development of new methodologies within the field of artificial intelligence (AI), allowing researchers to generate and analyze structured and unstructured data in a supervised way, as well as in a semi- or unsupervised way [3]. Data initiatives have flourished across the world, collecting firsthand data, aggregating various data sets, and developing simulation-based models. The Johns Hopkins University of Medicine has made a great effort in terms of data visualizations, which has disseminated throughout the world [4]. In doing so, it has undoubtedly helped to raise citizens' awareness, helped policy makers inform their population and avoid some potential fake news, or helped correct some misconceptions or confusion. The Johns Hopkins initiative has been followed by several others across the world, creating a large variety and diversity of data sets. This creation process allows researchers to benefit from different levels of granularity when it comes to the data dimensions such as geography, indicators, or methodologies; the open science characteristic is also an interesting aspect of the current research contributions [1].

With better access to these new technologies and methodologies, the breadth of expertise is more extensive; epidemiologists are leading the core of the research, but data scientists, biostatisticians, researchers in humanities, or social scientists can contribute and leverage their domain expertise using converging methodologies such as decision trees, text mining, or network analysis [5].

It is with these hypotheses in mind that we propose to replicate the spirit of the Allen Institute for AI initiative and to design an R package, whose main objective is to integrate easily in a researcher's workflow. This package is named EpiBibR.

EpiBibR stands for an "epidemiology-based bibliography for R." The R package is under the Massachusetts Institute of Technology license and, as such, is a free resource based on the open science principles (reproducible research, open data, and open code). The resource may be used by researchers whose domain is scientometrics but also by researchers from other disciplines. For instance, the scientific community in AI and data science may use this package to accelerate new research insights about COVID-19. The package follows the methodology put in place by the Allen Institute and its partners [6] to create the CORD-19 data set with some differences. The latter is accessible through downloads of subsets or a representational state transfer application programming interface. The data provide essential information such as authors, methods, data, and citations to make it easier for researchers to find relevant contributions to their research questions. Our package proposes 22 features for the 60,500 references (on June 26, 2020), and access to the data has been made as easy as possible to integrate efficiently in almost any researcher's pipeline.

Through this package, a researcher can connect the data to her research protocol based on the R language. With this workflow in mind, a researcher can save time on collecting data and can use an accessible language to perform complex analytical tasks, for instance, be it in R or Python. Indeed, it is usual that researchers use multiple languages (functional or not) to produce specific outputs. This workflow opens these data to analyses from the most extensive spectrum of potential options, enhancing multidisciplinary approaches applied to these data (biostatistics, bibliometrics, and text mining, among others).

The goal of this package in this emergency context is to limit the references to the medical domain (here the PubMed repository) but to then leverage the methodologies used across different disciplines. As we will address this point later, a further extension could be to add references from other disciplines to not only benefit from the wealth of methodologies but also their theories and concepts. For instance, to assess the spread of the disease, the literature—and theories—from researchers in demography would certainly be relevant.

Methods

Motivation

Across the world, a couple of initiatives have emerged whose goals are primarily to provide access to medical references. The main objective is to disseminate, as much as possible, the extensive research that has been done in the past (and recent history) to save some time and to improve the efficiency of further investigation. Research processes need to be efficient, and the time spent to perform this research needs to be relevant in this emergency. In addition, by proposing a (as much as possible) comprehensive data set of medical references on the coronavirus topic, the wisdom of crowds principle may play a role. A broader community beyond university researchers may use it and help shorten the time to the vaccine production. Researchers from pharmaceutical companies or other organizations may tap into these data to fine-tune their research and research processes.

A nonexhaustive list of the current bibliographic packages comprises the LitCovid data set from the National Library of Medicine (NLM) [7], the World Health Organization (WHO) data set [8], the "COVID-19 Research Articles Downloadable Database" from the Centers for Disease Control and Prevention (CDC) - Stephen B. Thacker CDC Library [9], and the "COVID-19 Open Research Dataset" by the Allen Institute for AI and their partners [6]. All these resources are essential and serve various complementary purposes. They are disseminated to their respective channels (ie, to their respective audiences). They are tailored to their specific needs. The LitCovid data set comprises 6530 references and can be downloaded from the United States NLM's website in a format that suits bibliographic software. It deals primarily with research about 2019-nCoV. The WHO's data set has around 9663 references, also specifically on COVID-19. The CDC's database is proposed in a software format as well (Microsoft Excel [Microsoft Corp] and bibliographic software formats) and comprises 17,636 references about COVID-19 and the other coronaviruses. The Allen Institute for AI's data set proposes over 190,000 references about COVID-19 as well as references about the other coronaviruses. It is accessible through different subsets of the overall database and a dedicated search engine. It also taps into a variety of academic article repositories.

In this context, the contributions made by the EpiBibR package are fourfold. First, with more than 60,500 references, EpiBibR is among the most extensive reference databases and is updated daily. The sheer number of references may be more suitable for a broader audience. Second, EpiBibR collects the data exclusively from PubMed to propose a controlled environment. Third, EpiBibR matches the keywords from the Allen Institute for AI's database to offer some consistency for researchers. Last, it is an R package and, as such, can be integrated into a workflow a little more efficiently than a file necessitating a specific software. Research teams can install the package in their systems and tap into it without the risks of version issues.

Beyond these four differentiation elements, EpiBibR is not better or worse than any other existing database. It just serves its audience and its purpose, like the other databases. It has not been created to replace a current database but, to the contrary, to complement these databases. We do believe that we need more initiatives in this domain at the world stage to support and integrate all the potential audiences and various workflows across the world. As a result, these initiatives would help accelerate research on coronaviruses overall and COVID-19 in particular.

Functionality

As previously mentioned, EpiBibR is an R package to access bibliographic information on COVID-19 and other coronaviruses references easily. The package can be found at [10]. The command to install it is remotes::install_github ('warint/'EpiBibR'). We advise making sure the latest version of the package has been installed on each researcher's system. The installation procedure can be found on the README file of this Github account. A full website with the various functions and examples are accessible from this page as well. The vignette has been created based on this paper to extend the use cases as more data are collected.

The references were collected via PubMed, a free resource that is developed and maintained by the National Center for Biotechnology Information at the United States NLM, located at the National Institutes of Health. PubMed includes over 30 million citations from biomedical literature.

More specifically, to collect our references, we adopted the procedure used by the Allen Institute for AI for their CORD-19 project. We applied a similar query on PubMed ("COVID-19" OR Coronavirus OR "Corona virus" OR "2019-nCoV" OR "SARS-CoV" OR "MERS-CoV" OR "Severe Acute Respiratory Syndrome" OR "Middle East Respiratory Syndrome") to build our bibliographic data.

To navigate through our data set, EpiBibR relies on a set of search arguments: author, author's country of origin, keyword in the title, keyword in the abstract, year, and the name of the journal. Each of them can genuinely help scientists and R users to filter references and find the relevant articles.

To simplify the workflow between our package and the research methodologies, the format of our data frame has been designed to integrate with different data pipelines, notably to facilitate the use of the R package Bibliometrix with our data [11].

The package comprises more than 60,500 references and 22 features (see Textbox 1).

Textbox 1. Field tags and their descriptions.

AU

Authors

TI

Document title

AB

Abstract

PY

Year

DT

Document type

MESH

Medical Subject Headings vocabulary

TC

Times cited

SO

Publication name (or source)

J9

Source abbreviation

JI

International Organization for Standardization source abbreviation

DI

Digital Object Identifier

ISSN

Source code

VOL

Volume

ISSUE

Issue number

LT

Language

C1

Author address

RP

Reprint address

ID

PubMed ID

DE

‘Authors’ Keywords

UT

Unique Article Identifier

AU_CO

‘Author’s country of origin

DB

Bibliographic database

EpiBibR allows researchers to search academic references using several arguments: author, author's country of origin, author + year, keywords in the title, keywords in the abstract, year, and source name. Researchers can also download the entire bibliographic data frame comprised of around 60,500 references with 22 metadata each.

In Textbox 2, we provide the descriptions of the functions available in the R language to collect the relevant information.

Textbox 2. Descriptions of the functions to collect the relevant references.

EpiBib_data <- EpiBib_references()

Download the entire bibliographic data frame

colson_articles <- EpiBib_author("Colson")

Search all the articles written by Philippe Colson

canada_articles <- EpiBib_country("canada")

Search by 'author's country of origin

yang2019 <- EpiBib_AU_YE(author = "yang", year = 2019)

Search by author and year

covid_articles <- EpiBib_title("covid")

Search by keywords in the title

coronavirus_articles <- EpiBib_abstract("coronavirus")

Search by keywords in the abstract

A2020_articles <- EpiBib_year(2020)

Search by year

bio_articles <- EpiBib_source("bio")

Search by source

Results

In what follows, a use case about how we can use such a data set is proposed. This section is not intended to offer a complete systematic literature review of the 60,500 references. This is the purpose of another article. However, we want to illustrate some powerful techniques that can be applied to this collection of references (for instance, social network analysis) while remaining at a very high level [12].

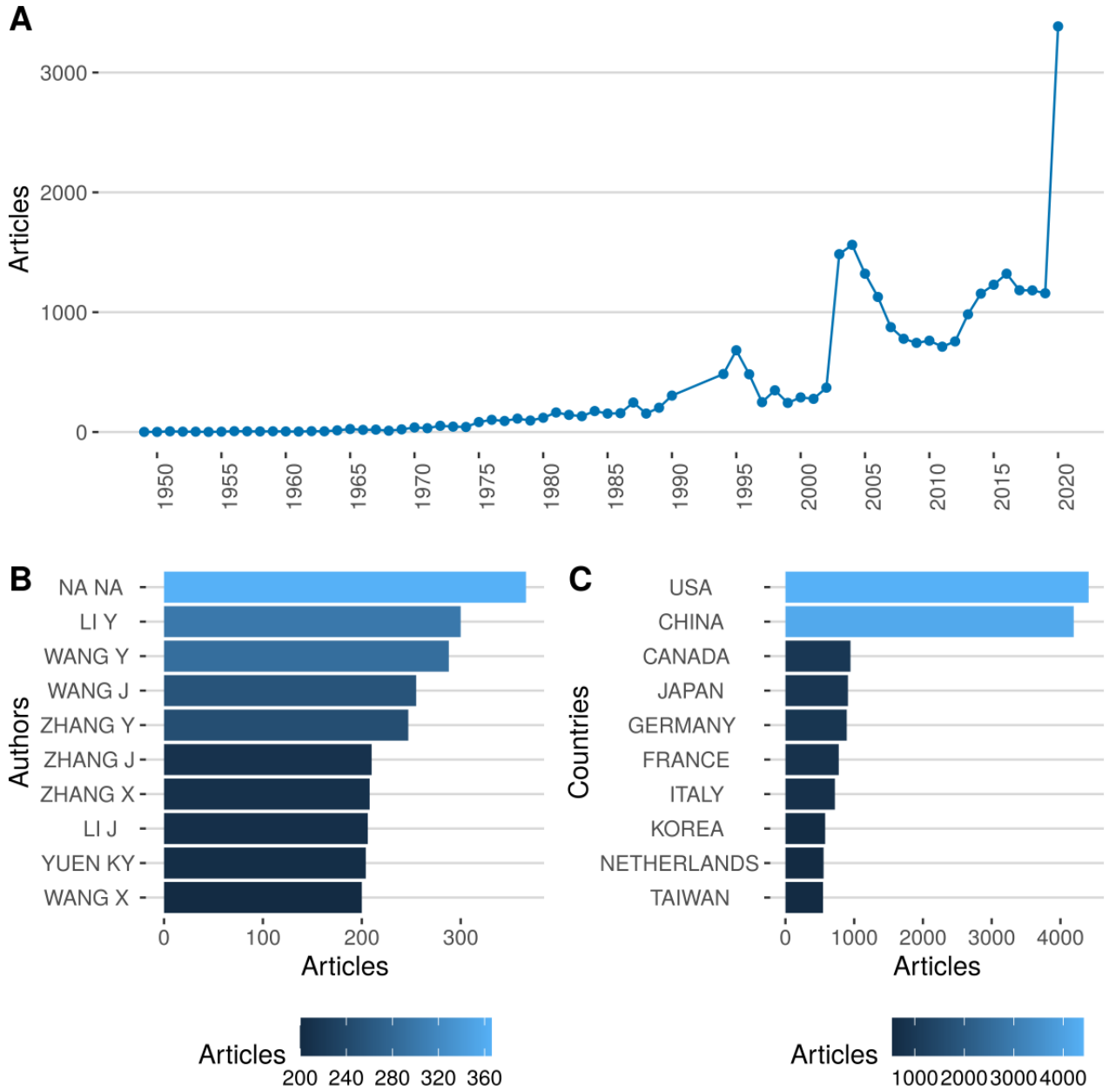
A systematic literature review consists of mainly four stages: (1) planning, (2) conducting, (3) analysis, and (4) synthesis and reporting. In the first stage, a preliminary study aims to build a corpus of articles citing the most relevant articles in the domain. The second stage is about producing a general review of the main topics used in the corpus. The third stage here is about making a cocitation analysis of the references in each corpus article. The last stage is about proposing a keywords co-occurrence analysis [13].

In our context, we propose a slightly modified perspective on a systematic literature review. The first stage is here at a different scale since we collect not just a few relevant articles to create the corpus but a vast, almost exhaustive, list of articles on a topic. It is worth noticing that the process of data selection consists more of what we define as an "algorithmic systematic literature review," sometimes also referred to as "automation" [14]. An algorithmic systematic literature review comes with lots of benefits. The proposed modified systematic literature review improves the more classical approach since it does not rely on a manual search and extraction, with the potential biases and limitations it might create. An algorithmic systematic literature review combines the strength of both approaches: the power of big data with the academic soundness of the systematic literature review process. As such, it does not replace the expert's analysis of the literature. On the contrary, it should be used to augment the expert's study [3].

The *bibliometrix* package allows a thorough bibliometric analysis using R. Our EpiBibR data have been designed to integrate easily with the *bibliometrix* package. A shiny application is also available, called biblioshiny() [11]. This package has been used extensively for various exercises mobilizing massive amounts of data [15].

Let us first propose a simple count of the references on the coronaviruses literature. In Figure 1, the historical development of research on coronaviruses can be analyzed as having three stages: exploration, initial development, and rapid development in the past year and the current year.

Figure 1. (A) Count of articles, (B) most productive authors, and (C) most productive countries.



In 2019, a little more than 1000 papers on coronaviruses had been produced, and in the first 5 months of 2020, close to 3400 papers were written on the topic. Those papers from the past 2 years seem to be an interesting, statistically representative sample. In Figure 1, we can also highlight the most productive authors as well as the most productive countries in terms of absolute counts. The most prolific authors provide interesting statistics since it is most likely to proxy research labs. In doing so, we can find which teams are working on which aspect of the coronaviruses. To illustrate our latter point, in Figure 2, we first propose a compelling visualization, called a Sankey diagram. It links authors, keywords, and sources on a connected map. It is the first way to create groups of researchers. The results could be used by policy makers to identify areas of research on this topic.

We can also use powerful techniques such as Social Network Theory to find potential clusters of topics, clusters of researchers, and groups of country collaborations. Figure 3 is an example of the latter. The United States and China produce the bulk of the research on coronaviruses.

Figure 2. Sankey diagram: authors <> keywords <> sources.

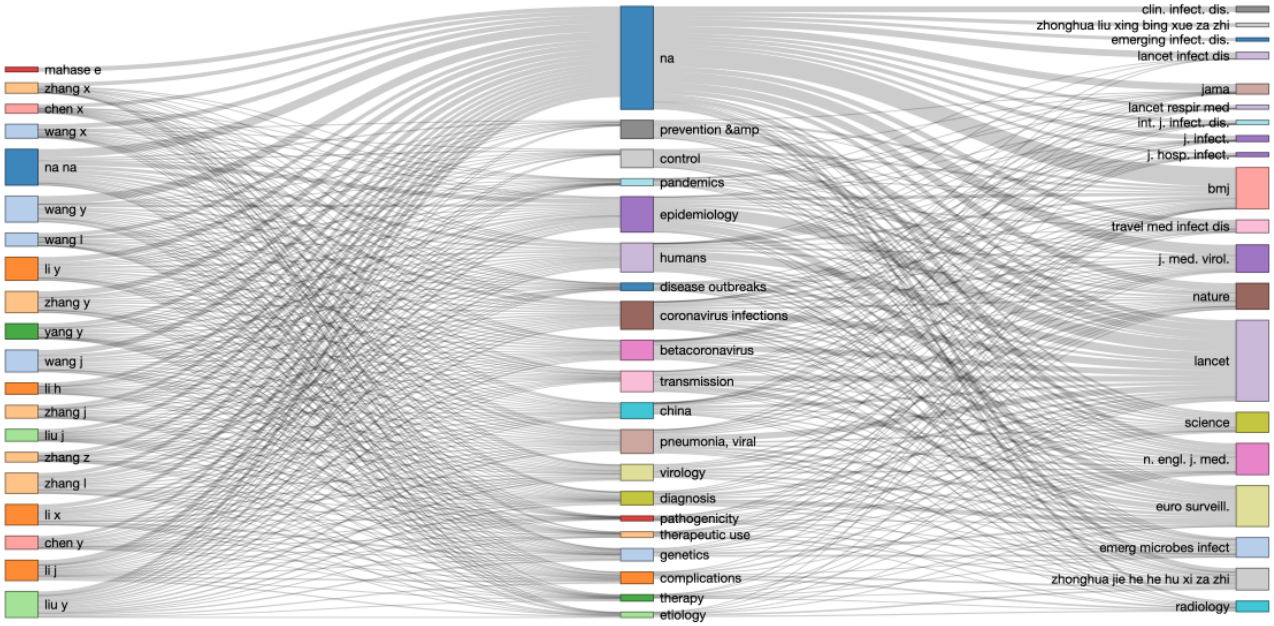


Figure 3. Country collaboration network.

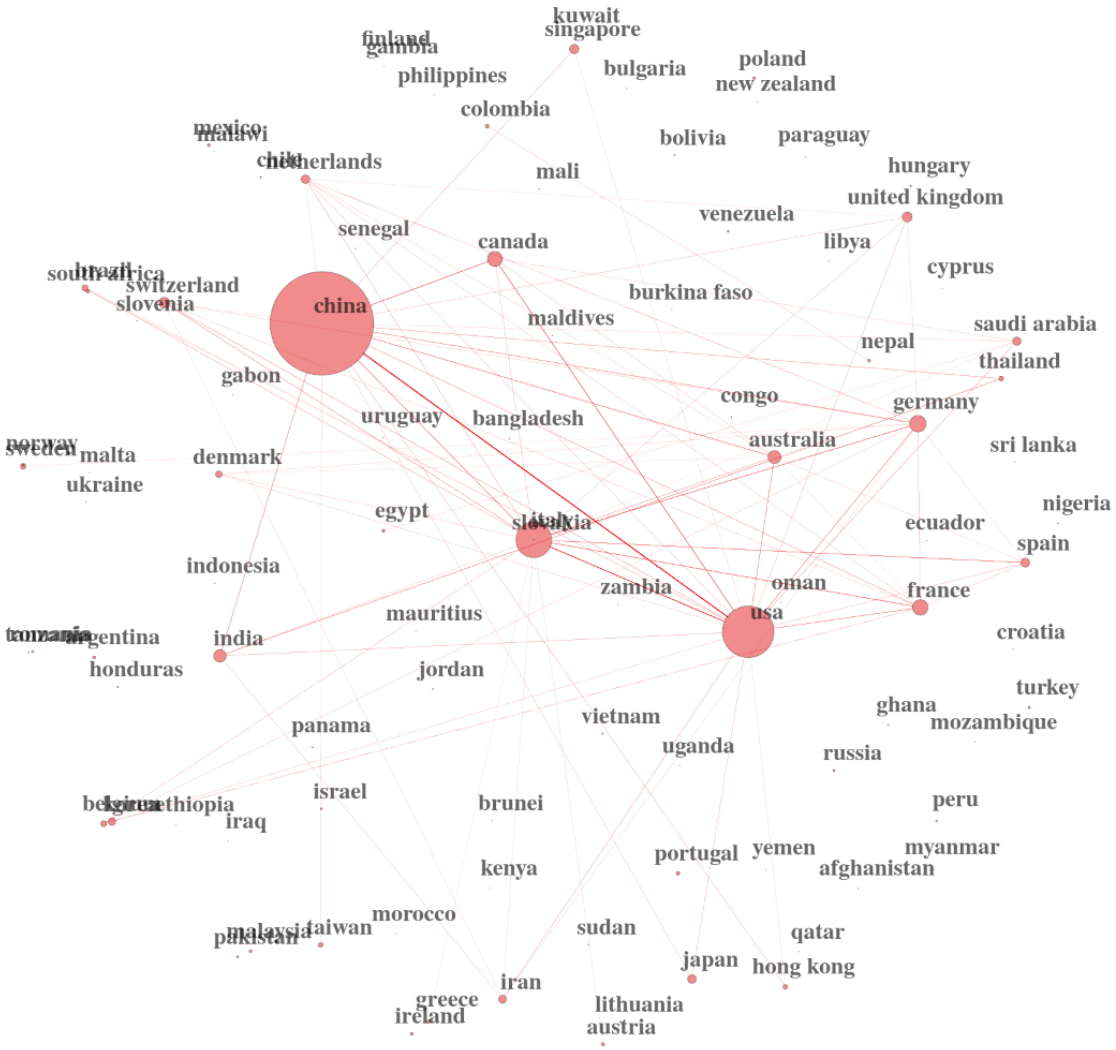


Figure 4 is an illustration of author collaboration networks. As previously mentioned, we remain at a very high level here. However, policy makers or public health officers, for instance, could use these techniques to find more granular networks either

just within the 60,500 references or by crossing with other databases. We could even imagine crossing with unstructured data for some specific purposes [16].

Figure 5 is about finding clusters of topics. This technique can be applied to subsamples of the 60,500 references to provide a more granular analysis.

Figure 6 proposes indeed to go further on the topic dimension. For instance, we can study the evolution over time of the author's keywords usages.

Figure 4. Author collaboration network.

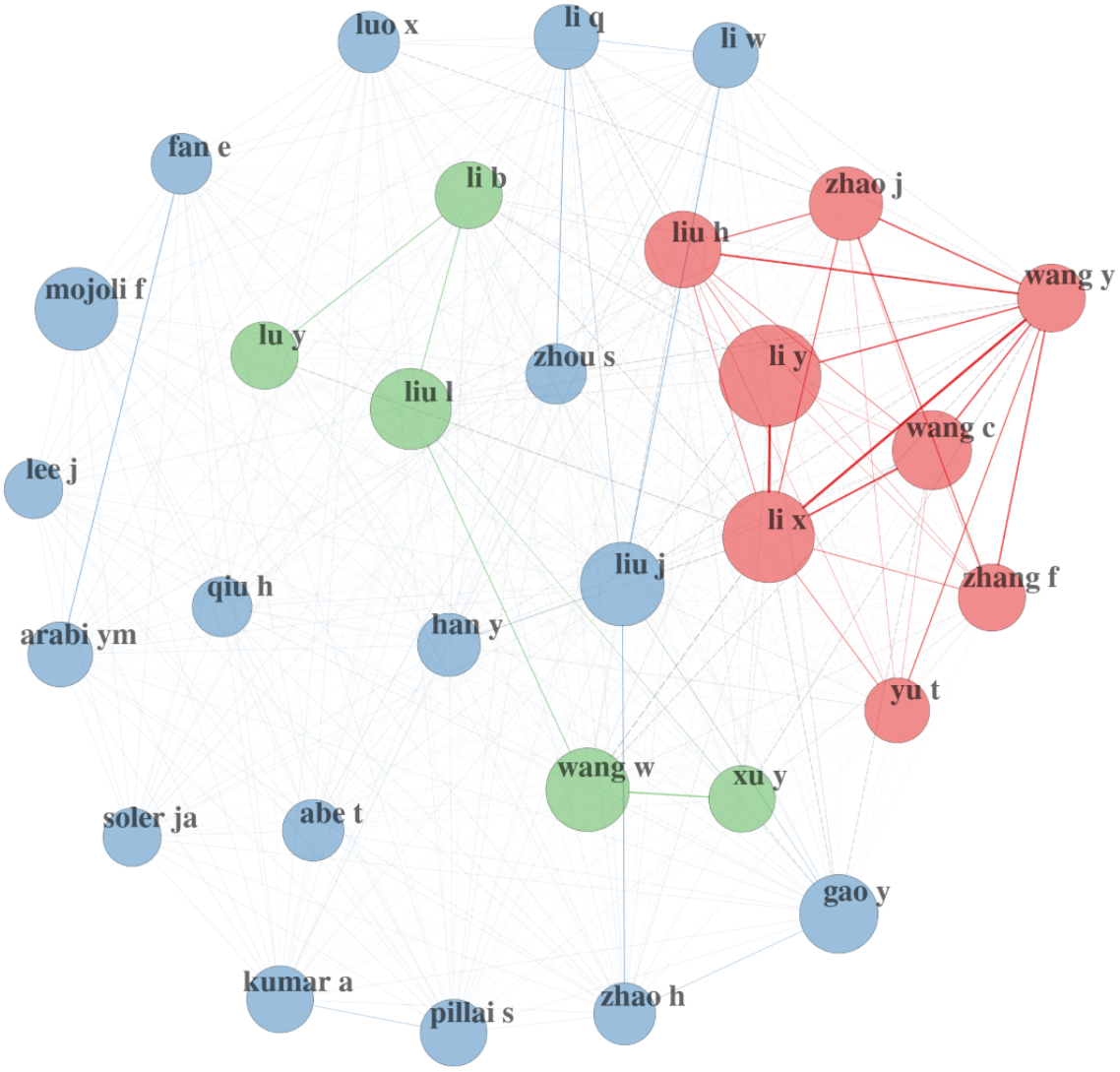


Figure 5. Thematic maps.

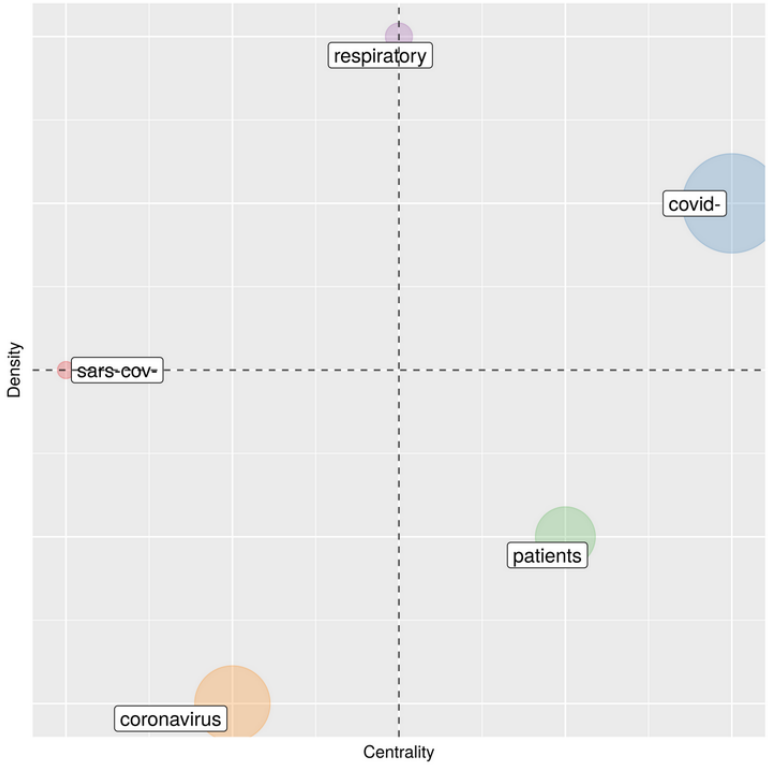
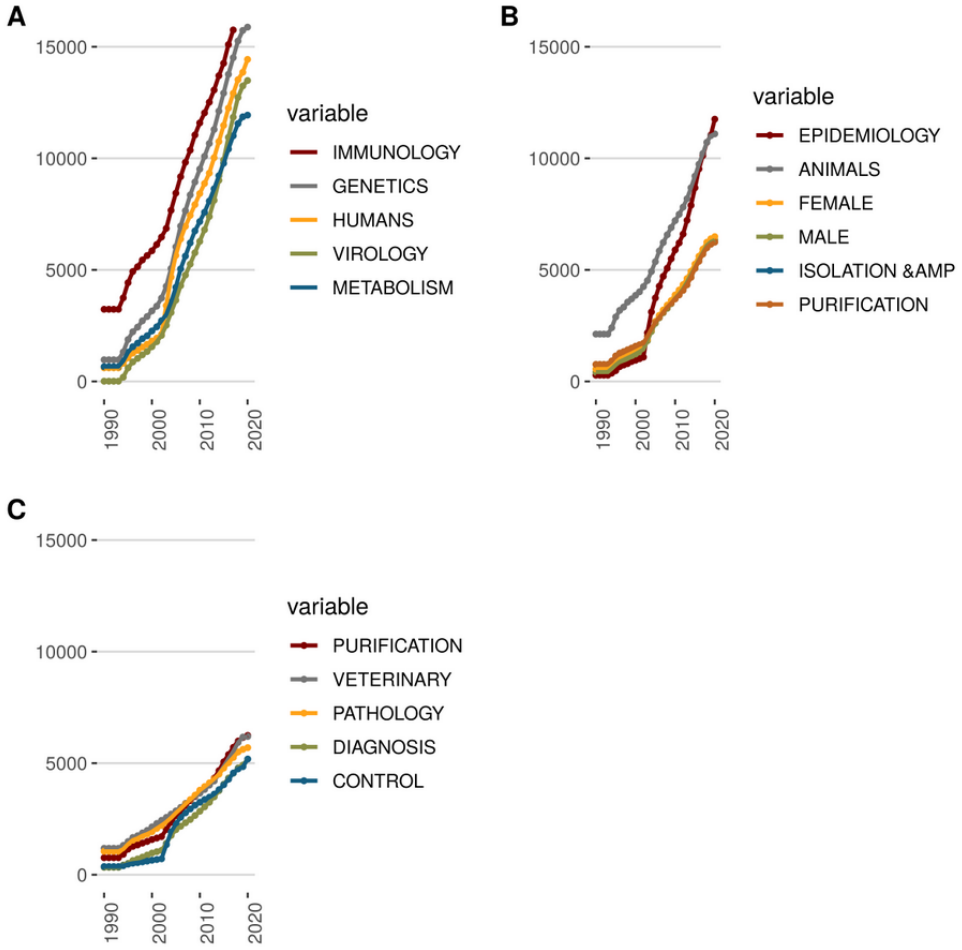


Figure 6. Author's keywords usage evolution over time.



As previously mentioned, this section is not intended to cover a systematic literature review but rather to illustrate some of the potential uses of powerful techniques or highlight some of the interesting questions that can be addressed. With a data set of 60,500 references, we have access to a wealth of information. Combined with the state of the technological development we have access to today, multiple questions can be answered. To go further, we could imagine analyzing document cocitations or reference burst detections [17].

Discussion

Further Developments

Recent developments in computing power, as well as data accessibility, offer new tools to develop policies to promote new capabilities or enhance existing skills as a way to encourage the further coevolution of new capabilities, echoing ideas put forward by Hirschman [18] more than 50 years ago. The difference is that now researchers, policy makers, and business analysts can analyze them in practice. It is capitalizing on the principles from the "wisdom of crowds" [19-22].

In this global pandemic, knowledge sharing and open data can have an impact on the solutions as well as the pace to discover the answers. With such a package, the easy dissemination through such an integrated workflow and low-level pipeline of tools may also help public policies. It allows the use of research evidence in health policy making [23].

Developing easy access to data and data modelization is of great importance for evidence-based policy making. In the past, there were lots of areas in public policy making where data were not accessible. As a result, decisions were made on assumptions coming from theoretical foundations or benchmarks from other sources. In our day and age, with more and more access to data across the world, being open data initiatives or not, evidence-based decisions are more and more possible. Numerous authors have demonstrated the role of data in informing better evidence-based policies [23-26].

This R package is updated daily when it comes to collecting the references and their metadata, and it will also be updated regularly to propose different use cases and new functionalities. We will update the modeling contribution of the package. For instance, we will integrate some of the *bibliometrix* package's functions directly in our package to ease the scientometrist's workflow. We will also include some models for network analysis and natural language processing–based studies.

Conclusion

In these trying times, we believe that working with reproducible research principles based on open science, disseminating scientific information, providing easy access to scientific production on this particular issue, and offering a rapid integration in researchers' workflows may help save time in this race against the virus, notably in terms of public health. In this context, we believe the bibliometric packages made available by research institutions, nongovernmental organizations, or individual researchers complement the other data packages and help provide a more comprehensive understanding of the pandemics. One of the objectives is to reduce "the barrier for researchers and public health officials in obtaining comprehensive, up-to-date data on this ongoing outbreak. With this package, epidemiologists and other scientists can directly access data from four sources, facilitating mathematical modelling and forecasting of the COVID-19 outbreak" [1].

This package aims at providing this easy access and integration in a researcher's workflow. It is specially designed to collect data and generate a data frame compatible with the *bibliometrix* package [11]. Such data sets may facilitate access to the right information. Moreover, the use of massive data sets crossed with robust data analyses may foster multidisciplinary perspectives, raising new questions and providing new answers [27-29]. Classification techniques can be used to go through the large volume of references and allow researchers to save time on this part of their research. Network analysis can be used to filter the data set. Text mining techniques can also help researchers calculate similarity indices and help them focus on the parts of the literature that are relevant for their research.

The package collects references that are interesting, for the most part, for the medical domain and allows multidisciplinary perspectives on this data set. It could be interesting to get views from other disciplines, for instance, mathematics, computer science, political science, economics, and environmental science. This is the result of the emergency in which humanity finds itself right now. We could also envisage later to add references from other disciplines such as social sciences and augment, or open, the perspectives on the issue. Not only would we benefit from a multidisciplinary perspective through the methodology dimension, as the goal is with our EpiBibR package, but we would also benefit from the multidisciplinary perspectives through the ontological concepts and theories of these added domains.

Acknowledgments

The author would like to thank Marine Leroi and Martin Paquette for their help in collecting the data and the maintenance of the package. The usual caveats apply.

Conflicts of Interest

None declared.

References

1. Wu T, Ge X, Yu G, Hu E. Open-source analytics tools for studying the COVID-19 coronavirus outbreak. medRxiv 2002:A. [doi: 10.1101/2020.02.25.20027433]
2. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb;395(10223):470-473. [doi: 10.1016/s0140-6736(20)30185-9]
3. Warin T, Duc R, Sanger W. Mapping innovations in artificial intelligence through patents: a social data science perspective. 2017 Dec Presented at: International Conference on Computational Science and Computational Intelligence (CSCI). . pp. 252?7; 2017; Las Vegas, NV p. 252-257. [doi: 10.1109/csci.2017.40]
4. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-04-21]
5. Warin T, Sanger W. Connectivity and closeness among international financial institutions: a network theory perspective. IJCM 2018;1(3):225. [doi: 10.1504/ijcm.2018.094479]
6. CORD-19. Semantic Scholar. URL: https://pages.semanticscholar.org/coronavirus-research [accessed 2020-04-21]
7. Chen Q, Allot A, Lu Z. Keep up with the latest coronavirus research. Nature 2020 Mar;579(7798):193. [doi: 10.1038/d41586-020-00694-1] [Medline: 32157233]
8. Global research on coronavirus disease (COVID-19). World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov [accessed 2020-04-24]
9. COVID-19 research articles downloadable database. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/library/researchguides/2019novelcoronavirus/researcharticles.html [accessed 2020-04-24]
10. EpiBibR. GitHub. URL: https://github.com/warint/EpiBibR
11. Aria M, Cuccurullo C. bibliometrix: an R-tool for comprehensive science mapping analysis. J Informetrics 2017 Nov;11(4):959-975 [FREE Full text] [doi: 10.1016/j.joi.2017.08.007]
12. Camacho D, Panizo-LLedot Á, Bello-Orgaz G, Gonzalez-Pardo A, Cambria E. The four dimensions of social network analysis: an overview of research methods, applications, and software tools. arXiv 2020:A [FREE Full text] [doi: 10.1016/j.inffus.2020.05.009]
13. Tani M, Papaluca O, Sasso P. The system thinking perspective in the open-innovation research: a systematic review. J Open Innovation Technol Market Complexity 2018 Aug 18;4(3):38. [doi: 10.3390/joitmc4030038]
14. Pulsiri N, Vatananan-Thesenvitz R. Improving systematic literature review with automation and bibliometrics. 2018 Presented at: 2018 Portland International Conference on Management of Engineering and Technology (PICMET); August 2018; Honolulu, HI. [doi: 10.23919/PICMET.2018.8481746]
15. Zhao L, Tang Z, Zou X. Mapping the knowledge domain of smart-city research: a bibliometric and scientometric analysis. Sustainability 2019 Nov 25;11(23):6648. [doi: 10.3390/su11236648]
16. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
17. Kleinberg J. Bursty and hierarchical structure in streams. 2002 Presented at: KDD02: The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; July 2002; Edmonton, Alberta. [doi: 10.1145/775047.775061]
18. Hirschman AO. The Strategy of Economic Development. New Haven, CT: Yale University Press; 1958.
19. Gal MS. The power of the crowd in the sharing economy. Law Ethics Hum Rights 2019;13:59. [doi: 10.1515/lehr-2019-0002]
20. Cai CW, Gippel J, Zhu Y, Singh AK. The power of crowds: grand challenges in the Asia-Pacific region. Aust J Manage 2019 Sep 18;44(4):551-570. [doi: 10.1177/0312896219871979]
21. Wang PDC, Soares VS, de Souza JM, Esteves MGP, Esteves NCL, Duarte FR. A Crowd Science framework to support the construction of a Gold Standard Corpus for Plagiarism Detection. 2019 Presented at: 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD); May 2019; Porto, Portugal. [doi: 10.1109/cscwd.2019.8791853]
22. Avasilcai S, Galateanu E. Co - creators in innovation ecosystems. Part II: crowdsprings ‘crowd in action. IOP Conf Ser: Mater Sci Eng 2018 Sep 18;400:062001. [doi: 10.1088/1757-899x/400/6/062001]
23. Payán DD, Lewis LB. Use of research evidence in state health policymaking: menu labeling policy in California. Prev Med Rep 2019 Dec;16:101004 [FREE Full text] [doi: 10.1016/j.pmedr.2019.101004] [Medline: 31709136]
24. Wolffe TA, Whaley P, Halsall C, Rooney AA, Walker VR. Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. Environ Int 2019 Sep;130:104871 [FREE Full text] [doi: 10.1016/j.envint.2019.05.065] [Medline: 31254867]
25. Giménez-Bertomeu VM, Domenech-López Y, Mateo-Pérez MA, de-Alfonseti-Hartmann N. Empirical evidence for professional practice and public policies: an exploratory study on social exclusion in users of primary care social services in Spain. Int J Environ Res Public Health 2019 Nov 20;16(23):4600 [FREE Full text] [doi: 10.3390/ijerph16234600] [Medline: 31756963]
26. Villumsen S, Faxvaag A, Nøhr C. Development and progression in Danish eHealth policies: towards evidence-based policy making. Stud Health Technol Inform 2019 Aug 21;264:1075-1079. [doi: 10.3233/SHTI190390] [Medline: 31438090]
27. Warin T, Leiter DB. An empirical study of price dispersion in homogenous goods markets. Middlebury College, Department of Economics 2007:A [FREE Full text]

28. Warin T, Leiter D. Homogenous goods markets: an empirical study of price dispersion on the internet. IJEBR 2012;4(5):514. [doi: 10.1504/ijebr.2012.048776]
29. De Marcellis-Warin N, Sanger W, Warin T. A network analysis of financial conversations on Twitter. IJWBC 2017;13(3):1. [doi: 10.1504/ijwbc.2017.10004118]

Abbreviations

AI: artificial intelligence
CDC: Centers for Disease Control and Prevention
COVID-19: coronavirus disease
MERS-CoV: Middle East respiratory syndrome–related coronavirus
NLM: National Library of Medicine
SARS-CoV: severe acute respiratory syndrome–related coronavirus
WHO: World Health Organization
2019-nCoV: novel coronavirus

Edited by G Eysenbach; submitted 24.04.20; peer-reviewed by CM Dutra, J Yang, S Abdulrahman; comments to author 12.06.20; revised version received 26.06.20; accepted 26.07.20; published 11.08.20.

Please cite as:
Warin T
Global Research on Coronaviruses: An R Package
J Med Internet Res 2020;22(8):e19615
URL: http://www.jmir.org/2020/8/e19615/
doi:10.2196/19615
PMID:32730218

Original Paper

Citizens' Adherence to COVID-19 Mitigation Recommendations by the Government: A 3-Country Comparative Evaluation Using Web-Based Cross-Sectional Survey Data

Abrar Al-Hasan[1], BSc, MBA, PhD; Dobin Yim[2], BS, MS, MBA, MA, PhD; Jiban Khuntia[3], ME, PhD

[1]College of Business Administration, Kuwait University, AlShadadiya University City, Kuwait

[2]Loyola University Maryland, Baltimore, MD, United States

[3]University of Colorado Denver, Denver, CO, United States

Corresponding Author:
Abrar Al-Hasan, BSc, MBA, PhD
College of Business Administration
Kuwait University
Information Systems and Operation Management Department
AlShadadiya University City, 13055
Kuwait
Phone: 965 51165005
Email: abrar.alhasan@ku.edu.kw

Abstract

Background: Social distancing is an effective preventative policy for the coronavirus disease (COVID-19) that is enforced by governments worldwide. However, significant variations are observed in following the policy across individuals and countries. Arguably, differences in citizens' adherence actions will be influenced by their perceptions about government's plans and the information available to guide their behaviors—more so in the digital age in the realm of mass influence of social media on citizens. Insights into the underlying factors and dynamics involved with citizens' adherence process will inform the policy makers to follow appropriate communication and messaging approaches to influence citizens' willingness to adhere to the recommendations.

Objective: The aim of this study is a comparative evaluation of citizens' adherence process to COVID-19–relevant recommendations by the government. The focus is on how three different countries' (United States, Kuwait, and South Korea) citizens, randomly sampled, respond to governments' pandemic guidance efforts. We draw insights into two categories of perceived government roles in managing the pandemic: (1) citizens' perceptions of government's role in responding to the pandemic and (2) citizens' perceptions of government's business reopening efforts. Undoubtedly, the internet and social media have burgeoned, with differing effects on shaping individuals' views and assessments of the COVID-19 situation; we argue and test for the effects of information sources, social media use, and knowledge on the adherence actions.

Methods: We randomly sampled web-based survey data collected by a global firm in May 2020 from citizens of the United States, Kuwait, and South Korea. A nonlinear ordered probit regression, controlling for several counterfactuals, was used for analysis. The focal estimated effects of the study were compared across countries using the weighted distance between the parameter estimates.

Results: The total sample size was 482 respondents, of which 207 (43%) lived in the United States, 181 (38%) lived in Kuwait, and 94 (20%) lived in South Korea. The ordered probit estimation results suggest that overall, perception of government response efforts positively influenced self-adherence (P<.001) and others' adherence (P<.001) to social distancing and sheltering. Perception of government business reopening efforts positively influenced others' adherence (P<.001). A higher intensity of general health information source for COVID-19 had a positive effect on self-adherence (P=.003). A higher intensity of social media source use for COVID-19 positively influenced others' adherence (P=.002). A higher intensity of knowledge on COVID-19 positively influenced self-adherence (P=.008) and negatively influenced others' adherence (P<.001). There were country-level variations—broadly, the United States and Kuwait had better effects than South Korea.

Conclusions: As the COVID-19 global pandemic continues to grow and governmental restrictions are ongoing, it is critical to understand people's frustration to reduce panic and promote social distancing to facilitate the control of the pandemic. This study finds that the government plays a central role in terms of adherence to restrictions. Governments need to enhance their efforts on

publicizing information on the pandemic, as well as employ strategies for improved communication management to citizens through social media as well as mainstream information sources.

(J Med Internet Res 2020;22(8):e20634) doi:10.2196/20634

KEYWORDS

COVID-19; adherence; social distancing; government perception; information sources; social media; knowledge

Introduction

The coronavirus disease (COVID-19) has spread worldwide as an epidemic with serious implications [1]. Since its beginnings in Wuhan, China in December 2019, the COVID-19 pandemic has led to more than 5 million cases and 328,000 deaths reported across the world as of May 2020 [2]. Although around 2 million people have recovered from the virus, the prevalence is still high.

There is no known cure for COVID-19 as of now. Although vaccines are being tested, no established ones are widely available yet to prevent COVID-19 [3]. Infected individuals do not often exhibit any symptoms. The disease often progresses swiftly and kills patients at a much higher rate than the typical flu [4]. A few treatment options are possible for a patient with COVID-19. One of them is a ventilator that assists breathing. Because of limited testing and treatment availability, individuals must adopt preventive measures so that they do not get the virus [5]. Simple measures involve practicing good respiratory hygiene by washing hands with soap and water for at least 20 seconds at frequent intervals and avoiding touching the eyes, nose, or mouth with unwashed hands. More drastic efforts include social distancing, which is a set of nonpharmaceutical interventions or measures taken to prevent the spread of COVID-19 by maintaining a physical distance between people and reducing the number of times people come into close contact with each other; maintaining a safe distance from others; and sheltering practices by staying at home to avoid all contacts [3,6].

Governments in countries have instituted social distancing using policy directives and recommendations [7]. However, government oversights to influence citizens to social distancing and sheltering must wrestle with citizen's willingness to comply or adhere to the recommendations relevant to the disease prevention and mitigation. For example, whether a citizen is willing to comply with the social distancing recommendations provided by the government and maintain a safe distance from others or stay at home. Another example is citizen adherence to wearing masks in public as a precaution against COVID-19 [8]. Significant variations are observed in citizens' adherence to the COVID-19–relevant social distancing policy and masking in public across countries. The variation of agreeing that the guidelines are beneficial yet not following the guidelines by all citizens is a controversial issue in many countries [9,10]. A plausible reason for the variation in adherence to the policy recommendations by citizens is a culmination of informative and social influence [11] that is not explicitly understood in practice and academics well enough.

This study asks the research questions to what extent citizen's perceptions about government efforts in the COVID-19 situation influence the adherence actions and to what extent information sources, social media (SM) use, and subsequent knowledge influence the adherence actions. Insights into the underlying factors and dynamics involved with citizens' adherence process will inform the policy makers to follow appropriate communication and messaging strategies to influence citizens' willingness to adhere to the recommendations.

Citizen's perceptions of the government's response to the COVID-19 situation and subsequent recovery efforts are critical in the context of implicit or explicit response to the social distancing recommendations. The adherence decision may be reliant on several factors. In the context of personal health compliance behavior, individuals follow health protection suggestions based on the perception of severity and vulnerability of a situation [12,13]. Given a health-relevant policy recommendation, individuals will also assess the efficacy of the recommendations and whether they have abilities to manage the process. Given that the COVID-19 context is a stressful situation, the compliance action will depend on the cognitive appraisal processes of the situation and the differing sensitivity, vulnerability, interpretations, and reactions to the situation [14,15]. Based on these arguments, we propose to test the conjectures:

- Conjecture 1: Citizens' perceptions of the government's role to respond to the pandemic have a positive influence on adherence intention.
- Conjecture 2: Citizens' perceptions of the government's business reopening efforts have a positive influence on adherence intention.

The COVID-19 situation is not limited to *individuals* only; it extends to the social fabric. The policy recommendations have social implications as limiting interactions disrupts social group activities such as parties, get-togethers, and even shopping in malls. Thus, individual behavior would be affected by social norms and practices, as espoused by social psychology research [16,17], and individuals will try to understand and move through complex situations [18,19]. The COVID-19 situation is "difficult" in the sense that there is no treatment, and thus, individuals are more susceptible to learning from different sources and will subsequently manage the situation based on the knowledge gained from these information sources. The information available to people will also influence such behavior—more so in the current context of social media influences on people [20]. Therefore, we argue the variation in the adherence intentions at the individual level is an outcome of the information and knowledge gathering process. Thus, the common belief shaped by an individual is based on the information and knowledge that he gathers from different

sources. Given that social media is emerging as a highly critical information source to influence individual beliefs and perceptions, it is beyond doubt that such an influence would be quite effective in the COVID-19 situation [20-22]. Thus, we propose to test the following conjectures:

- Conjecture 3: A higher intensity of health information source (HIS) used for COVID-19 influences citizens' adherence intentions.
- Conjecture 4: A higher intensity of social media source use for COVID-19 information influences citizens' adherence intentions.
- Conjecture 5: A higher intensity of knowledge accrual for COVID-19 information influences citizens' adherence intentions.

Although the individual level variations depend on the perceptions about the government actions and knowledge gained from different sources, the individual is also embedded within the normative influence of a country and its cultural fabric to follow certain social and cultural norms [23]. As already observed, governmental efforts toward the COVID-19 situation have differing outcomes across countries. To explore this normative influence, we conduct a comparative evaluation of citizens' adherence process to COVID-19–relevant recommendations by the governments in three different countries: the United States, Kuwait, and South Korea. In summary, this study examines the impact of government efforts during the COVID-19 pandemic on citizens' adherence, as well as the impact of information sources, social media use, and subsequent COVID-19 knowledge on citizens' adherence in three different countries with differing cultures.

Methods

Recruitment

A focus group of 10 people in Kuwait was initially conducted. The focus group was asked to comment and provide insight into the subject matter. The feedback was that the cultural aspects are significant factors in terms of self adherence and belief of others adhering to governmental recommendations. Thus, we tried to expand to other countries with different cultures to understand the cultural impact on adherence better. However, although the chosen countries are different in culture, constrained by resources, the sampling strategy was limited to the countries that the authors are familiar with to gain firsthand experience to explain any similarities and differences.

A global survey-deploying firm collected the data for this study using online platforms. The firm recruited respondents from the United States, Kuwait, and South Korea in May 2020. The firm sampled respondents using age, gender, ethnicity, and a geographic region–based strata and quota matching process. Participation in the survey was free and voluntary; the respondents filled in electronic informed consent that was shown on the first page of the survey. The firm protects the confidentiality of anonymous respondents.

Data Collection

Data was collected using a survey instrument, as shown in Multimedia Appendix 1 Table A1. The questions asked participants about the cause and current state of the COVID-19 situation, their opinion on the government's role during the COVID-19 pandemic, and their information sources for the COVID-19 situation. The survey items included simple information-seeking questions, along with several existing validated scales from prior studies [24-30].

The survey instrument was pilot tested using a sample of 48 respondents, leading to minor refinements to a few items. A total of 535 participants took the survey. Because of missing responses to the items, 53 observations were excluded, resulting in a sample size of 482. Responses were coded, validated, and analyzed using STATA version 16 (StataCorp).

Sample Demographics

Table 1 shows the descriptive statistics and pairwise correlations among the key variables used in this study. A detailed descriptive statistic, correlation tables, and distribution details of several demographic controls used in the models are available in Multimedia Appendix 1 (Table A2, Textbox A1, Table A5, and Figures A1-A5).

Table 1. Summary statistics and pairwise correlations among key variables (N=482).

Variable	Mean (SD)	Min	Max	1	2	3	4	5	6	7
Self-adherence (1)	4.28 (1.13)	1.00	5.00	1.00						
Other's adherence (2)	3.16 (1.14)	1.00	5.00	0.24	1.00					
Response (3)	0.00 (0.80)	–2.66	0.75	0.69	0.31	1.00				
Reopen agreement (4)	2.34 (1.21)	1.00	5.00	–0.10	0.11	–0.22	1.00			
HIS[a]-general (5)	0.01 (0.56)	–1.80	2.14	0.11	0.02	0.00	0.13	1.00		
SM[b]-general (6)	–0.01 (0.92)	–2.58	1.95	–0.12	0.19	–0.03	0.01	–0.15	1.00	
Knowledge (7)	0.01 (2.33)	–7.37	5.53	0.34	–0.08	0.33	–0.20	–0.02	–0.11	1.00

[a]HIS: health information source.

[b]SM: social media.

Study Variables

The main dependent variables in this study are self-adherence and others' adherence. Self-adherence was measured using the question of whether they would comply with the social distancing measures. Others' adherence was measured by asking whether the respondent believes that others would comply with social distancing measures. Table 1 shows that, on average, self-adherence was 4.28 out of 5, showing that most people are adhering to social distancing recommendations. However, the mean for others' adherence was less, 3.16, showing lower levels of others' adherence to social distancing.

Five independent variables were of interest in this study. First, the independent variable response reflects on the individual's perception that the government response to the COVID-19 pandemic is effective. This variable was operationalized using four questions on the effectiveness, rights, and responsibilities of the government to respond to the COVID-19 situation (see Multimedia Appendix 1 Table A1). The internal consistency of the items was tested using Cronbach α (.81), and the standardized score was generated for the response variable. As Table 1 shows, the variable response has a mean of 0.00 and a standard deviation of 0.80.

The second independent variable, reopen agreement, reflects the individual's perception of whether the government has a right to decide when to reopen businesses. The variable was measured with a single item on a five-point scale. The mean of this variable was 2.34, with 1.21 standard deviation, indicating the disagreement bias of respondents toward government plans to reopen businesses.

The respondents were asked where they obtained health information on COVID-19, whether from social media, TV, newspapers, friends and family, or doctors and health care specialists. The HIS-general variable was coded using the scheme mentioned in Multimedia Appendix 1 Table A1. Furthermore, the SM-general variable was coded to reflect on the individual's intensity of social media platform use (further details on HIS-general and SM-general are provided in Multimedia Appendix 1 Textbox A2, Table A3, and Table A4).

The knowledge variable was coded to reflect the respondents' overall knowledge of COVID-19, as mentioned in Multimedia Appendix 1 Table A1. The mean for knowledge was 0.01, with a minimum of –7.37 and a maximum of 5.53, showing poor overall knowledge on COVID-19.

In addition to these key variables of interest, several control variables, as mentioned in Multimedia Appendix 1 Table A1, are included to account for counterfactual explanations relevant to our models.

Econometric Analysis

The empirical model specifies how individuals express their opinion toward adherence to social distancing and sheltering guidance set by the government. We specified self-adherence and others' adherence as two dependent variables. The set of independent variables we focused on include response, reopen agreement, HIS-general, SM-general, and knowledge. A set of control variables to enhance the robustness of our empirical model include demographics characteristics of the survey participants, such as gender, age group, household income, and ethnicity. The empirical model is described below:

$$Y(\textit{Self-Adherence, Others' Adherence})_i = \beta_0 + \beta_1 \times \textit{Response}_i + \beta_2 \times \textit{Reopen Agreement}_i + \beta_3 \times \textit{HIS-General}_i + \beta_4 \times \textit{SM-General}_i + \beta_5 \times \textit{Knowledge}_i + \textit{Controls}_i + \varepsilon_i$$

where *Controls* include gender, age groups, household income, and ethnicity. The country dummy was included in the full sample model but was removed for subsample analyses. Since the dependent variables are ordinal values, we used an ordered probit model to estimate the parameters of the key variables with robust standard errors (see Multimedia Appendix 1 Textbox A3 for details on the model). Using ordered probit regression, we estimated to what extent our set of key variables influence self-adherence and others' adherence separately.

Results

Table 2 presents the key estimation results (detailed estimations are provided in Multimedia Appendix 1 Table A6). The first four columns (1-4) in the table show the parameter estimates for the self-adherence dependent variable for the United States, Kuwait, South Korea, and the full sample (all). The columns 5-8 in Table 2 show the parameter estimates for the others' adherence regression for the United States, Kuwait, South Korea, and the full sample (all).

The first set of findings are relevant to the effect of the response variable on the outcomes. We found that the coefficients of response were positive and statistically significant at $P<.001$ across all columns. This suggests that individuals who believe that government response efforts are positive follow adherence guidance on social distancing and sheltering. At the same time, higher perception that government response efforts are positive leads individuals to believe that others will also follow adherence guidance on social distancing and sheltering.

We compared the coefficients of the estimated coefficients using a model-based chi-square comparison test with Bonferroni adjustments. Table 3 presents the results. We found that the positive effect of response on adherence shows that individuals who reside in the United States and Kuwait are more likely to follow adherence than South Korea.

Table 2. Ordered probit regression.[a]

Variables	DV[b]: self-adherence								DV: others' adherence							
	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)	
	US	*P* value	Kuwait	*P* value	South Korea	*P* value	All	*P* value	US	*P* value	Kuwait	*P* value	South Korea	*P* value	All	*P* value
Response	1.192 (0.154)[c]	<.001	0.929 (0.205)	<.001	2.342 (0.279)	<.001	1.108 (0.097)	<.001	0.353 (0.124)	.003	0.520 (0.133)	.001	1.494 (0.174)	<.001	0.636 (0.076)	<.001
Reopen agreement	–0.021 (0.083)	.82	–0.109 (0.106)	.33	–0.091 (0.120)	.53	–0.008 (0.054)	.89	0.106 (0.081)	.14	0.233 (0.088)	.004	0.381 (0.148)	.005	0.174 (0.050)	.001
HIS[d]-general	0.356 (0.174)	.06	0.532 (0.211)	.02	0.697 (0.289)	.049	0.309 (0.105)	.003	0.200 (0.114)	.17	–0.149 (0.131)	.33	0.537 (0.401)	.03	0.009 (0.089)	.92
SM[e]-general	0.010 (0.124)	.94	–0.132 (0.137)	.45	0.209 (0.274)	.51	0.002 (0.073)	.98	0.238 (0.113)	.03	0.329 (0.124)	.006	0.568 (0.287)	.05	0.254 (0.068)	<.001
Knowledge	0.121 (0.049)	.02	–0.041 (0.073)	.53	0.114 (0.088)	.18	0.074 (0.028)	.008	–0.009 (0.044)	.82	–0.155 (0.050)	.001	–0.054 (0.072)	.45	–0.086 (0.024)	<.001
Observations, n	207	N/A[f]	181	N/A	94	N/A	482	N/A	207	N/A	181	N/A	94	N/A	482	N/A
Pseudo R^2	0.211	N/A	0.222	N/A	0.505	N/A	0.268	N/A	0.054	N/A	0.134	N/A	0.328	N/A	0.092	N/A
χ^2 (*df*)	84.22 (21)	N/A	48.437 (21)	N/A	115.427 (21)	N/A	219.146 (21)	N/A	30.05 (21)	N/A	108.713 (21)	N/A	132.198 (21)	N/A	108.604 (21)	N/A

[a]Models include all controls: age, gender, household income, ethnicity.

[b]DV: dependent variable.

[c]Standard errors in parentheses.

[d]HIS: health information source.

[e]SM: social media.

[f]N/A: not applicable.

Table 3. Comparison of coefficients across countries.[a]

Variables	DV[b]: self-adherence						DV: others' adherence					
	US vs Kuwait	*P* value	US vs South Korea	*P* value	Kuwait vs South Korea	*P* value	US vs Kuwait	*P* value	US vs South Korea	*P* value	Kuwait vs South Korea	*P* value
Response	1.05	.84	13.09	.002	16.73	.01	0.85	>.99	28.68	<.001	19.83	<.001
Reopen agreement	0.43	.97	0.23	.99	0.01	>.99	1.13	>.99	2.68	.51	0.74	>.99
HIS[c]-general	0.42	.97	1.03	.84	0.21	.99	4.03	.22	0.66	>.99	2.66	.52
SM[d]-general	0.60	.95	0.44	.97	1.25	.78	0.30	>.99	1.16	>.99	0.59	>.99
Knowledge	3.39	.29	0.00	>.99	1.85	.62	4.78	.14	0.29	>.99	1.31	>.99
All countries	6.09	.29	13.36	.02	20.15	.001	10.14	.07	33.52	<.001	30.55	<.001

[a]Chi-square values reported with Bonferroni adjustment.

[b]DV: dependent variable.

[c]HIS: health information source.

[d]SM: social media.

The second set of findings were related to the effect of reopen agreement on the outcomes. As shown in Table 2, we found that the coefficient of reopen agreement was not significant for self-adherence for all the models. However, the coefficients of reopen agreement were positive and statistically significant for others' adherence for Kuwait (*P*=.008) and South Korea (*P*=.01) but not significant for the United States (*P*=.19). This set of findings suggests that in Kuwait and South Korea, an individual's opinion on the government's reopening efforts influence only their belief on others' following adherence guidance on social distancing and sheltering practices. In contrast, in the case of the United States, the agreement on the reopening of businesses does not have any effect on any adherence outcomes. Interestingly, across countries, there is not

a significant comparative difference in this effect, as the comparative chi-square values are not significant, as shown in Table 3.

A third finding was that the coefficients of HIS-general were positive and statistically significant (P=.003) for self-adherence only and not on others' adherence, suggesting that individuals use various information sources to guide their actions on adherence but not on others. Fourth, the coefficients of SM-general were positive and statistically significant for the United States (P=.04), South Korea (P=.048), and Kuwait (P=.008) for others' adherence only, suggesting that individuals use social media to formulate their belief on others' adherence to social distancing and sheltering behavior. Comparisons of coefficients for HIS-general and SM-general across three countries did not show statistically significant results.

Finally, the coefficients of knowledge on COVID-19 show different effects on self-adherence and others' adherence. For US residents, knowing more about the virus is positively associated with self-adherence (column 1 of Table 2). However, residents in Kuwait who know more about the virus are less likely to believe that others will adhere to social distancing and sheltering guidance (column 6 of Table 2). The full sample analysis shows that the direction of the coefficients was positive for self-adherence and negative for other's adherence, suggesting that this effect is consistent overall but differs across countries.

Discussion

Findings and Implications

In general, this study finds that, in all three countries, government response efforts, business reopening agreements, as well as the intensity of information source use, social media use, and knowledge about the COVID-19 pandemic all influence either self-adherence or belief in others adhering. Table 4 summarizes the findings from this study.

Table 4. Summary of findings.

Variables	Self-adherence				Others' adherence				Findings
	US	Kuwait	South Korea	All	US	Kuwait	South Korea	All	
Response	Pos[a]	Pos	Pos	Pos	Pos	Pos	Pos	Pos	C[b]1: Supported for self-adherence and others' adherence; US and Kuwait have better outcomes than South Korea
Reopen agreement	NS[c]	NS	NS	NS	NS	Pos	Pos	Pos	C2: Partially supported for others' adherence; no comparative difference results across countries
HIS[d]-general	Pos	Pos	Pos	Pos	NS	NS	NS	NS	C3: Supported for self-adherence; no comparative difference results across countries
SM[e]-general	NS	NS	NS	NS	Pos	Pos	Pos	Pos	C4: Supported for others' adherence; no comparative difference results across countries
Knowledge	Pos	NS	NS	Pos	NS	Neg[f]	NS	Neg	C5: Partially supported for self-adherence (positive for the US only) and others' adherence (negative for Kuwait only)

[a]POS: positive association.

[b]C: conjecture.

[c]NS: not significant.

[d]HIS: health information source.

[e]SM: social media.

[f]Neg: negative association.

The first finding inherently suggests a *persuasion effect* for governments' adherence guidelines on individuals. This persuasion effect in terms of conveying information and communicating risks to the public during the COVID-19 crisis are becoming complicated due to the large amount of misinformation, uncertainty surrounding the source and spread of the virus, and the absence of a vaccine [31-35]. However, previous research has shown that trusting the health system has a significant impact on public willingness to receive and adhere to health instructions [36]. This persuasion effect is highly effective when the government provides guidelines' context (ie, detailed explanations, implications, and consequences of the guidelines to the individuals). However, the magnitude of the persuasion effect differs across the three countries sampled in this study. Several reasons can be attributed to explain the differences, with a plausible one pointing to the expectations of citizens that the government is doing everything right for them. In the case of the United States and Kuwait, the citizens expect that the governments need to provide enough information and rationale to take away their individual and social freedom in the form of sheltering and distancing measures. However, citizens of South Korea are more willing to follow government guidelines during a national crisis.

The second finding suggests that individuals expect that others will adhere to guidelines given that they agree to the reopening stipulations—although they themselves may not do so. A plausible argument points to human's self-interests. To instantiate, individuals would expect that life should go back to normal in the post–COVID-19 situation [37]. The COVID-19 situation has led to countrywide unemployment, economic

downturn, and a lack of supplies—leading to a stressful situation for individuals who carry on with their livelihoods. Although reopening provides avenues for economic recovery, there is no guarantee that people will follow the suggested safe-distancing and other guidelines. Research finds that there is a positive and significant correlation between overall risk perception and economic threat perception [36]. People that perceive a high personal economic threat may feel that the government is not managing the crisis well; this thereby impacts their adherence. Therefore, individuals are a bit cautious but hopeful that recovery attempts will increase others' adherence to the guidelines.

The third set of findings emphasizes the information-based decision-making process of individuals in the COVID-19 situation. As much as we have validated the influence of the perceptions of the individuals on the government's actions, the adherence decision is influenced by the information that is available to the individual. Moreover, following our coding scheme and the sensitivity checks of the impact of general versus specific information sources on the self-adherence outcomes, it can be inferred that channeling information to citizens should follow a broader spectrum [34]. In other words, individuals are influenced by broader information sources such as friends, families, doctors, and social sources rather than relying entirely on only popular channels such as newspapers and television. This is the same irrespective of the individual's country, referring to any lack of comparative effects across the countries. This finding is in line with research done on the social influence that shows that health behaviors spread through perceived norms with whom common identities are most shared [38]. Social networks can amplify the spread of behavior regardless of whether the behavior is harmful or beneficial to the pandemic [39,40]. Therefore, sending out information such as *nudges* [40] that is related to what others within the community are doing would influence an individual adherence level. Nudges and normative information may be an alternative method that the government can use to impose the spread of policies better.

The fourth set of findings highlights some interesting roles of social media on the individual's adherence decisions in the COVID-19 situation. Although previous studies have found an impact of social media use on individual behavior during the COVID-19 pandemic [41,42], this study examines the impact of social media use on adherence. This study finds that the influence of social media is significant on others' adherence but not on self-adherence. We contrasted and compared this with the previously mentioned information source effects. As it was observed, information sources influence self-adherence, but social media influences others' adherence. These findings would suggest that shaping social norms is effective through social media. In this effect, the prominence of social media platforms such as Twitter, Instagram, Snapchat, and WhatsApp is additionally validated in the analysis, compared to Facebook and YouTube. Plausibly, some social media platforms such as Twitter and Instagram are relatively informative; some are more socially inclusive ones; WhatsApp, Snapchat, and a few others like YouTube and Facebook are entertainment-oriented platforms. These distinctions are not exclusive but may help to explain the higher impact of informative social media versus entertaining social media to shape an expected social norm regarding adhering decisions.

The final set of findings and implications are about the effect of the knowledge level on the symptoms, treatments, and risks associated with COVID-19 on adherence intentions. A recent study has found a positive and significant correlation between awareness and knowledge, and adherence to social distancing [43]. This study supports this finding and further finds that a higher knowledge level is influential in shaping the decisions on self-adherence for the United States, whereas it raises a skepticism on others' adherence decisions for Kuwait. A possible reason may be due to differences in the cultural norm as well as the information surrounding the culture or community. Disseminating accurate information about what most people are doing is helpful and health-promoting; however, if the surrounding community is not health-promoting, providing purely descriptive normative information may backfire by reducing positive behaviors among people who already engage in them [40]. Furthermore, when it comes to health information dissemination, often less is better than giving too much, as deeper knowledge may sway different perceptions or lead to anxiety and panic [44,45]. Thus, a balanced approach in disseminating details about COVID-19 and possible implications is essential while keeping in mind that not all information would lead to positive outcomes.

Practice and Policy Implications

A set of implications and recommendations for public health officials and policy makers can be drawn from this study. This study points to the gap in the responsible behavior of individuals while following self-adherence versus expecting others' decisions. In that, the current government efforts to mitigate and the expectation that everything should be back to normal have differing consequences. Existing work suggests that the efficacy and outcomes of policy recommendations depend on the individuals' beliefs and subsequent action [46,47]. The first step in this process is the strong belief of whether the recommended action will help in mitigating the threat or situation. The key to minimizing rejection and maximizing acceptance of recommendation, thus, is reliant on how the recommendations are framed and the subsequent ways the relevant information is disseminated and communicated. Care needs to be taken in framing information dissemination and communication channels in managing the current COVID-19 pandemic.

The understanding of the aspects relevant to the persuasion process of information dissemination to citizens so that not only negative and fear-based appeals and messages are spread but also positive and emotion-based messages that can lead to better appeals is important. As it is already seen, popular press and television channels may not do justice to the messages unless the source design of the messages constructs the message appropriately. The snowballing effects of fear-based appeals and subsequent rejection would have a highly negative consequence.

The study also highlights the importance of information sources in terms of compliance intention and, thus, the importance of managing misinformation and misleading channels. Health

authorities have warned that widespread misinformation about COVID-19 is a severe concern causing xenophobia worldwide [48]. Given the spread of information through the channels, it is not easy for individuals to discern individual responsibilities and expectations of others, without which the situation may get quickly aggravated and spread panic. Thus, as much as channels are used to disseminate, the prominence of social media in this pandemic cannot be underestimated [40], and effective social media strategies need to be deployed by policy makers to defuse the fear, anxiety, and panic-inducing effects of information by nudging positive health behaviors within communities [14,49].

This study has implications for health systems across countries that are facing heightened levels of stress because of COVID-19. Efforts to contain the spread of the disease have virtually failed, leading to ventilator-based treatment of patients with COVID-19. Health systems are working to increase supply by finding any way possible to create capacity for patients with COVID-19 while following tactics such as stopping elective procedures and surgeries.

In this context, knowledge-based decisions by citizens play an important role, such as to know whether to go to a hospital for COVID-19 symptoms or not. Knowledge of exact symptoms and risks may empower citizens to make decisions at their ends and stay away from hospitals, and avail possible treatments at their end.

Thus, this study unravels a set of insights, reflecting and collecting data on the differing adherence outcomes across countries. The insights inform four policy recommendations: (1) governments should tell citizens precisely what individuals need to do for themselves; (2) governments should also tell what individuals should expect from others and what others need to do when dealing with other individuals; (3) messages from policy makers to citizens must be done through social media and similar channels rather than only through press and TV; and (4) messages need to provide knowledge about the disease along with information that should calm the citizens, and messages should nudge citizens on what others in their community are doing in terms of health-promoting behavior.

Limitations

This study examines factors that influence citizens adhering to social distancing and their belief in others' adherence of social distancing at a point in time. However, the citizen might go back and forth in the adherence process, and the knowledge level or source of information may change over time. This is a limitation of this study, as the data set used is a cross-sectional survey. Future studies could examine how a citizen's adherence changes over time. In addition, using the random sampling process of the public in the United States, Kuwait, and South Korea samples may include fewer familiar respondents to the study context. In particular, the questionnaire was on the internet. Therefore, respondents are all users of the internet. The study does not examine noninternet users, which could have differential impacts. Thus, the generalization of the sample to a uniform national culture characteristic is a limitation of this study. Future studies could conduct both internet and noninternet surveys, and examine the difference in terms of adherence behavior.

Conclusions

The rapid spread of COVID-19 has generated great public interest and media coverage around the outbreak, and it has created social unrest and economic downturn. Strict measures of mitigation have been implemented to avoid the breakdown of the health system [50] and reduce the deaths caused by the virus. Efforts to contain the spread of the disease have virtually failed with efforts to find a way to keep the economies sustained. In this context, the social distancing policy deems to be a lifesaver not only for individuals but also for economies across the world.

As the COVID-19 global pandemic continues to grow and governmental restrictions are ongoing, it is critical to understand people's frustration to reduce panic and promote social distancing to facilitate the control of the pandemic. This study finds that the government plays a central role in terms of adherence to restrictions and provides several insights that are important for the government to frame the message, information, and communication for the citizens.

The pandemic may end socially or medically—in the sense that, prior to the vaccine or effective treatment, exhausted and frustrated people will return to regular life [51]. As several countries are lifting restrictions, listening to people's voices of frustrations, an equal and opposite criticism is often voicing that these steps are premature. This conflict is not easy to resolve and points to the need for understanding public voices and opinions more carefully than the medical end of a pandemic. This study is a step in exploring more nuances to the citizen's mindsets underlying the voices, and we expect that the recommendations based on the insights will provide effectiveness to the social distancing and masking measures, and will help to curb COVID-19.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire and further detailed analysis.
[DOC File , 366 KB - jmir_v22i8e20634_app1.doc]

References

1. Coronavirus disease 2019 (COVID-19): symptoms and causes. Mayo Clinic. URL: https://www.mayoclinic.org/diseases-conditions/coronavirus/symptoms-causes/syc-20479963 [accessed 2020-05-20]
2. COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). John Hopkins University. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-05-01]
3. Q&A on coronaviruses (COVID-19). World Health Organization. URL: https://www.who.int/news-room/q-a-detail/q-a-coronaviruses [accessed 2020-05-10]
4. Prompetchara E, Ketloy C, Palaga T. Immune responses in COVID-19 and potential vaccines: lessons learned from SARS and MERS epidemic. Asian Pac J Allergy Immunol 2020 Mar;38(1):1-9 [FREE Full text] [doi: 10.12932/AP-200220-0772] [Medline: 32105090]
5. Fisher D, Heymann D. Q&A: the novel coronavirus outbreak causing COVID-19. BMC Med 2020 Feb 28;18(1):57 [FREE Full text] [doi: 10.1186/s12916-020-01533-w] [Medline: 32106852]
6. Guest JL, Del Rio C, Sanchez T. The three steps needed to end the COVID-19 pandemic: bold public health leadership, rapid innovations, and courageous political will. JMIR Public Health Surveill 2020 Apr 06;6(2):e19043 [FREE Full text] [doi: 10.2196/19043] [Medline: 32240972]
7. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. Corrigendum to "World Health Organization declares Global Emergency: A review of the 2019 Novel Coronavirus (COVID-19)" [Int. J. Surg. 76 (2020) 71-76]. Int J Surg 2020 May;77:217 [FREE Full text] [doi: 10.1016/j.ijsu.2020.03.036] [Medline: 32305321]
8. Brooks JT, Butler JC, Redfield RR. Universal masking to prevent SARS-CoV-2 transmission-the time is now. JAMA 2020 Jul 14:A. [doi: 10.1001/jama.2020.13107] [Medline: 32663243]
9. Igielnik R. Most Americans say they regularly wore a mask in stores in the past month; fewer see others doing it. Pew Research Center. 2020. URL: https://pewrsr.ch/2Bu7qIi [accessed 2020-07-20]
10. Saey T. Why scientists say wearing masks shouldn't be controversial. Science News. URL: https://www.sciencenews.org/article/covid-19-coronavirus-why-wearing-masks-controversial [accessed 2020-07-20]
11. Farooq A, Laato S, Islam AKMN. Impact of online information on self-isolation intention during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2020 May 06;22(5):e19128 [FREE Full text] [doi: 10.2196/19128] [Medline: 32330115]
12. Rogers RW. A protection motivation theory of fear appeals and attitude Change1. J Psychol 1975 Sep;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
13. Floyd DL, Prentice-Dunn S, Rogers RW. A meta-analysis of research on protection motivation theory. J Appl Social Pyschol 2000 Feb;30(2):407-429. [doi: 10.1111/j.1559-1816.2000.tb02323.x]
14. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
15. Monat A, Lazarus R, editors. Stress and Coping: An Anthology, 3rd ed. US: Columbia University Press; 1991:-231.
16. Aronson E, Akert R, Wilson T. Sozialpsychologie. Germany: Pearson; 2010:3827372763.
17. Cialdini RB, Goldstein NJ. Social influence: compliance and conformity. Annu Rev Psychol 2004;55:591-621. [doi: 10.1146/annurev.psych.55.090902.142015] [Medline: 14744228]
18. Lucas T, Alexander S, Firestone IJ, Baltes BB. Self‐efficacy and independence from social influence: discovery of an efficacy–difficulty effect. Soc Influence 2006 Mar;1(1):58-80. [doi: 10.1080/15534510500291662]
19. Baron RS, Vandello JA, Brunsman B. The forgotten variable in conformity research: Impact of task importance on social influence. J Pers Soc Psychol 1996;71(5):915-927. [doi: 10.1037/0022-3514.71.5.915]
20. Hernández-García I, Giménez-Júlvez T. Assessment of health information about COVID-19 prevention on the internet: infodemiological study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717 [FREE Full text] [doi: 10.2196/18717] [Medline: 32217507]
21. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10):2000199 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
22. Bhagavathula AS, Aldhaleei WA, Rahmani J, Mahabadi MA, Bandari DK. Knowledge and perceptions of COVID-19 among health care workers: cross-sectional study. JMIR Public Health Surveill 2020 Apr 30;6(2):e19160 [FREE Full text] [doi: 10.2196/19160] [Medline: 32320381]
23. Vora D, Martin L, Fitzsimmons SR, Pekerti AA, Lakshman C, Raheem S. Multiculturalism within individuals: a review, critique, and agenda for future research. J Int Bus Stud 2018 Dec 6;50(4):499-524. [doi: 10.1057/s41267-018-0191-3]
24. Macdonell K, Chen X, Yan Y, Li F, Gong J, Sun H, et al. A protection motivation theory-based scale for tobacco research among Chinese youth. J Addict Res Ther 2013 Jul 08;4:154 [FREE Full text] [doi: 10.4172/2155-6105.1000154] [Medline: 24478933]
25. Milne S, Sheeran P, Orbell S. Prediction and intervention in health-related behavior: a meta-analytic review of protection motivation theory. J Appl Social Pyschol 2000 Jan;30(1):106-143 [FREE Full text] [doi: 10.1111/j.1559-1816.2000.tb02308.x]

26. Xiao H, Li S, Chen X, Yu B, Gao M, Yan H, et al. Protection motivation theory in predicting intention to engage in protective behaviors against schistosomiasis among middle school students in rural China. PLoS Negl Trop Dis 2014 Oct;8(10):e3246 [FREE Full text] [doi: 10.1371/journal.pntd.0003246] [Medline: 25329829]
27. Yan Y, Jacques-Tiura AJ, Chen X, Xie N, Chen J, Yang N, et al. Application of the protection motivation theory in predicting cigarette smoking among adolescents in China. Addict Behav 2014 Jan;39(1):181-188 [FREE Full text] [doi: 10.1016/j.addbeh.2013.09.027] [Medline: 24157424]
28. Hassani L, Dehdari T, Hajizadeh E, Shojaeizadeh D, Abedini M, Nedjat S. Development of an instrument based on the protection motivation theory to measure factors influencing women's intention to first pap test practice. Asian Pac J Cancer Prev 2014;15(3):1227-1232 [FREE Full text] [doi: 10.7314/apjcp.2014.15.3.1227] [Medline: 24606445]
29. Xiao H, Peng M, Yan H, Gao M, Li J, Yu B, et al. An instrument based on protection motivation theory to predict Chinese adolescents' intention to engage in protective behaviors against schistosomiasis. Glob Health Res Policy 2016;1:15 [FREE Full text] [doi: 10.1186/s41256-016-0015-6] [Medline: 29202064]
30. Corona virus disease 2019 (COVID-19) frequently asked questions. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/faq.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fprepare%2Fchildren-faq.html [accessed 2020-05-02]
31. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med 2020 Apr;26(4):450-452 [FREE Full text] [doi: 10.1038/s41591-020-0820-9] [Medline: 32284615]
32. Larson HJ. Blocking information on COVID-19 can fuel the spread of misinformation. Nature 2020 Apr;580(7803):306. [doi: 10.1038/d41586-020-00920-w] [Medline: 32231320]
33. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
34. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
35. Bastani P, Bahrami MA. COVID-19 related misinformation on social media: a qualitative study from Iran. J Med Internet Res 2020 Apr 05:A. [doi: 10.2196/18932] [Medline: 32250961]
36. Gesser-Edelsburg A, Cohen R, Hijazi R, Abed Elhadi Shahbari N. Analysis of public perception of the Israeli government's early emergency instructions regarding COVID-19: online survey study. J Med Internet Res 2020 May 15;22(5):e19370 [FREE Full text] [doi: 10.2196/19370] [Medline: 32392172]
37. James G. Coronavirus: when will the outbreak end and life get back to normal? BBC. URL: https://www.bbc.com/news/health-51963486 [accessed 2020-05-20]
38. Centola D. An experimental study of homophily in the adoption of health behavior. Science 2011 Dec 02;334(6060):1269-1272 [FREE Full text] [doi: 10.1126/science.1207055] [Medline: 22144624]
39. Christakis NA, Fowler JH. Social network sensors for early detection of contagious outbreaks. PLoS One 2010 Sep 15;5(9):e12948 [FREE Full text] [doi: 10.1371/journal.pone.0012948] [Medline: 20856792]
40. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
41. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public engagement and government responsiveness in the communications about COVID-19 during the early epidemic stage in China: infodemiology study on social media data. J Med Internet Res 2020 May 26;22(5):e18796 [FREE Full text] [doi: 10.2196/18796] [Medline: 32412414]
42. Sesagiri Raamkumar A, Tan SG, Wee HL. Measuring the outreach efforts of public health authorities and the public response on Facebook during the COVID-19 pandemic in early 2020: cross-country comparison. J Med Internet Res 2020 May 19;22(5):e19334 [FREE Full text] [doi: 10.2196/19334] [Medline: 32401219]
43. Qazi A, Qazi J, Naseer K, Zeeshan M, Hardaker G, Maitama JZ, et al. Analyzing situational awareness through public opinion to predict adoption of social distancing amid pandemic COVID-19. J Med Virol 2020 Jul;92(7):849-855 [FREE Full text] [doi: 10.1002/jmv.25840] [Medline: 32266990]
44. Fu S, Li H, Liu Y, Pirkkalainen H, Salo M. Social media overload, exhaustion, and use discontinuance: examining the effects of information overload, system feature overload, and social overload. Inf Processing Manage 2020 Nov;57(6):102307. [doi: 10.1016/j.ipm.2020.102307]
45. Matthes J, Karsay K, Schmuck D, Stevic A. “Too much to handle”: impact of mobile social networking sites on information overload, depressive symptoms, and well-being. Comput Hum Behav 2020 Apr;105(6):106217. [doi: 10.1016/j.chb.2019.106217]
46. Mukhtar S. Mental health and emotional impact of COVID-19: applying health belief model for medical staff to general public of Pakistan. Brain Behav Immun 2020 Jul;87:28-29 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.012] [Medline: 32283289]
47. Guidry JPD, Carlyle KE, LaRose JG, Perrin P, Messner M, Ryan M. Using the health belief model to analyze Instagram posts about Zika for public health communications. Emerg Infect Dis 2019 Jan;25(1):179-180. [doi: 10.3201/eid2501.180824] [Medline: 30561302]

48. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
49. Seymour B, Getman R, Saraf A, Zhang LH, Kalenderian E. When advocacy obscures accuracy online: digital pandemics of public health misinformation through an antifluoride case study. Am J Public Health 2015 Mar;105(3):517-523. [doi: 10.2105/AJPH.2014.302437] [Medline: 25602893]
50. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: digital topic modeling approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]
51. Kolata G. How pandemics end. Forbes India. URL: https://www.forbesindia.com/article/coronavirus/how-pandemics-end/59305/1 [accessed 2020-05-14]

Abbreviations

COVID-19: coronavirus disease
HIS: health information source
SM: social media

Edited by G Eysenbach; submitted 24.05.20; peer-reviewed by B Akinwunmi, C García; comments to author 16.07.20; revised version received 21.07.20; accepted 26.07.20; published 11.08.20.

Please cite as:
Al-Hasan A, Yim D, Khuntia J
Citizens' Adherence to COVID-19 Mitigation Recommendations by the Government: A 3-Country Comparative Evaluation Using Web-Based Cross-Sectional Survey Data
J Med Internet Res 2020;22(8):e20634
URL: http://www.jmir.org/2020/8/e20634/
doi:10.2196/20634
PMID:32716896

Original Paper

Evaluation of Korean-Language COVID-19–Related Medical Information on YouTube: Cross-Sectional Infodemiology Study

Hana Moon[1], MD, PhD; Geon Ho Lee[1], MD, MSc

Department of Family Medicine, School of Medicine, Daegu Catholic University, Daegu, Republic of Korea

Corresponding Author:
Geon Ho Lee, MD, MSc
Department of Family Medicine
School of Medicine
Daegu Catholic University
33, Duryugongwon-ro 17-gil, Nam-gu
Daegu, 42472
Republic of Korea
Phone: 82 53 650 4696
Email: totoslee@cu.ac.kr

Abstract

Background: In South Korea, the number of coronavirus disease (COVID-19) cases has declined rapidly and much sooner than in other countries. South Korea is one of the most digitalized countries in the world, and YouTube may have served as a rapid delivery mechanism for increasing public awareness of COVID-19. Thus, the platform may have helped the South Korean public fight the spread of the disease.

Objective: The aim of this study is to compare the reliability, overall quality, title–content consistency, and content coverage of Korean-language YouTube videos on COVID-19, which have been uploaded by different sources.

Methods: A total of 200 of the most viewed YouTube videos from January 1, 2020, to April 30, 2020, were screened, searching in Korean for the terms “Coronavirus,” “COVID,” “Corona,” “Wuhan virus,” and “Wuhan pneumonia.” Non-Korean videos and videos that were duplicated, irrelevant, or livestreamed were excluded. Source and video metrics were collected. The videos were scored based on the following criteria: modified DISCERN index, Journal of the American Medical Association Score (JAMAS) benchmark criteria, global quality score (GQS), title–content consistency index (TCCI), and medical information and content index (MICI).

Results: Of the 105 total videos, 37.14% (39/105) contained misleading information; independent user–generated videos showed the highest proportion of misleading information at 68.09% (32/47), while all of the government-generated videos were useful. Government agency–generated videos achieved the highest median score of DISCERN (5.0, IQR 5.0-5.0), JAMAS (4.0, IQR 4.0-4.0), GQS (4.0, IQR 3.0-4.5), and TCCI (5.0, IQR 5.0-5.0), while independent user–generated videos achieved the lowest median score of DISCERN (2.0, IQR 1.0-3.0), JAMAS (2.0, IQR 1.5-2.0), GQS (2.0, IQR 1.5-2.0), and TCCI (3.0, IQR 3.0-4.0). However, the total MICI was not significantly different among sources. “Transmission and precautionary measures” were the most commonly covered content by government agencies, news agencies, and independent users. In contrast, the most mentioned content by news agencies was “prevalence,” followed by “transmission and precautionary measures.”

Conclusions: Misleading videos had more likes, fewer comments, and longer running times than useful videos. Korean-language YouTube videos on COVID-19 uploaded by different sources varied significantly in terms of reliability, overall quality, and title–content consistency, but the content coverage was not significantly different. Government-generated videos had higher reliability, overall quality, and title–content consistency than independent user–generated videos.

(J Med Internet Res 2020;22(8):e20775) doi:10.2196/20775

KEYWORDS

COVID-19; YouTube; social media; misinformation; public health surveillance; health communication; consumer health information; health education; infectious disease outbreaks; infodemiology; infoveillance; infodemic; internet; multimedia

Introduction

The coronavirus disease (COVID-19) is rapidly spreading all over the world. South Korea has noticeably flattened the curve of infection and recorded low fatality rates despite an explosion of cases after the 31st patient was confirmed [1]. The South Korean government's key strategies were transparency in sharing information, mass screening, tracking of suspected cases, and the reallocation of medical resources [1].

However, without public cooperation, these governmental measures would not have been effective. South Korea's success, thus far, prompts us to seek answers to the following questions:

- "What explains the early adoption of social distancing among Koreans?"
- "Why did they come to accept the cost to their privacy and voluntarily cooperate with authorities?"
- "What persuaded them to change their behaviors in such a short time?"

YouTube may have served as a valuable tool in providing information on COVID-19, creating understanding among South Koreans and urging them to cooperate with the authorities in implementing precautionary measures. South Korea is one of the best connected countries in the world, with 88.5% of the population accessing the internet via smartphones in 2019 [2]. Even on public transportation such as buses, subways, and trains, South Koreans enjoy free public Wi-Fi. Thus, commuters can easily access online video platforms without worrying about their phone bill. YouTube has become the most popular video platform in South Korea in terms of monthly average use time [3]. Bearing in mind that health conscious consumers often search for health information online, South Koreans may have accessed YouTube for information on COVID-19 [4]. Therefore, due to its popularity, YouTube may have provided timely and relevant information on COVID-19 to the public.

The popularity of a YouTube video has been the focus of marketing researchers since the platform became one of the leading types of advertising media [5]. Researchers primarily ask, "What key factors determine video popularity?" Many factors have been suggested, such as the video's title, thumbnail, subtitles, video upload date and time, delivery style, running time, or social network of the YouTube channel [6]. There have also been attempts to build computer-based video popularity prediction models [7-9]. However, the measurement of popularity is not easy to determine because video metrics are constantly evolving. For example, it seems reasonable to consider a video clip with 1 million views in 1 week popular compared to a video with the same views that have been accumulating for 5 years. In medical informatics, the video power index (VPI) has been proposed to solve this problem using a simple formula [10]. VPI captures the views, likes, and dislikes over the number of days that the video has been posted.

In contrast with marketing researchers who have studied social media to promote their brand or merchandise, medical researchers have been using social media for public health surveillance [11]. Social media can be used as a tool to provide a snapshot of the public's interest in and response to ongoing health issues [12]. A wide variety of health topics have been analyzed so far, including infectious diseases, mental disorders, and chronic diseases [11]. For example, previous studies have quantitatively analyzed Twitter data to assess people's attitudes toward mental illness [13,14]. In addition to YouTube, various sources were used to retrieve health-related data, including search engines, blogs, forums, and social media platforms such as Google, Bing, Baidu, Yahoo, Twitter, and Facebook [15]. The frequency of relevant keywords or trending hashtags, as well as numbers of views, likes, and shares, is typically measured. Manual coding and computer-based methods including content analysis, text mining, natural language processing, and topic modeling have been used to determine the most talked about topics [11,16]. Moreover, there were also attempts to predict infectious disease outbreaks or quickly detect a person who has depression. Demand-based infoveillance studies using internet search queries have primarily focused on predicting infectious disease outbreaks, such as Zika, influenza, dengue, and the measles virus [17-19]. Other studies analyzed user's behavior on social media and proposed a model that was based on machine learning for the early detection of depression and suicidal risk [20,21].

Content analysis has been widely used in YouTube studies involving online medical information. The question of who is supplying what information has been addressed in various fields of research, including ulcerative colitis, tinnitus, sleep apnea, cervical cancer, and orthodontics [22-26]. In contrast with Twitter studies that use computer-assisted content coding, the content in these studies was extracted from videos and manually coded. Notably, the medical information and content index (MICI) has been a commonly used tool in infectious disease studies since it was proposed to systematically assess the content coverage of Ebola hemorrhagic fever [27]. MICI assesses five key components for understanding infectious disease: (1) prevalence, (2) transmission, (3) clinical symptoms, (4) screening and testing, and (5) treatment and outcome.

Recently, YouTube has garnered attention from researchers in medical communication and education. Traditionally, written medical information has been criticized for its low accessibility. Extensive use of medical jargon, with which laypeople may be unfamiliar, hinders the delivering of complete messages [28]. Such accessibility problems have also been a barrier in informed or shared decision making [29]. In contrast, YouTube videos offer easy-to-understand information because videos can contain multimedia elements such as graphics, animations, and voice-overs using verbal expressions. In this regard, YouTube could be a user-friendly tool to educate the public on health-related topics.

However, there are concerns about the reliability and quality of online information. Viewers may be exposed to misinformation because YouTube does not have a verification process that videos must pass before being published. The spread of misinformation via social media is amplified by the filter bubble and the echo chamber effect. Many people access news from their social media feeds, which are curated by an algorithm for each person. This filtered information exposes users less to opposing viewpoints and isolates them in their own bubbles [30]. Furthermore, the echo chamber effect means that people

prefer to read articles that confirm and strengthen their original opinions. Users also prefer to interact with like-minded people, allowing misinformation more opportunities to go viral [31]. It is the responsibility of each YouTube user to be aware of the veracity of the information they watch and share.

Moreover, the reliability and quality of medical information are of the highest importance. Inaccurate information may lead to physical harm or irreversible damage. Medical misinformation may have life-threatening consequences for vulnerable populations such as patients with cancer, children, and older people. For example, patients with cancer taking alternative medicine are more likely to refuse evidence-based therapies and have higher mortality rates than patients who do not take alternative medicine [32]. Moreover, thriving antivaccine movements on social media have made parents hesitant to have their children vaccinated, possibly contributing to a reduction in vaccination rates and leading to multiple measles outbreaks [33].

Researchers, therefore, have tried to compare misleading and useful information as well as evaluate the reliability and quality of consumer medical information. Some studies have attempted to examine the differences in video metrics between misleading and useful videos, including the number of views, likes, dislikes, and running time [34-37]. The results have been controversial. Various measurement tools have been suggested to evaluate the reliability and quality of consumer medical information, such as DISCERN, Health on the Net code, the Journal of the American Medical Association Score (JAMAS) benchmark criteria, and the global quality score (GQS) [10,38-43]. Other studies have compared videos based on their sources (ie, government agencies, news agencies, health care professionals, and independent users) and then on the reliability and quality of these sources. Their results indicate concerns about the reliability of information that is neither monitored nor filtered. Although government- or professionally generated videos were more likely to be reliable and accurate, they were falling behind in nurturing their YouTube presence [34,44,45]. In contrast, individual user-generated videos were superior in number [35]; however, they were more likely to be inaccurate and incomplete [46]. Consequently, consumers are exposed to misleading medical information.

The impact of disseminating misinformation is particularly important in the context of public health emergencies. False beliefs or misperceptions disseminated via YouTube can spread mistrust toward authorities, generate confusion, and heighten public anxiety. Furthermore, fake news can cause people to engage in undesirable behaviors such as panic buying of food, medicines, and toilet paper due to fears about the pandemic [47]. Panic buying and hoarding can make it difficult for physically challenged or older people to buy essential products. Therefore, several studies have analyzed the spread of medical information through YouTube videos on outbreaks of several infectious diseases such as the H1N1 influenza, Ebola virus disease, and Zika virus disease [27,35,48,49].

Social media has been analyzed as a source of information on infectious disease in a number of studies; however, there is limited research on COVID-19. Twitter and Weibo have been analyzed to understand the impacts of COVID-19 and social distancing on mental health [50,51]. Although several studies have analyzed the content of COVID-19 videos in English, Spanish, and Chinese, this study is, as far as we are aware, the first to evaluate the Korean-language content of COVID-19 videos on YouTube [36,44,52]. Korean, which is a distinct language and not a dialect of Chinese or Japanese, is the only official language of Korea; therefore, it is important to evaluate the Korean video content.

This study investigated three main research questions:

1. Are there differences in video metrics between misleading and useful Korean-language videos about COVID-19 on YouTube?
2. Do Korean-language YouTube videos on COVID-19 that are uploaded by different sources vary significantly in reliability, overall quality, title–content consistency, and content coverage?
3. Do government videos have higher reliability, overall quality, title–content consistency, and content coverage compared to independent user–generated videos?

Methods

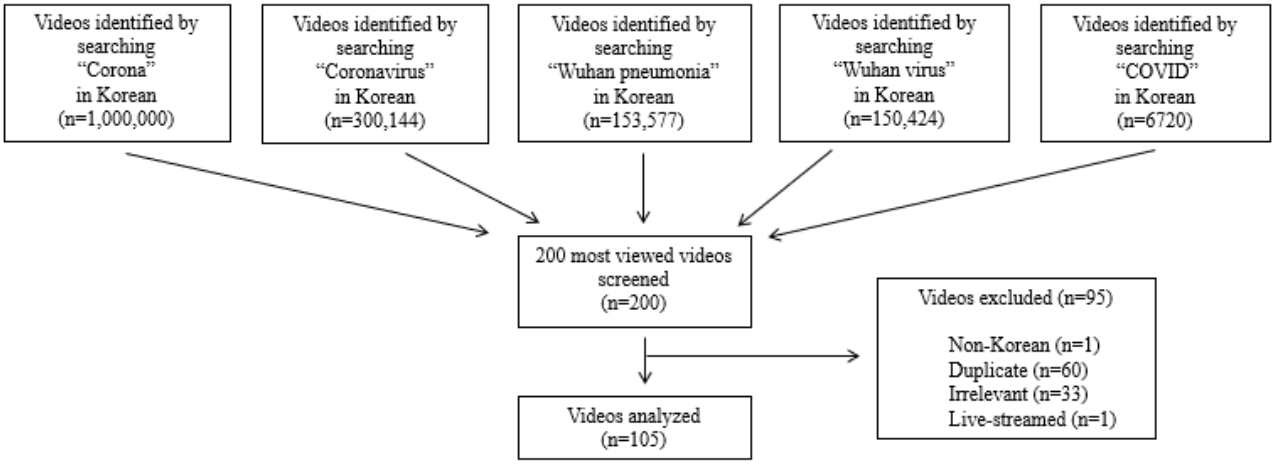
Recruitment

Data were obtained from publicly available YouTube videos. A total 200 of the most extensively viewed videos were screened from the YouTube video application programming interface. The search criteria used to obtain the videos comprised of the following terms in Korean: “Coronavirus,” “COVID,” “Corona,” “Wuhan virus,” and “Wuhan pneumonia.” Although the authors are aware that the last two keywords are inappropriate as they could create the impression of discrimination, these terms had to be included in the search to identify the videos uploaded on YouTube during the initial phase of the COVID-19 pandemic from January to early February 2020. Prior to the World Health Organization announcing the official name of the virus on February 11, 2020, “Wuhan virus” and “Wuhan pneumonia” were the commonly used terms describing the virus on social media [53].

Videos uploaded from January 1, 2020, to April 30, 2020, were included. Exclusion criteria were videos in languages other than Korean and videos that were duplicated, irrelevant, or livestreamed. Date of upload; source; number of views, likes, dislikes, and comments; the view ratio (number of views / days); and the likes ratio (likes ∗ 100 / [likes + dislikes]) as of May 1, 2020, were collected.

As shown in Figure 1, a total of 1,610,865 videos were obtained. Of the 200 most widely viewed videos, 95 videos were excluded based on the exclusion criteria. A total of 105 videos with 126,633,036 views were included in the study. Of the 105 included videos, 47 (44.76%) were from independent users, 39 (37.14%) were from news agencies, and 11 (10.48%) were from health care professionals. Government agencies contributed only 8 (7.62%) of the total videos.

Figure 1. Data collection flow.



Ethics Statement

This study was reviewed and approved as exempt research by the corresponding author's Institutional Review Board (CR-20-102-L).

Assessment of Popularity

The VPI was used to assess video popularity [10]. The formula for VPI is the (ratio of likes * ratio of views / 100). The view ratio (number of views / days); and the likes ratio (likes ∗ 100 / [likes + dislikes]) as of May 1, 2020, were assessed.

Categorization of Source

Videos were categorized based on their source, which comprised of government agencies, news agencies, health care professionals, and independent users. Government agencies include the Korea Centers for Disease Control and Prevention (KCDC), central disaster management headquarters, Cheongwadae (the executive office and official residence of the South Korean president), the regional health departments, medical associations, academic institutions, and hospitals. News agencies include news clips from newspapers or a broadcast television network. Health care professionals include physicians, nurses, pharmacists, or other health care providers, who do not represent the opinion of their affiliation. Independent users are individuals who are not health care professionals and have no established affiliation, and we included YouTubers from both Korea and other countries who made Korean-language videos.

Assessment of Usefulness

Videos were classified as useful or misleading, and useful videos were defined as those with scientifically correct information. This study's authors are physicians who have provided patient care in Korea during the COVID-19 pandemic and are aware of the uncertainty regarding the disease; therefore, KCDC guidelines as of May 1, 2020, were used as a standard to determine whether a video's content was scientifically correct. Misleading videos were defined as those with scientifically unproven information, and we considered a video misleading if it contained any misleading information at all because it still had the potential to disseminate misinformation. The complete list of misleading information is available in Textbox 1.

Textbox 1. Detailed statements about misleading videos.

- The coronavirus disease (COVID-19) outbreak is not a concern for the public.
- COVID-19 does not exist, and it is a hoax perpetrated to make money.
- COVID-19 has been weaponized.
- Conspiracy theories are concerned with where it originated, how it spread, and who was responsible for it.
- Authorities are hiding the truth.
- COVID-19 was manufactured with the goal of human depopulation.
- One type of food or supplement is recommended over another as a cure or for the prevention of COVID-19.
- COVID-19 has occurred as part of nature's revenge on humankind for cruelty to wild animals.
- People are recommended to not follow the guidelines of the Korea Centers for Disease Control and Prevention.
- People are recommended to follow scientifically unproven measures.
- The COVID-19 pandemic was prophesied by clairvoyants.
- COVID-19 was started to insert microchips into people under the guise of a vaccine.
- COVID-19 was created to build a global surveillance system.

Assessment of Reliability

Videos classified as useful were further analyzed for reliability, overall quality, title–content consistency, and content. There is still no consensus on how to evaluate medical information contained in videos; therefore, we adopted the evaluation tools commonly used in previous studies on online consumer health information, such as tools for websites (JAMAS, GQS) and written patient information on treatment choices (DISCERN) [10,27,36,39,41-43]. The reliability of videos was determined using both the modified DISCERN index and the JAMAS benchmark criteria. DISCERN assesses clarity, reliability, bias, reference supplementation, and areas of uncertainty [36,42]. One point is awarded for each criterion, and the maximum total score is 5 points. JAMAS benchmark criteria consist of authorship, attribution, disclosure, and current status [10,42,43]. One point is awarded for each criterion, and the maximum total score is 4 points. The full list of the modified DISCERN index and JAMAS benchmark criteria is available in Multimedia Appendices 1 and 2, respectively.

Assessment of Overall Quality

The overall quality of the videos was determined by the GQS, which is a 5-point Likert scale ranging from 1 (poor quality) to 5 (high quality) [10,41,43]. It consists of the flow of information, ease of use, and usefulness. The full list of GQS criteria is available in Multimedia Appendix 3.

Assessment of Title–Content Consistency

We created a novel scoring system for assessing title–content consistency because there was no validated evaluation tool for this. The title–content consistency index (TCCI) is a 5-point Likert scale ranging from 1 (poor consistency) to 5 (high consistency), which rates the sensationalist style of a video, that is, the gap between title and content. This index was developed based on previous research on junk news and clickbait [53,54]. The full list of TCCI is available in Multimedia Appendix 4.

Content Analysis

We used the MICI for content analysis [27,36]. MICI is a 5-point Likert scale to assess five components of medical information: (1) prevalence, (2) transmission and precautionary measures, (3) signs and symptoms, (4) testing, and (5) treatment and outcome. The full list of MICI is available in Multimedia Appendix 5.

Statistical Analysis

All videos were reviewed and evaluated by two independent authors. Discrepancies between the authors were resolved by discussion. Interreviewer agreement for categorical variables was analyzed using the Cohen kappa coefficient. Interreviewer agreement for continuous variables was analyzed using intraclass correlation coefficient (ICC) estimates and their 95% confidence intervals based on average measures, absolute agreement, and two-way random model. The kappa coefficients of agreement regarding the classification of source and the usefulness of the videos were k=1 in both cases, indicating full agreement between the authors. ICCs regarding DISCERN, JAMAS, GQS, TCCI, and MICI were 0.47 (95% CI –0.29 to 0.81), 0.93 (95% CI 0.80 to 0.98), 0.71 (95% CI 0.17 to 0.90), 0.89 (95% CI 0.70 to 0.96), and 0.91 (95% CI 0.73 to 0.97), respectively. The Shapiro-Wilk test was used to assess the normality of data. Normally distributed continuous variables were presented as mean (SD). Nonnormally distributed continuous variables were presented as median (IQR). Categorical variables were compared by chi-square test or Fisher exact tests and presented as a number (percentage). Student *t* test for continuous variables and chi-square test for categorical variables were used to compare misleading and useful videos. Kruskal-Wallis test and Mann-Whitney test with Bonferroni correction were used to compare the four sources. All analyses were conducted with the R statistical package version 3.6.3 (R Foundation for Statistical Computing). A *P* value<.05 was considered statistically significant.

Results

Misleading Versus Useful Video

The characteristics of misleading and useful videos are contained in Table 1. Of the total videos, 37.14% (39/105) contained misleading information. Independent user–generated videos showed the highest proportion of misleading information at 68.09% (32/47), while all the government-generated videos were useful. The mean number of likes was 1.47 times higher in misleading videos (18,266 vs 12,389, *P*=.03). The mean number of comments was 1.42 times higher in useful videos (2203 vs 3224, *P*=.02). Misleading videos had almost twice the running time than useful videos (795 seconds vs 405 seconds, *P*=.03). There was no significant difference in the mean views, dislikes, and VPI between the two groups (*P*=.11, *P*=.08, *P*=.31, respectively).

Table 1. The characteristics of misleading and useful videos.[a]

Variables[b]	Misleading videos	Useful videos	Total	*P* value
Videos, n (%)	39 (37.14)	66 (62.86)	105 (100)	N/A[c]
Video metrics, mean (SD)				
Views	972,020 (800,267)	1,344,307 (1,537,507)	1,206,029 (1,320,654)	.11
Likes	18,266 (12,772)	12,389 (13,519)	14572 (13,490)	.03
Dislikes	774 (656)	545 (622)	630 (641)	.08
Comments	2203 (1532)	3224 (2983)	2845 (2581)	.02
Days	79 (22)	70 (27)	73 (26)	.09
Length (seconds)	795 (1046)	405 (340)	550 (713)	.03
VPI[d]	67,708 (159,243)	129,469 (443,911)	106,529 (365,137)	.31
Source, n (%)				
Independent users	32 (68.09)	15 (31.91)	47 (44.76)	N/A
News agencies	4 (10.26)	35 (89.74)	39 (37.14)	N/A
Health care professionals	3 (27.27)	8 (72.73)	11 (10.48)	N/A
Government agencies	0 (0.00)	8 (100)	8 (7.62)	N/A

[a]Student *t* test was used to compare misleading and useful videos.

[b]Continuous variables were presented as mean (SD), and categorical variables were presented as n (%).

[c]N/A: not applicable.

[d]VPI: video power index.

Video Metrics, Reliability, Overall Quality, and Title–Content Consistency of the Useful Videos by Source

Video metrics, reliability, overall quality, title–content consistency, and content of the useful videos by source are presented in Table 2. The distribution of views, likes, comments, and length was significantly different across the four sources (independent users: *P*=.04, news agencies: *P*=.005, health care professionals: *P*=.03, and government agencies: *P*=.002). Videos by government agencies had the shortest median running time of 41 seconds, with the highest views. VPI, as a measurement of popularity, was not significantly different among sources. Government agency–generated videos achieved the highest median score of DISCERN (5.0, IQR 5.0-5.0), JAMAS (4.0, IQR 4.0-4.0), and GQS (4.0, IQR 3.0-4.5), while independent user–generated videos achieved the lowest median score of DISCERN (2.0, IQR 1.0-3.0), JAMAS (2.0, IQR 1.5-2.0), and GQS (2.0, IQR 1.5-2.0). These differences were statistically significant (*P*<.001, *P*<.001, *P*<.001, respectively). The median scores of TCCI were 3.0 (IQR 3.0-4.0) in independent users, 5.0 (IQR 4.0-5.0) in news agencies, 3.0 (IQR 2.0-4.0) in health care professionals, and 5.0 (IQR 5.0-5.0) in governmental agencies (*P*<.001).

Table 2. Video metrics, reliability, overall quality, title–content consistency, and content of the useful videos by source.

Variables[a]	Independent users	News agencies	Health care professionals	Government agencies	Total	*P* value[b]
Videos, n (%)	15 (22.73)	35 (55.03)	8 (12.12)	8 (12.12)	66 (100.00)	N/A[c]
Video metrics, median (IQR)						
Views	744,824 (553,149-1,050,672)	906,731 (768,863-1,282,426)	928,563 (768,613-1,083,527)	2,418,742 (856,649-3,896,445)	888,772 (726,536-1,292,581)	.04
Likes	9643 (5063-18,041)	7523 (4727-12,831)	20,874 (14,977-27,702)	1037 (157-15,169)	9756 (4598-16,432)	.005
Dislikes	348 (190-494)	305 (195-469)	789 (396-894)	104 (19-1280)	355 (179-698)	.10
Comments	1251 (779-2516)	2938 (1810-5570)	2031 (1055-3242)	179 (35-3921)	2377 (1251-4151)	.03
Days	56 (32-84)	85 (47-98)	72 (67-94)	74 (54-82)	75 (49-96)	.22
Length (seconds)	422 (246-518)	207 (165-536)	791 (371-822)	41 (36-176)	249 (162-602)	.002
VPI[d]	15,689 (5786-112,775)	22,326 (4568-50,043)	33,709 (16,773-54,613)	9652 (755-208,260)	18,092 (5780-56,760)	.54
Reliability, median (IQR)						
DISCERN	2.0 (1.0-3.0)	4.0 (4.0-5.0)	4.5 (3.0-5.0)	5.0 (5.0-5.0)	4.0 (3.0-5.0)	<.001
JAMAS[e]	2.0 (1.5-2.0)	3.0 (3.0-4.0)	2.5 (2.0-3.0)	4.0 (4.0-4.0)	3.0 (2.0-4.0)	<.001
Overall quality, median (IQR)						
GQS[f]	2.0 (1.5-2.0)	3.0 (2.0-3.0)	3.5 (2.5-4.5)	4.0 (3.0-4.5)	3.0 (2.0-3.0)	<.001
Title-content consistency, median (IQR)						
TCCI[g]	3.0 (3.0-4.0)	5.0 (4.0-5.0)	3.0 (2.0-4.0)	5.0 (5.0-5.0)	5.0 (3.0-5.0)	<.001
Content						
Frequency, n (%)						
Prevalence	10 (66.67)	27 (77.14)	5 (62.50)	2 (25)	44 (66.67)	N/A
Transmission and precautionary measures	11 (73.33)	23 (65.71)	7 (87.50)	8 (100)	49 (74.24)	N/A
Signs and symptoms	4 (26.67)	17 (48.57)	3 (37.50)	7 (87.50)	31 (46.97)	N/A
Testing	7 (46.67)	15 (42.86)	4 (50)	4 (50)	30 (45.45)	N/A
Treatment and outcome	3 (20)	16 (45.71)	4 (50)	1 (12.50)	24 (36.36)	N/A
Total score of MICI[h], median (IQR)	5.0 (3.0-7.0)	5.0 (3.5-6.5)	7.0 (2.5-8.0)	5.0 (5.0-9.0)	5.0 (3.00-7.0)	.77
Individual scores of the MICI components, median (IQR)						
Prevalence	1.0 (0.0-1.0)	1.0 (1.0-1.0)	1.0 (0.0-2.5)	0.0 (0.0-0.5)	1.0 (0.0-1.0)	.18
Transmission and precautionary measures	2.0 (0.5-3.0)	2.0 (0.0-3.0)	2.5 (1.5-3.5)	3.0 (2.0-3.5)	2.0 (0.0-3.0)	.23
Signs and symptoms	0.0 (0.0-1.0)	0.0 (0.0-2.0)	0.0 (0.0-2.0)	2.0 (2.0-2.5)	0.0 (0.0-2.0)	.03
Testing	0.0 (0.0-2.0)	0.0 (0.0-2.0)	0.5 (0.0-1.0)	1.0 (0.0-2.5)	0.0 (0.0-2.0)	.86
Treatment and outcome	0.0 (0.0-0.0)	0.0 (0.0-2.0)	0.5 (0.0-2.0)	0.0 (0.0-0.0)	0.0 (0.0-1.0)	.23

[a]Continuous variables were presented as median (IQR), and categorical variables were presented as n (%).

[b]Kruskal-Wallis tests were used to calculate *P* values.

[c]N/A: not applicable.

[d]VPI: video power index.

[e]JAMAS: Journal of the American Medical Association Score.

[f]GQS: global quality score.

[g]TCCI: title–content consistency index.

[h]MICI: medical information and content index.

Content Analysis of the Useful Videos by Source

Of the useful videos, 74.24% (49/105) provided information on "transmission and precautionary measures," 66.67% (44/105) contained information on "prevalence," 46.97% (31/105) contained "signs and symptoms," 45.45% (30/105) contained "testing," and 36.36% (24/105) contained "treatment and outcome." "Transmission and precautionary measures" were the most discussed topic by government agencies, news agencies, and independent users. Every video by government agencies covered "transmission and precautionary measures," and 7 out of 8 (87.50%) videos mentioned "signs and symptoms." On the other hand, the most mentioned topic by news agencies was "prevalence," followed by "transmission and precautionary measures." The total score of MICI was not significantly different among sources (P=.77). The highest median score among individual MICI components was shown in "transmission and precautionary measures" of government agency–generated videos (3.0, IQR 2.0-3.5).

Discussion

Principal Findings

This study is the first of its kind to evaluate the Korean-language content of COVID-19 videos on YouTube. Previous COVID-19 YouTube studies have captured videos using a relevant filter at the time of the search or have been mainly descriptive [36,52]. In this study, we analyzed the 200 most popular videos uploaded between January 1, 2020, and April 30, 2020, which comprised 126,633,036 views. We conducted content analysis and assessed the reliability, overall quality, and title–content consistency of the videos.

One must be cautious when labeling content as misinformation; however, the majority of YouTube studies in the field of emerging infectious disease did not show detailed criteria or specific examples of misleading videos. Some of them showed two or three examples of conspiracy theories, and others merely mentioned that they classified videos as misleading if they conveyed at least one scientifically unproven piece of information [35,55,56]. In contrast, one study on the Ebola virus provided specific examples of misleading videos [46]. COVID-19–related YouTube studies are not unlike previous studies on other infectious diseases, and one briefly mentioned that they reviewed published references as the standard for what is known about COVID-19 [57]. Notably, a study performed by Li et al [44] provided statements recorded from YouTube videos classified as misleading. Li et al [44] also created a novel five-point scoring system to assess the usefulness of a video. However, this score is not designed to distinguish misleading videos from useful ones but to measure how much of the video content is useful. Unlike previous studies, we set clear criteria to distinguish misleading from useful videos and provided a complete list of misleading information in Textbox 1.

Out of 105 videos, 39 (37.14%) were found to be misleading. This percentage is higher than that from a previous study that evaluated English-language videos addressing COVID-19 (19/69, 27.5%) [44]. Useful videos did not exceed misleading ones in popularity, which suggests that the chance of a layperson being exposed to inaccurate information is quite high. Unfortunately, fake news spreads six times faster than verified news and receives higher viewer interaction [58,59]. Biggs et al [37] reported that misleading videos are more viewed than useful videos because the useful ones have longer running times; however, recent studies on COVID-19–related YouTube videos have returned inconsistent results. Although useful videos gained more views than misleading ones in this study, the difference was not statistically significant (1,344,307 vs 972,020, P=.11). Furthermore, useful videos earned more comments but fewer likes than misleading videos (3224 vs 2203, P=.02 and 12,389 vs 18,266, P=.03, respectively). Previously, there was an attempt to compare Mandarin videos regarding COVID-19 to English ones. They reported that misleading Mandarin videos gained more views than useful ones, but the result was the opposite for English videos (Mandarin: 91,949 in useful videos, 151,868 in misleading videos, P=.30; English: 288,545 in useful videos, 1621 in misleading videos, P<.001) [36]. Another study on English-language COVID-19–related videos found that there were no significant differences in views, likes, and dislikes between useful and misleading videos (P=.50, P=.79, P=.10, respectively) [44].

In this study, most of the misleading information was delivered by independent users (32/39, 82.05%). Moreover, some of these videos generated a lot of interaction from viewers. For example, one video suggested that because of COVID-19, a microchip will be inserted into people under the guise of a vaccine to build a global surveillance system; this video gained 330,672 views with 2454 comments. Another video mentioned a conspiracy theory that COVID-19 is a biological weapon, and it was manufactured for human depopulation. Another video with 1,478,262 views claimed that COVID-19 was predicted in works of fiction or in movies. One video posted by a shaman was entirely misleading. It consisted of question-answer pairs, and a shaman answered questions about COVID-19 such as, "When will the COVID-19 pandemic end?" and "When will a COVID-19 vaccine be available?"

The mere fact that a video uploader is a doctor or health expert does not imply that their videos provide accurate medical information. There were 3 out of 11 videos posted by health care professionals that were misleading; 2 videos addressed misleading information throughout the entire running time. There was 1 video that alleged the existence of a conspiracy theory as to when COVID-19 originated and who is responsible for the virus. Another video recommended one vitamin supplement as a cure for COVID-19. In contrast, one video posted by a pharmacist provided partially useful information, but it was still classified as misleading since videos containing any misleading information could potentially disseminate

misinformation. This video included helpful information during the first half of the running time, such as characteristics of the viral disease, transmission, hand hygiene, and face masks. However, during the second half, the video recommended several foods such as ginger, onions, green tea, and black beans as immune boosters against COVID-19. Although good nutrition is key to staying healthy, this video was classified as misleading because no food or dietary supplement alone can prevent COVID-19.

Most videos by news agencies in this study provided scientifically accurate information (35/39, 89.74%). Video clips of television-based news were often posted on YouTube, amplifying the impact of traditional media. Given the inclusion criterion for this study, it can be reasoned that consumers may have the social media literacy skills to choose appropriate videos among the hundreds of thousands of videos that can be viewed on YouTube.

Government-generated videos were effective delivery tools. Although they comprised only 7.62% (8/105) of the total videos, they were all useful and gained the highest median views (P=.04). They also had the shortest median running time at 41 seconds (P=.002) and showed higher reliability and overall quality (all P<.001). These findings are consistent with previous studies. A systematic review of health care information on YouTube found that government agency–generated videos had credible information [45]. Similarly, in a study performed by Li et al [44], government videos only contributed 2.89% (2/69) of COVID-19–related English videos, but they contained only useful information and showed higher reliability compared to consumer videos (DISCERN: 4.57 vs 2.12, P=.008; JAMAS: 2.71 vs 1.50, P=.03) [44].

However, credible videos with high quality were not popular. Considering that videos generated by government agencies received the least number of likes and comments (P=.005, P=.03, respectively), they failed to encourage viewer interaction and engagement. They also showed the lowest VPI as a measurement of popularity, but the VPI was not significantly different (P=.54). Government or news agencies were also more likely to post videos with a proper title (median TCCI of 5.0, IQR 4.0-5.0 and 5.0, IQR 5.0–5.0, respectively). In contrast, independent users were more likely to post clickbait videos with sensationalist headlines or eye-catching, attention-grabbing thumbnails with large gaps between the title and their content (median TCCI 3.0, IQR 3.0-4.0). Viewers are more likely to select emotionally appealing titles, regardless of the correctness of the content [60].

Several studies have reported content differences among analyzed COVID-19–related YouTube videos. Basch et al [52] reported that “quarantine and travel restrictions” was the most discussed item in English and Spanish videos (89/89 and 84/89, respectively), and “precautionary measures” was covered in less than one-third (0/100 to 31/100) of the videos. In a study performed by Khatri et al [36], only 10% (2/21) of Mandarin videos covered “testing” compared to 53.19% (25/47) of the English videos [36]. In our study, 45.45% (30/66) of videos covered “testing.” On the other hand, “transmission” was the most mentioned subject in both Mandarin and English videos, which is similar to our finding (49/66, 74.24% in our study; 43/47, 91.49% in English; 17/21, 81% in Mandarin).

Through content analysis, we can understand the characteristics of popular COVID-19–related videos in Korea. As shown in Table 2, the most common content was “transmission and precautionary measures.” Even all the videos published by the government agencies covered “transmission and precautionary measures.” Various personal protective strategies were emphasized in these videos, including washing hands, wearing a mask, maintaining a distance of 1-2 meters, staying home when sick, and avoiding gatherings. These strategies may encourage people to practice preventive behaviors such as personal hygiene and social distancing. Asymptomatic carriers were also mentioned in these videos, and the comments on these videos show that people share similar concerns. For example:

> *I am young and healthy. It looks like I will probably be okay, but what if I spread the virus to my parents without knowing it? They are old and weak. I’ve got to stay at home.*

The second most common content area in Korean videos regarding COVID-19 covered “signs and symptoms.” These videos may encourage people who develop suspicious symptoms to be screened as soon as possible. Early detection of the symptoms of COVID-19 may enable people to visit the hospital early on, improving treatment responses and leading to decreased mortality.

Furthermore, videos frequently mentioned call centers run by KCDC or the regional health department. In addition, they recommended having a consultation from the call center first before visiting a hospital if an individual had suspicious symptoms or had come in contact with a patient with COVID-19. This could prevent the spread of the virus from patients who are infected to health care providers.

Of the videos uploaded by independent users, two were “patient experience” videos. One video is filmed by a patient lying in a hospital bed who has COVID-19. The patient, who was having difficulty breathing, was trying to talk to the viewers about not taking any risks regarding COVID-19. The other video includes a survivor of COVID-19 who shares his personal story and experience with viewers. Viewers shared their opinions through the comments under these videos.

> *Thank you for sharing your story. I hope you will get better soon.*
>
> *It looks like so much pain. It is so sad and scary.*
>
> *When you go outside, please wear a mask. It is for our family and friends.*

Science is not supposed to be a popularity contest, but governments should exert more effort to disseminate accurate and complete information via social media to ameliorate the negative health consequences of misinformation. Peer-reviewed or expert-approved videos are expected to provide credible medical information [61]. However, only a few of them that were uploaded by government agencies, universities, hospitals, and medical associations were included in this study because they did not rank within the 200 most viewed videos. The efforts of health care professionals cannot efficiently compete with

malicious people and bots, who are able to automatically post millions of messages [62,63].

Limitations

This study includes several limitations. First, this is a cross-sectional study, so it is limited to capturing YouTube scenes at a certain point in time. Common search terms or the most viewed videos may change over time, and longitudinal changes in video metrics such as views, likes, dislikes, or comments were not captured. Second, there is no validated tool for evaluating video-based medical information. Therefore, we adopted the evaluation tools commonly used in previous studies on online consumer health information. A comprehensive evaluation tool for the medical content of videos needs to be developed. Third, we could not tell whether watching a video clip led to a change in health behavior. For example, some people enjoy watching conspiracy videos to pass time and may be able to distinguish which stories are valid. Thus, they may end up following the guidelines of authorities regardless of their viewing such videos. More research is needed to evaluate the relationship between exposure to misinformation and health behaviors. Fourth, we did not collect any data on the viewers. All the YouTube video metrics were collected anonymously, so we could not grasp the viewers' demographic characteristics.

Conclusions

Misleading videos had more likes, fewer comments, and longer running times than useful videos. Korean-language YouTube videos on COVID-19 uploaded by different sources varied significantly in terms of reliability, overall quality, and title–content consistency, but the content coverage was not significantly different. Government-generated videos had higher reliability, overall quality, and title–content consistency than independent user–generated videos.

Although there is concern about the spread of misinformation via YouTube, the educational value of this website cannot be ignored. YouTube can be a powerful tool to keep the public informed during a crisis in a controlled and reassuring manner. However, to do so, accurate information must be made available on such platforms. Therefore, governments should have a stronger presence on social media and produce more online videos to reach a wider audience. First, to accomplish this, policy makers should support health institutions financially so they can use social networking platforms to their full potential. For example, an educational program could be developed to teach health care providers how to make YouTube videos and engage with the audience on social media. Second, health care professionals should cooperate with social media influencers, not compete with them, to reach more people. For example, the top-10 most subscribed YouTubers could create and upload a video in collaboration with a physician guest that provides a combination of entertainment and information on COVID-19. Thus, YouTube users would be able to obtain more high-quality information than false data. Third, compelling content should be used, as it can grab viewers' attention. For example, visually attractive recordings can receive more views, and emotionally persuasive videos with exemplar stories are more likely to catch viewers' attention than those with statistical evidence [64]. In this regard, effective communication via YouTube would contribute to reducing the risk of undesirable behavior, such as panic buying, and help the public distinguish between valid information and fake news. YouTube could serve as a rapid and inexpensive platform for reaching more people with accurate information during a public health crisis.

Authors' Contributions

HM and GHL designed the study, coded the videos, and conducted statistical analyses. HM collected the data and wrote the manuscript. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Modified DISCERN index.
[DOC File , 32 KB - jmir_v22i8e20775_app1.doc]

Multimedia Appendix 2
The Journal of the American Medical Association Score benchmark criteria.
[DOC File , 33 KB - jmir_v22i8e20775_app2.doc]

Multimedia Appendix 3
Global quality score.
[DOC File , 33 KB - jmir_v22i8e20775_app3.doc]

Multimedia Appendix 4
Title–content consistency index.
[DOC File , 33 KB - jmir_v22i8e20775_app4.doc]

Multimedia Appendix 5
Medical information and content index.
[DOC File , 51 KB - jmir_v22i8e20775_app5.doc]

References

1. Oh J, Lee J, Schwarz D, Ratcliffe HL, Markuns JF, Hirschhorn LR. National response to COVID-19 in the Republic of Korea and lessons learned for other countries. Health Syst Reform 2020 Jan 01;6(1):e1753464. [doi: 10.1080/23288604.2020.1753464] [Medline: 32347772]
2. Smartphone usage in South Korea - statistics & facts. Statista. URL: https://www.statista.com/topics/5340/smartphone-usage-in-south-korea/
3. Jin-young C. YouTube ranks 1st in S. Korea for number of users. Businesskorea. 2018 Jan 05. URL: http://www.businesskorea.co.kr/news/articleView.html?idxno=20222 [accessed 2020-05-17]
4. Fox S. The social life of health information. Pew Research Center. 2014 Jan 15. URL: http://pewrsr.ch/1hYp7Oy
5. Guttmann A. Social media platforms used by marketers worldwide 2019. Statista. URL: https://www.statista.com/statistics/259379/social-media-platforms-used-by-marketers-worldwide/
6. Welbourne DJ, Grant WJ. Science communication on YouTube: factors that affect channel and video popularity. Public Underst Sci 2016 Aug;25(6):706-718. [doi: 10.1177/0963662515572068] [Medline: 25698225]
7. Rizoiu MA, Xie L, Sanner S, Cebrian M, Yu H, Van Hentenryck P. Expecting to be HIP: Hawkes intensity processes for social media popularity. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Apr Presented at: WWW '17: 26th International World Wide Web Conference; April 2017; Perth Australia p. 735-744. [doi: 10.1145/3038912.3052650]
8. Wu J, Zhou Y, Chiu DM, Zhu Z. Modeling dynamics of online video popularity. IEEE Trans Multimedia 2016 Sep;18(9):1882-1895. [doi: 10.1109/tmm.2016.2579600]
9. Borghol Y, Mitra S, Ardon S, Carlsson N, Eager D, Mahanti A. Characterizing and modelling popularity of user-generated videos. Performance Eval 2011 Nov;68(11):1037-1055. [doi: 10.1016/j.peva.2011.07.008]
10. Erdem MN, Karaca S. Evaluating the accuracy and quality of the information in kyphosis videos shared on YouTube. Spine 2018 Nov 15;43(22):E1334-E1339. [doi: 10.1097/BRS.0000000000002691] [Medline: 29664816]
11. Barros JM, Duggan J, Rebholz-Schuhmann D. The application of internet-based sources for public health surveillance (infoveillance): systematic review. J Med Internet Res 2020 Mar 13;22(3):e13680 [FREE Full text] [doi: 10.2196/13680] [Medline: 32167477]
12. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
13. Alvarez-Mon MA, Llavero-Valero M, Sánchez-Bayona R, Pereira-Sanchez V, Vallejo-Valdivielso M, Monserrat J, et al. Areas of interest and stigmatic attitudes of the general public in five relevant medical conditions: thematic and quantitative analysis using Twitter. J Med Internet Res 2019 May 28;21(5):e14110 [FREE Full text] [doi: 10.2196/14110] [Medline: 31140438]
14. Viguria I, Alvarez-Mon M, Llavero-Valero M, Asunsolo Del Barco A, Ortuño F, Alvarez-Mon M. Eating disorder awareness campaigns: thematic and quantitative analysis using Twitter. J Med Internet Res 2020 Jul 14;22(7):e17626 [FREE Full text] [doi: 10.2196/17626] [Medline: 32673225]
15. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
16. Williams SA, Terras M, Warwick C. How Twitter is studied in the medical professions: a classification of Twitter papers indexed in PubMed. Med 2 0 2013;2(2):e2 [FREE Full text] [doi: 10.2196/med20.2269] [Medline: 25075237]
17. Milinovich GJ, Avril SMR, Clements ACA, Brownstein JS, Tong S, Hu W. Using internet search queries for infectious disease surveillance: screening diseases for suitability. BMC Infect Dis 2014 Dec 31;14:690 [FREE Full text] [doi: 10.1186/s12879-014-0690-1] [Medline: 25551277]
18. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
19. Majumder MS, Santillana M, Mekaru SR, McGinnis DP, Khan K, Brownstein JS. Utilizing nontraditional data sources for near real-time estimation of transmission dynamics during the 2015-2016 Colombian Zika virus disease outbreak. JMIR Public Health Surveill 2016 Jun 01;2(1):e30 [FREE Full text] [doi: 10.2196/publichealth.5814] [Medline: 27251981]
20. Cacheda F, Fernandez D, Novoa FJ, Carneiro V. Early detection of depression: social network analysis and random forest techniques. J Med Internet Res 2019 Jun 10;21(6):e12554 [FREE Full text] [doi: 10.2196/12554] [Medline: 31199323]
21. Ramírez-Cifuentes D, Freire A, Baeza-Yates R, Puntí J, Medina-Bravo P, Velazquez D, et al. Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis. J Med Internet Res 2020 Jul 07;22(7):e17758 [FREE Full text] [doi: 10.2196/17758] [Medline: 32673256]
22. Adhikari J, Sharma P, Arjyal L, Uprety D. YouTube as a source of information on cervical cancer. N Am J Med Sci 2016 Apr;8(4):183-186 [FREE Full text] [doi: 10.4103/1947-2714.179940] [Medline: 27213142]

23. Knösel M, Jung K. Informational value and bias of videos related to orthodontics screened on a video-sharing Web site. Angle Orthod 2011 May;81(3):532-539. [doi: 10.2319/091710-541.1] [Medline: 21261492]
24. Baker DM, Marshall JH, Lee MJ, Jones GL, Brown SR, Lobo AJ. YouTube as a source of information for patients considering surgery for ulcerative colitis. J Surg Res 2017 Dec;220:133-138. [doi: 10.1016/j.jss.2017.06.094] [Medline: 29180175]
25. Basch CH, Yin J, Kollia B, Adedokun A, Trusty S, Yeboah F, et al. Public online information about tinnitus: a cross-sectional study of YouTube videos. Noise Health 2018;20(92):1-8 [FREE Full text] [doi: 10.4103/nah.NAH_32_17] [Medline: 29457600]
26. Singh SK, Liu S, Capasso R, Kern RC, Gouveia CJ. YouTube as a source of information for obstructive sleep apnea. Am J Otolaryngol 2018;39(4):378-382. [doi: 10.1016/j.amjoto.2018.03.024] [Medline: 29605236]
27. Nagpal SJS, Karimianpour A, Mukhija D, Mohan D, Brateanu A. YouTube videos as a source of medical information during the Ebola hemorrhagic fever epidemic. Springerplus 2015;4:457 [FREE Full text] [doi: 10.1186/s40064-015-1251-9] [Medline: 26322263]
28. Park EW. Medical jargon used in health care communication of family physician. Korean J Fam Med 2010;31(6):453. [doi: 10.4082/kjfm.2010.31.6.453]
29. Links AR, Callon W, Wasserman C, Walsh J, Beach MC, Boss EF. Surgeon use of medical jargon with parents in the outpatient setting. Patient Educ Couns 2019 Jun;102(6):1111-1118 [FREE Full text] [doi: 10.1016/j.pec.2019.02.002] [Medline: 30744965]
30. Pariser E. The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think. London: Penguin; 2011.
31. Törnberg P. Echo chambers and viral misinformation: modeling fake news as complex contagion. PLoS One 2018;13(9):e0203958 [FREE Full text] [doi: 10.1371/journal.pone.0203958] [Medline: 30235239]
32. Johnson SB, Park HS, Gross CP, Yu JB. Complementary medicine, refusal of conventional cancer therapy, and survival among patients with curable cancers. JAMA Oncol 2018 Oct 01;4(10):1375-1381 [FREE Full text] [doi: 10.1001/jamaoncol.2018.2487] [Medline: 30027204]
33. Hussain A, Ali S, Ahmed M, Hussain S. The anti-vaccination movement: a regression in modern medicine. Cureus 2018 Jul 03;10(7):e2919 [FREE Full text] [doi: 10.7759/cureus.2919] [Medline: 30186724]
34. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis--a wakeup call? J Rheumatol 2012 May;39(5):899-903. [doi: 10.3899/jrheum.111114] [Medline: 22467934]
35. Pandey A, Patni N, Singh M, Sood A, Singh G. YouTube as a source of information on the H1N1 influenza pandemic. Am J Prev Med 2010 Mar;38(3):e1-e3. [doi: 10.1016/j.amepre.2009.11.007] [Medline: 20171526]
36. Khatri P, Singh S, Belani N, Yeong Y, Lohan R, Lim Y, et al. YouTube as source of information on 2019 novel coronavirus outbreak: a cross sectional study of English and Mandarin content. Travel Med Infect Dis 2020;35:101636 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101636] [Medline: 32205267]
37. Biggs TC, Bird JH, Harries PG, Salib RJ. YouTube as a source of information on rhinosinusitis: the good, the bad and the ugly. J Laryngol Otol 2013 Aug;127(8):749-754. [doi: 10.1017/S0022215113001473] [Medline: 23866821]
38. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
39. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the internet. JAMA 1997 Apr 16;277(15):1244. [doi: 10.1001/jama.1997.03540390074039]
40. Boyer C, Gaudinat A, Baujard V, Geissbühler A. Health on the Net Foundation: assessing the quality of health web pages all over the world. Stud Health Technol Inform 2007;129(Pt 2):1017-1021. [Medline: 17911869]
41. Langille M, Bernard A, Rodgers C, Hughes S, Leddin D, van Zanten SV. Systematic review of the quality of patient information on the internet regarding inflammatory bowel disease treatments. Clin Gastroenterol Hepatol 2010 Apr;8(4):322-328. [doi: 10.1016/j.cgh.2009.12.024] [Medline: 20060070]
42. Radonjic A, Fat Hing NN, Harlock J, Naji F. YouTube as a source of patient information for abdominal aortic aneurysms. J Vasc Surg 2020 Feb;71(2):637-644. [doi: 10.1016/j.jvs.2019.08.230] [Medline: 31611104]
43. Ferhatoglu MF, Kartal A, Filiz A, Kebudi A. Comparison of New Era's education platforms, YouTube® and WebSurg®, in sleeve gastrectomy. Obes Surg 2019 Nov;29(11):3472-3477. [doi: 10.1007/s11695-019-04008-x] [Medline: 31172453]
44. Li HO, Bailey A, Huynh D, Chan J. YouTube as a source of information on COVID-19: a pandemic of misinformation? BMJ Glob Health 2020 May;5(5):e002604 [FREE Full text] [doi: 10.1136/bmjgh-2020-002604] [Medline: 32409327]
45. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015 Sep;21(3):173-194. [doi: 10.1177/1460458213512220] [Medline: 24670899]
46. Pathak R, Poudel DR, Karmacharya P, Pathak A, Aryal MR, Mahmood M, et al. YouTube as a source of information on Ebola virus disease. N Am J Med Sci 2015 Jul;7(7):306-309 [FREE Full text] [doi: 10.4103/1947-2714.161244] [Medline: 26258077]
47. Kuruppu GN, De Zoysa A. COVID-19 and panic buying: an examination of the impact of behavioural biases. SSRN J 2020 May 13:A. [doi: 10.2139/ssrn.3596101]

48. Basch CH, Fung IC, Hammond RN, Blankenship EB, Tse ZTH, Fu K, et al. Zika virus on YouTube: an analysis of English-language video content by source. J Prev Med Public Health 2017 Mar;50(2):133-140. [doi: 10.3961/jpmph.16.107] [Medline: 28372356]
49. Basch CH, Basch CE, Ruggles KV, Hammond R. Coverage of the Ebola virus disease epidemic on YouTube. Disaster Med Public Health Prep 2015 Oct;9(5):531-535. [doi: 10.1017/dmp.2015.77] [Medline: 26088275]
50. Li S, Wang Y, Xue J, Zhao N, Zhu T. The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users. Int J Environ Res Public Health 2020 Mar 19;17(6):2032 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]
51. Su Y, Xue J, Liu X, Wu P, Chen J, Chen C, et al. Examining the impact of COVID-19 lockdown in Wuhan and Lombardy: a psycholinguistic analysis on Weibo and Twitter. Int J Environ Res Public Health 2020 Jun 24;17(12):4552 [FREE Full text] [doi: 10.3390/ijerph17124552] [Medline: 32599811]
52. Basch C, Hillyer G, Meleo-Erwin Z, Jaime C, Mohlman J, Basch C. Preventive behaviors conveyed on YouTube to mitigate transmission of COVID-19: cross-sectional study. JMIR Public Health Surveill 2020 Apr 02;6(2):e18807 [FREE Full text] [doi: 10.2196/18807] [Medline: 32240096]
53. Narayanan V, Barash V, Kelly J, Kollanyi B, Neudert L, Howard P. Polarization, partisanship and junk news consumption over social media in the US. The Computational Propaganda Project. 2018. URL: https://comprop.oii.ox.ac.uk/research/polarization-partisanship-and-junk-news/
54. Bolton DM, Yaxley J. Fake news and clickbait - natural enemies of evidence-based medicine. BJU Int 2017 May;119 Suppl 5:8-9. [doi: 10.1111/bju.13883] [Medline: 28544296]
55. Dubey D, Amritphale A, Sawhney A, Dubey D, Srivastav N. Analysis of YouTube as a source of information for West Nile Virus infection. Clin Med Res 2014 Dec;12(3-4):129-132 [FREE Full text] [doi: 10.3121/cmr.2013.1194] [Medline: 24573700]
56. Gonzalez-Estrada A, Cuervo-Pardo L, Ghosh B, Smith M, Pazheri F, Zell K, et al. Popular on YouTube: a critical appraisal of the educational quality of information regarding asthma. Allergy Asthma Proc 2015;36(6):e121-e126. [doi: 10.2500/aap.2015.36.3890] [Medline: 26534743]
57. D'Souza RS, D'Souza S, Strand N, Anderson A, Vogt MNP, Olatoye O. YouTube as a source of medical information on the novel coronavirus 2019 disease (COVID-19) pandemic. Glob Public Health 2020 Jul;15(7):935-942. [doi: 10.1080/17441692.2020.1761426] [Medline: 32397870]
58. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
59. Burger P, Kanhai S, Pleijter A, Verberne S. The reach of commercially motivated junk news on Facebook. PLoS One 2019;14(8):e0220446 [FREE Full text] [doi: 10.1371/journal.pone.0220446] [Medline: 31369596]
60. Kim HS. Attracting views and going viral: how message features and news-sharing channels affect health news diffusion. J Commun 2015 Jun 01;65(3):512-534 [FREE Full text] [doi: 10.1111/jcom.12160] [Medline: 26441472]
61. Bezner SK, Hodgman EI, Diesen DL, Clayton JT, Minkes RK, Langer JC, et al. Pediatric surgery on YouTube™: is the truth out there? J Pediatr Surg 2014 Apr;49(4):586-589. [doi: 10.1016/j.jpedsurg.2013.08.004] [Medline: 24726118]
62. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
63. Jamison AM, Broniatowski DA, Quinn SC. Malicious actors on Twitter: a guide for public health researchers. Am J Public Health 2019 May;109(5):688-692. [doi: 10.2105/AJPH.2019.304969] [Medline: 30896994]
64. Hastall MR, Knobloch-Westerwick S. Severity, efficacy, and evidence type as determinants of health message exposure. Health Commun 2013;28(4):378-388. [doi: 10.1080/10410236.2012.690175] [Medline: 22809248]

Abbreviations

COVID-19: coronavirus disease
GQS: global quality score
ICC: intraclass correlation coefficient
JAMAS: Journal of the American Medical Association Score
KCDC: Korea Centers for Disease Control and Prevention
MICI: medical information and content index
TCCI: title–content consistency index
VPI: video power index

Edited by G Eysenbach; submitted 04.06.20; peer-reviewed by C Basch, D Frohlich; comments to author 25.06.20; revised version received 11.07.20; accepted 27.07.20; published 12.08.20.

Please cite as:
Moon H, Lee GH
Evaluation of Korean-Language COVID-19–Related Medical Information on YouTube: Cross-Sectional Infodemiology Study
J Med Internet Res 2020;22(8):e20775
URL: http://www.jmir.org/2020/8/e20775/
doi:10.2196/20775
PMID:32730221

Original Paper

Natural Language Processing for Rapid Response to Emergent Diseases: Case Study of Calcium Channel Blockers and Hypertension in the COVID-19 Pandemic

Antoine Neuraz[1,2,3], MD; Ivan Lerner[1,2], MD; William Digan[2,4], MSc; Nicolas Paris[5], MSc; Rosy Tsopra[2,4], MD, PhD; Alice Rogier[2,4], MSc; David Baudoin[2,4], MSc; Kevin Bretonnel Cohen[6], PhD; Anita Burgun[1,2,4], MD, PhD; Nicolas Garcelon[2,7], PhD; Bastien Rance[2,4], PhD; AP-HP/Universities/INSERM COVID-19 Research Collaboration; AP-HP COVID CDR Initiative[8]

[1]Department of Biomedical Informatics, Necker-Enfant Malades Hospital, Assistance Publique – Hôpitaux de Paris (AP-HP), Paris, France
[2]Centre de Recherche des Cordeliers, INSERM UMRS 1138 Team 22, Université de Paris, Paris, France
[3]LIMSI, CNRS, Université Paris Saclay, Orsay, France
[4]Department of Biomedical Informatics, Georges Pompidou European Hospital, Assistance Publique – Hôpitaux de Paris (AP-HP), Paris, France
[5]DSI WIND, Assistance Publique – Hôpitaux de Paris (AP-HP), Paris, France
[6]School of Medicine, University of Colorado, Denver, CO, United States
[7]Institut Imagine, INSERM U1163, Université Paris Descartes, Université de Paris, Paris, France
[8]Please see acknowledgements for list of collaborators

Corresponding Author:
Antoine Neuraz, MD
Department of Biomedical Informatics
Necker-Enfant Malades Hospital
Assistance Publique – Hôpitaux de Paris (AP-HP)
Bat Imagine, Bureau 145
149 rue de Sèvres
Paris, 75015
France
Phone: 33 0624622355
Email: antoine.neuraz@aphp.fr

Abstract

Background: A novel disease poses special challenges for informatics solutions. Biomedical informatics relies for the most part on structured data, which require a preexisting data or knowledge model; however, novel diseases do not have preexisting knowledge models. In an emergent epidemic, language processing can enable rapid conversion of unstructured text to a novel knowledge model. However, although this idea has often been suggested, no opportunity has arisen to actually test it in real time. The current coronavirus disease (COVID-19) pandemic presents such an opportunity.

Objective: The aim of this study was to evaluate the added value of information from clinical text in response to emergent diseases using natural language processing (NLP).

Methods: We explored the effects of long-term treatment by calcium channel blockers on the outcomes of COVID-19 infection in patients with high blood pressure during in-patient hospital stays using two sources of information: data available strictly from structured electronic health records (EHRs) and data available through structured EHRs and text mining.

Results: In this multicenter study involving 39 hospitals, text mining increased the statistical power sufficiently to change a negative result for an adjusted hazard ratio to a positive one. Compared to the baseline structured data, the number of patients available for inclusion in the study increased by 2.95 times, the amount of available information on medications increased by 7.2 times, and the amount of additional phenotypic information increased by 11.9 times.

Conclusions: In our study, use of calcium channel blockers was associated with decreased in-hospital mortality in patients with COVID-19 infection. This finding was obtained by quickly adapting an NLP pipeline to the domain of the novel disease; the adapted pipeline still performed sufficiently to extract useful information. When that information was used to supplement existing

structured data, the sample size could be increased sufficiently to see treatment effects that were not previously statistically detectable.

(J Med Internet Res 2020;22(8):e20773) doi:10.2196/20773

KEYWORDS

medication information; natural language processing; electronic health records; COVID-19; public health; response; emergent disease; informatics

Introduction

Outbreaks of novel diseases can create enormous strain on public health systems. Since the time of Snow's pioneering work [1] on the epidemiology of the London cholera outbreak of 1854, it has been clear that information is key to the successful abatement of these substantial public health challenges. Currently, health care systems have access to quantities of data that would have been unimaginable in Snow's time. Because these data are in electronic format, they can be manipulated and exploited rapidly. However, a novel disease poses special challenges for informatics solutions. Biomedical informatics relies for the most part on structured data; structured data require a preexisting data or knowledge model; and a novel disease will not have a preexisting knowledge model. This poses a formidable obstacle to leveraging informatics solutions to address the type of public health crisis the world is facing at the time of writing. One solution to the lack of structured information is natural language processing (NLP).

Biomedical text mining, or the use of textual data, in electronic health records (EHRs) has often been proposed as a method for converting unstructured data to the structured data that is needed in public health informatics. One of the advantages of biomedical text mining is that it can be developed rapidly [2], which can permit the leveraging of electronic health records of patients with a novel disease as quickly as they are entered into the EHR. However, although this has often been suggested [3], there has never been an opportunity to actually test that claim in real time. Thus, the current novel coronavirus disease (COVID-19) pandemic, with all of its challenges, presents an opportunity to advance the state of public health informatics. In this paper, we tested this possibility with a case study on the effects of use of calcium channel blockers (CCBs) in patients with high blood pressure on the risk of death from COVID-19 infection. An association between CCB and the outcome of COVID-19 infection has already been suggested [4] but has not previously been explored in a large multicenter clinical study.

Methods

Data Source and NLP Pipeline

The data used in this study were obtained from 39 different hospitals in the Paris metropolitan area in the Assistance Publique – Hôpitaux de Paris (AP-HP) system. Focusing on this region of the country and on a large number of hospitals afforded a diversity of patient demographics that would not be available in most other parts of the country. As of May 4, 2020, the Entrepôt de Données de Santé (EDS)-COVID data set contained 84,966 electronic records of suspected or confirmed patients with COVID-19 (see Table 1 for further details on the data set). The records comprise structured fields and free text documents, including clinical notes and narratives. Most of the textual documents do not follow a specific structure and contain different types of patient information, such as patient history, family history, laboratory results, drug history, and prescriptions. Therefore, they represent an excellent test case for the real abilities of text mining. We used the following pipeline:

- Typical preprocessing steps (ie, text cleaning and sentence detection) were applied to the full data set (see Multimedia Appendix 1 for a detailed description).
- Drug names and details of administration (dose, route of administration, frequency, and duration) were extracted via a deep learning approach based on bidirectional encoder representations from transformers (BERT) contextual embeddings [5] (NLP Medication).
- Specific phenotypes associated with COVID-19 (eg, obesity, smoking status), scores (eg, sequential organ failure assessment score) and physiological measures (eg, BMI), were extracted via a list of 60 regular expressions (NLP RegExp).
- All signs, symptoms, and comorbidities included in the Unified Medical Language System (UMLS) [6] were extracted with the quickUMLS algorithm [7] (NLP UMLS).

A visual depiction of the pipeline is provided in Multimedia Appendix 2.

The NLP medication extraction model was a bidirectional long short-term memory with a conditional random field (BiLSTM-CRF) [8] layer on top of a vector representation of tokens using BERT [5]. We fine-tuned multilingual BERT on a set of 10 million clinical texts from EHRs. The model was trained on the APMed corpus, a manually annotated corpus of French clinical texts described in [9]. We used the FLAIR [10] implementation with 2 layers of 1024 units for the LSTMs with an asynchronous stochastic gradient descent (ASGD) optimizer and a reduction of the learning rate on plateau.

The NLP regular expression for the extraction of specific phenotypes was a set of 60 regular expressions developed manually and iteratively by medical informatics experts and physicians. We evaluated their precision at the sentence level using a random sample of 100 positive sentences for each regular expression. Examples of these expressions can be found in Multimedia Appendix 3.

All the terms extracted by the NLP pipeline, regardless of the method, were automatically annotated according to their modality (negated or hypothetical) and experiencer in the text, as described in previous work [11]. The outputs of the NLP pipeline were normalized to the Observational Medical

Outcomes Partnership (OMOP) common data model (CDM) [12] and were fed back to the database system on a daily basis.

Data Availability

Data supporting this study can be made available on request, on condition that the research project is accepted by the scientific and ethics committee of the AP-HP health data warehouse [13].

Clinical Application: Long-Term CCB Use and Outcomes of COVID-19 in Patients With High Blood Pressure

The clinical goal of this case study was to evaluate the potential effects of CCBs on in-hospital mortality related to COVID-19 [4]. To achieve this goal, we used two different sources of data. The first source was two elements of structured data: International Classification of Disease, Tenth Revision (ICD-10) codes and medication prescriptions from an electronic prescription system. The second source was information on medications and comorbidities extracted by the NLP pipeline from nonstructured fields in the EHR. The inclusion criterion for patients was COVID-19 disease confirmed by reverse transcriptase–polymerase chain reaction (RT-PCR).

We considered a patient as receiving long-term treatment with CCBs (Multimedia Appendix 4) if there were at least two mentions (in structured data or extracted with NLP, respectively) in the last 6 months. We qualified cases as having comorbidities through one occurrence of an ICD-10 code (Multimedia Appendix 5) or two NLP mentions in the last 6 months.

The measured outcome was in-hospital mortality. We used a multivariate Cox proportional hazard model [14] that was adjusted according to age, gender, and the presence of obesity, diabetes, and cancer. The level of significance was set as $P=.05$, and all statistical tests were two-sided. We used R statistical software v.3.6.2 (R Project) with the Survival package.

Results

NLP Pipeline

As Table 1 shows, NLP markedly expanded the quantity of medication and phenotype information available for the analysis. The number of data points for medication increased by 7.2 times (*NLP medication*)/(*structured medication*), and the number of phenotypes increased by 15.2 times ((*NLP RegExp* + *NLP UMLS*)/(*ICD-10 codes*). Among the 84,966 patients with records present in the EDS-COVID cohort (Table 1), 45,593 (53.7%) contained drug information in their narrative EHR documents, whereas only 19,791 (23.3%) of the patients had medication information available in the structured fields in the EHR.

For specific phenotypes with existing ICD-10 codes (Figure 1), information was only available in clinical free-text fields for the majority of patients: 7133/8526 (60.2%) for diabetes, and 2138/2871 (74.5%) for obesity. Some items were absent from the structured data but could be recovered using the NLP extraction pipeline, such as COVID-19–specific symptoms such as ageusia (2449 patients) and anosmia (2732 patients).

In terms of quality, the extraction of medication names showed an F1 score of 93.8% (91.6% after normalization) in all sections. When focusing on the admission and discharge treatment sections, the F1 score was 96.7% (96.0% after normalization). The detailed results are shown in Multimedia Appendix 6. Regarding the phenotypes extracted by regular expressions in our case study, hypertension showed a precision of 99%, and obesity, diabetes, and cancer showed precisions of 94%, 80%, and 91%, respectively.

Table 1. Description of the information extracted using the NLP pipeline in the EDS-COVID cohort (N=84,966).

Source	Patient records (N=84,966), n (%)	Documents (N=1,524,057), n (%)	Data points, n
NLP[a] Medication	45,593 (53.7)	696,125 (45.7)	5,995,945
NLP RegExp[b]	44,498 (52.4)	711,900 (46.7)	5,449,932
NLP UMLS[c]	44,035 (51.8)	833,610 (54.7)	19,626,172
Structured medication	19,791 (23.3)	N/A[d]	826,554
ICD-10[e] codes	38,993 (45.9)	N/A	1,643,819

[a]NLP: natural language processing.

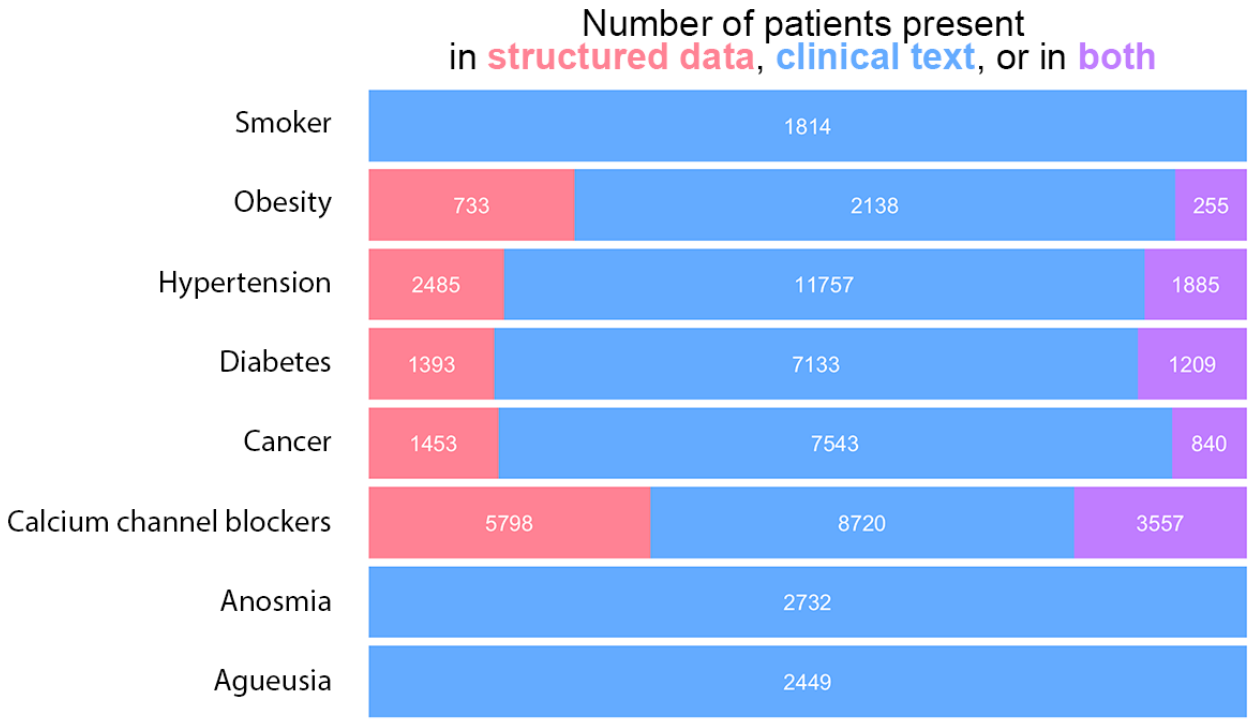
[b]RegExp: regular expression.

[c]UMLS: Unified Medical Language System.

[d]N/A: not applicable.

[e]ICD-10: International Classification of Disease, Tenth Revision.

Figure 1. Quantity of patients with information for a selection of items depending on the source of data.



Case Study

Of the 84,966 total patients, 3965 (4.7%) were included using the NLP pipeline, of which only 1343 (15.9%) could be included if the study were limited to the use of structured data; this increased the number of patients added for the case study increased by 2.95 times (Multimedia Appendix 7). A detailed description of the population of patients who tested positive for COVID-19 with a history of high blood pressure can be found in Multimedia Appendix 8). In terms of the temporal depth of CCB treatment information, Figure 2 shows that a higher volume of information was obtained from text fields compared to structured data.

When using only structured data, we observed an adjusted hazard ratio (aHR) of 0.83 (95% CI 0.67-1.05) for treatment with CCBs; this result was not statistically significant (*P*=.12). When including NLP data, the aHR became 0.82 (95% CI 0.71-0.94), which represents a statistically significant reduction of the risk of death (*P*=.005). Similar results can be observed that support an increased risk of mortality with the presence of diabetes and cancer as comorbidities (Table 2).

Figure 2. Quantity of information about calcium channel blockers for the two data sources over time. NLP: natural language processing.

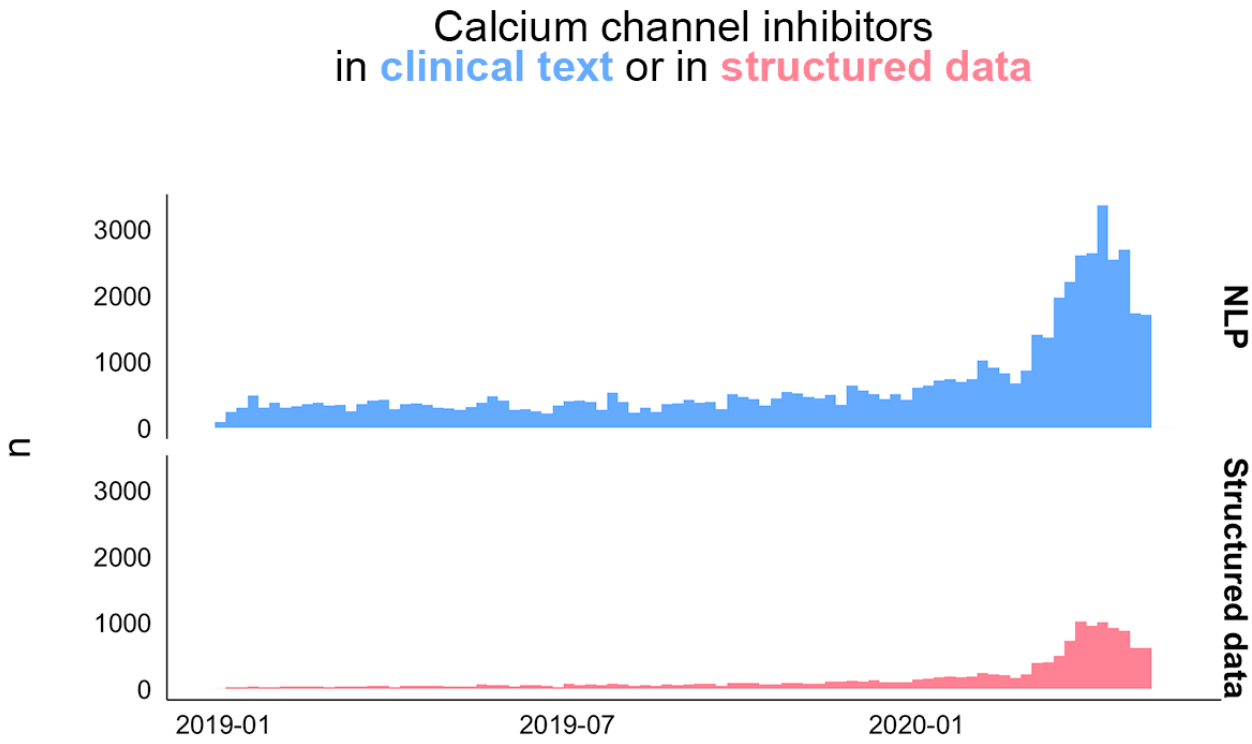


Table 2. Results of the multivariate Cox survival model.

Characteristic	Structured data			NLP[a]		
	aHR[b]	95% CI	*P* value	HR[c]	95% CI	*P* value
Calcium channel blockers	0.83	0.67-1.05	.12	0.82	0.71-0.94	.005
Age (years)						
45-64	Reference	N/A[d]	N/A	N/A	N/A	N/A
18-44	0.20	0.03-1.46	.11	0.35	0.15-0.80	.01
65-74	1.50	0.99-2.27	.053	1.95	1.54-2.47	<.001
75-84	1.68	1.14-2.48	.009	2.94	2.35-3.69	<.001
85+	2.45	1.66-3.61	<.001	3.99	3.16-5.03	<.001
Gender						
Female	Reference	N/A	N/A	N/A	N/A	N/A
Male	1.59	1.27-2.00	<.001	1.53	1.32-1.77	<.001
Obesity	1.07	0.81-1.42	.60	1.13	0.90-1.41	.30
Diabetes	1.22	0.98-1.52	.08	1.25	1.09-1.45	.002
Cancer	1.20	0.96-1.49	.11	1.34	1.15-1.56	<.001

[a]NLP: natural language processing.

[b]aHR: adjusted hazard ratio.

[c]HR: hazard ratio.

[d]N/A: not applicable.

Discussion

In this paper, we investigated the potential utility of biomedical NLP in the context of a rapidly emerging novel disease. To do this, we asked a specific question: Does the leveraging of unstructured textual information via NLP yield clinically actionable information? To answer this question, we used NLP to extract information about hypertension and a medication for treating it from the EHRs of patients with COVID-19. The results showed that an NLP pipeline can be adapted quickly to the domain of a novel disease, it can perform well enough to extract useful information, and when that information is used to supplement the structured data that is already available, the sample size can be increased sufficiently to see treatment effects that were not previously statistically detectable.

Several agencies, notably the European Medicines Agency, have highlighted the benefits of using real-world data for research, in particular for the generation of complementary evidence and new hypotheses [15]. During the peak of the COVID-19 pandemic, the time available for clinicians to enter EHR data was greatly reduced. Medical informatics became vital to manage the crisis in hospitals and acquire better knowledge of the disease. The NLP pipeline was implemented within two weeks at the beginning of the COVID-19 epidemic in France, building on previous developments in artificial intelligence and text mining at AP-HP. More specifically, combining nonspecific preexisting developments (eg, negation, family history, and hypothesis detection) to tailored extractions (ie, regular expressions) allowed us to obtain rapid results of sufficient quality.

Approximately 60 internal research projects exploring EDS-COVID data were submitted for Institutional Review Board approval within the first eight weeks of COVID-19 epidemic. More than half of these projects studied variables such as symptoms (eg, ageusia), radiological signs (eg, crazy paving), comorbidities (eg, obesity), and drug history (eg, hydroxychloroquine), requiring extraction of information from narrative reports in EHRs.

The case study described in this paper shows the possible impact of using information extracted from text in the EHR for COVID-19 research. More precisely, the conclusions of the above study would have been different if information from unstructured fields had been excluded. In our case study, the addition of information from NLP did not dramatically change the hazard ratio from the analyses; however, it allowed us to include more patients and therefore narrowed the CIs and increased the statistical power. Note that the increased statistical power is mainly due to the increase in the number of patients included and the quantity of data available. Further analyses are required to assess the validity of the associations detected here, given that some confounding biases may remain and provoke false positive results. Reproducing the analysis with an external population or performing falsification testing [16] could help improve the validity of these findings.

Acknowledgments

The authors thank the EDS AP-HP COVID consortium integrating the AP-HP Health Data Warehouse team as well as all the AP-HP staff and volunteers who contributed to the implementation of the EDS-COVID database and operating solutions for the database. The authors would like to acknowledge John Bennett for his thorough editing. This work was supported by state funding from the French National Research Agency (Agence Nationale de la Recherche, ANR) under the “Investissements d’Avenir” program (reference: ANR-10-IAHU-01) and an ANR PractikPharma grant (ANR-15-CE23-0028). The collaborators associated with AP-HP/Universities/INSERM COVID-19 Research Collaboration: AP-HP COVID CDR Initiative, Paris, France, are as follows: Pierre-Yves Ancel, Alain Bauchet, Nathanaël Beeker, Vincent Benoit, Mélodie Bernaux, Ali Bellamine, Romain Bey, Aurélie Bourmaud, Stéphane Breant, Anita Burgun, Fabrice Carrat, Charlotte Caucheteux, Julien Champ, Sylvie Cormont, Christel Daniel, Julien Dubiel, Catherine Duclos, Loic Esteve, Marie Frank, Nicolas Garcelon, Alexandre Gramfort, Nicolas Griffon, Olivier Grisel, Martin Guilbaud, Claire Hassen-Khodja, François Hemery, Martin Hilka, Anne Sophie Jannot, Jerome Lambert, Richard Layese, Judith Leblanc, Léo Lebouter, Guillaume Lemaitre, Damien Leprovost, Ivan Lerner, Kankoe Levi Sallah, Aurélien Maire, Marie-France Mamzer, Patricia Martel, Arthur Mensch, Thomas Moreau, Antoine Neuraz, Nina Orlova, Nicolas Paris, Bastien Rance, Hélène Ravera, Antoine Rozes, Elisa Salamanca, Arnaud Sandrin, Patricia Serre, Xavier Tannier, Jean-Marc Treluyer, Damien van Gysel, Gaël Varoquaux, Jill Jen Vie, Maxime Wack, Perceval Wajsburt, Demian Wassermann and Eric Zapletal.

Authors' Contributions

AN, IL, AB, NG, and BR contributed to the conception or design of the work. AN, IL, WD, NP, RT, NG, and BR acquired, analyzed, or interpreted the data. AN, IL, WD, NP, AR, DB, NG, and BR created the new software used in the work. AN, IL, AB, NG, RT, BR, and KBC drafted the work or substantively revised it.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary methods.

[DOCX File , 14 KB - jmir_v22i8e20773_app1.docx]

Multimedia Appendix 2
Description of the natural language processing pipeline.
[DOCX File , 53 KB - jmir_v22i8e20773_app2.docx]

Multimedia Appendix 3
Examples of regular expression for the extraction of phenotypes.
[DOCX File , 13 KB - jmir_v22i8e20773_app3.docx]

Multimedia Appendix 4
Definition of calcium channel blockers (name, ATC number).
[DOCX File , 12 KB - jmir_v22i8e20773_app4.docx]

Multimedia Appendix 5
Definition of phenotypes (name, ICD0-10 code).
[DOCX File , 12 KB - jmir_v22i8e20773_app5.docx]

Multimedia Appendix 6
Performance of the medication information extraction model before and after normalization of the entities.
[DOCX File , 14 KB - jmir_v22i8e20773_app6.docx]

Multimedia Appendix 7
Flowchart of the use case: patients who tested positive for COVID-19 who have hypertension.
[DOCX File , 227 KB - jmir_v22i8e20773_app7.docx]

Multimedia Appendix 8
Characteristics of the population of COVID positive patients with hypertension in EDS-COVID.
[DOCX File , 13 KB - jmir_v22i8e20773_app8.docx]

References

1. Snow J. On the Mode of Communication of Cholera. London, UK: Wilson and Ogilvy; 1855.
2. Chapman W, Dowling J, Ivanov O, Gesteland P, Olszewski R, Espino J, et al. Evaluating natural language processing applications applied to outbreak and disease surveillance. In: Proceedings of 36th symposium on the interface: computing science and statistics 2004. 2004 Presented at: 36th Symposium on the Interface: Computing Science and Statistics 2004; May 26-29, 2004; Baltimore, MD.
3. Elkin PL, Froehling DA, Wahner-Roedler DL, Brown SH, Bailey KR. Comparison of natural language processing biosurveillance methods for identifying influenza from encounter notes. Ann Intern Med 2012 Jan 03;156(1 Pt 1):11-18. [doi: 10.7326/0003-4819-156-1-201201030-00003] [Medline: 22213490]
4. Zhang L, Sun Y, Zeng H, Peng Y, Jiang X, Shang W, et al. Calcium channel blocker amlodipine besylate is associated with reduced case fatality rate of COVID-19 patients with hypertension. medRxiv 2020 Apr 14:preprint. [doi: 10.1101/2020.04.08.20047134]
5. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXivcs. 2018 Oct 10. URL: http://arxiv.org/abs/1810.04805 [accessed 2018-11-17]
6. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med 2018 Feb 06;32(04):281-291. [doi: 10.1055/s-0038-1634945]
7. Okazaki N, Tsujii J. Simple and Efficient Algorithm for Approximate Dictionary Matching. In: Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010).: Coling 2010 Organizing Committee; 2010 Presented at: 23rd International Conference on Computational Linguistics (Coling 2010); August 2010; Beijing, China URL: https://www.aclweb.org/anthology/C10-1096
8. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural Architectures for Named Entity Recognition. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.: Association for Computational Linguistics; 2016 Presented at: 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2016; San Diego, CA p. A. [doi: 10.18653/v1/n16-1030]
9. Jouffroy J, Feldman S, Lerner I, Rance B, Burgun A, Neuraz A. MedExt: combining expert knowledge and deep learning for medication extraction from French clinical texts. ResearchGate 2020 Jan:preprint. [doi: 10.2196/preprints.17934]

10. Akbik A, Bergmann T, Blythe D, Rasul K, Schweter S, Vollgraf R. FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations). Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations) Internet Minneapolis, Minnesota: Association for Computational Linguistics; 2019 Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations); June 2019; Minneapolis, MI. [doi: 10.18653/v1/n19-4010]
11. Garcelon N, Neuraz A, Benoit V, Salomon R, Burgun A. Improving a full-text search engine: the importance of negation detection and family history context to identify cases in a biomedical data warehouse. J Am Med Inform Assoc 2017 May 01;24(3):607-613. [doi: 10.1093/jamia/ocw144] [Medline: 28339516]
12. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform 2015;216:574-578 [FREE Full text] [Medline: 26262116]
13. Soumettre un projet de recherche au Comité Scientifique et Ethique de l'Entrepôt de Données de Santé. Assistance Publique — Hôpitaux de Paris. URL: https://recherche.aphp.fr/eds/recherche/ [accessed 2020-08-11]
14. Cox DR. Regression Models and Life-Tables. J R Stat Soc Series B Stat Methodol 2018 Dec 05;34(2):187-202. [doi: 10.1111/j.2517-6161.1972.tb00899.x]
15. EMA Regulatory Science to 2025: Strategic reflection. European Medicines Agency. 2018. URL: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf [accessed 2020-08-11]
16. Pizer SD. Falsification Testing of Instrumental Variables Methods for Comparative Effectiveness Research. Health Serv Res 2016 Apr;51(2):790-811 [FREE Full text] [doi: 10.1111/1475-6773.12355] [Medline: 26293167]

Abbreviations

aHR: adjusted hazard ratio
AP-HP: Assistance Publique – Hôpitaux de Paris
ASGD: asynchronous stochastic gradient descent
BiLSTM-CRF: bidirectional long short-term memory with a conditional random field
CCB: calcium channel blocker
CDM: common data model
COVID-19: coronavirus disease
EDS: Entrepôt de Données de Santé
EHR: electronic health record
ICD-10: International Classification of Disease, Tenth Revision
NLP: natural language processing
RT-PCR: reverse transcriptase–polymerase chain reaction
OMOP: Observational Medical Outcomes Partnership
UMLS: Unified Medical Language System

Edited by G Eysenbach; submitted 02.06.20; peer-reviewed by H Kalicoglu, S Zheng, D Pförringer, N Shah; comments to author 23.06.20; revised version received 02.07.20; accepted 26.07.20; published 14.08.20.

Please cite as:
Neuraz A, Lerner I, Digan W, Paris N, Tsopra R, Rogier A, Baudoin D, Cohen KB, Burgun A, Garcelon N, Rance B, AP-HP/Universities/INSERM COVID-19 Research Collaboration; AP-HP COVID CDR Initiative
Natural Language Processing for Rapid Response to Emergent Diseases: Case Study of Calcium Channel Blockers and Hypertension in the COVID-19 Pandemic
J Med Internet Res 2020;22(8):e20773
URL: http://www.jmir.org/2020/8/e20773/
doi:10.2196/20773
PMID:32759101

Original Paper

The Infection Rate of COVID-19 in Wuhan, China: Combined Analysis of Population Samples

Hui-Qi Qu[1], MD, PhD; Zhangkai Jason Cheng[1,2], PhD; Zhifeng Duan[1,3], PhD; Lifeng Tian[1], PhD; Hakon Hakonarson[1,4], MD, PhD

[1]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, United States

[2]Department of Radiation Oncology, Sun Yat-sen University Cancer Center, Guangzhou, China

[3]Institute of Allergy and Immunology, School of Medicine, Shenzhen University, Shenzhen, China

[4]Divisions of Human Genetics and Pulmonary Medicine, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Hakon Hakonarson, MD, PhD
Center for Applied Genomics
The Children's Hospital of Philadelphia
3615 Civic Center Boulevard
Philadelphia, PA
United States
Phone: 1 267 426 0088
Email: Hakonarson@email.chop.edu

Abstract

Background: The coronavirus disease (COVID-19) pandemic began in Wuhan, China, in December 2019. Wuhan had a much higher mortality rate than the rest of China. However, a large number of asymptomatic infections in Wuhan may have never been diagnosed, contributing to an overestimated mortality rate.

Objective: This study aims to obtain an accurate estimate of infections in Wuhan using internet data.

Methods: In this study, we performed a combined analysis of the infection rate among evacuated foreign citizens to estimate the infection rate in Wuhan in late January and early February.

Results: Based on our analysis, the combined infection rate of the foreign evacuees was 0.013 (95% CI 0.008-0.022). Therefore, we estimate the number of infected people in Wuhan to be 143,000 (range 88,000-242,000), which is significantly higher than previous estimates. Our study indicates that a large number of infections in Wuhan were not diagnosed, which has resulted in an overestimated case fatality rate.

Conclusions: Increased awareness of the original infection rate of Wuhan is critical for proper public health measures at all levels, as well as to eliminate panic caused by overestimated mortality rates that may bias health policy actions by the authorities.

(J Med Internet Res 2020;22(8):e20914) doi:10.2196/20914

KEYWORDS

infectious disease; COVID-19; infection rate; China; Wuhan; fatality; public health; diagnosis

Introduction

In December 2019, the first cases of coronavirus disease (COVID-19) were reported in Wuhan, China, a megacity with a population of approximately 11 million people. To prevent the spread of this highly infectious disease, the government initiated a city-wide lockdown on January 23, 2020. However, despite these efforts, COVID-19 spread to many countries across the world, reaching pandemic levels, and continues to be a serious public health concern due to its high mortality rate. According to the large-sample analysis by Wu and McGoogan [1], China's case fatality rate (CFR) was 2.3%—that is, 1023 deaths from 44,672 confirmed cases as of February 11, 2020, with a significant proportion of cases originating from Wuhan. The large number of infected people in Wuhan put a tight strain on essential medical resources. The city had a much higher mortality rate (according to Feb 10th statistics: CFR=4.05% [748 deaths/18,454 diagnoses]; Apr 24th statistics: CFR=7.69% [3869 deaths/50,333 diagnoses]) than the rest of China. The overall CFR of 2.3% for China was likely overestimated, due

to strained medical resources and a large number of undiagnosed patients. According to a recent study, 78% of those who had been infected were asymptomatic [2]. Therefore, a large number of asymptomatic infections in Wuhan might have never been diagnosed, which contributed to the overestimated CFR. An accurate estimation of the infection rate is therefore important to assess Wuhan's CFR precisely.

Methods

Using Markov Chain Monte Carlo methods, Wu et al [3] estimated that 75,815 individuals (95% CI 37,304-130,330) had been infected in Wuhan as of January 25, 2020. Following this, a number of foreign governments evacuated their citizens and performed thorough etiological tests on them. This group of evacuees can serve as a "random" sample to estimate the infection rate in Wuhan. With internet search as an important source of epidemiologic information on COVID-19 [4], we performed a combined analysis of the infection rates of these population samples using publicly available data (Table 1), instead of a simple pooled calculation, considering potential differences in lifestyles and pathogen exposure across different populations. The combined analysis was done using the Comprehensive Meta-Analysis Software (Biostat, Inc).

Table 1. Number of infected people from different countries.

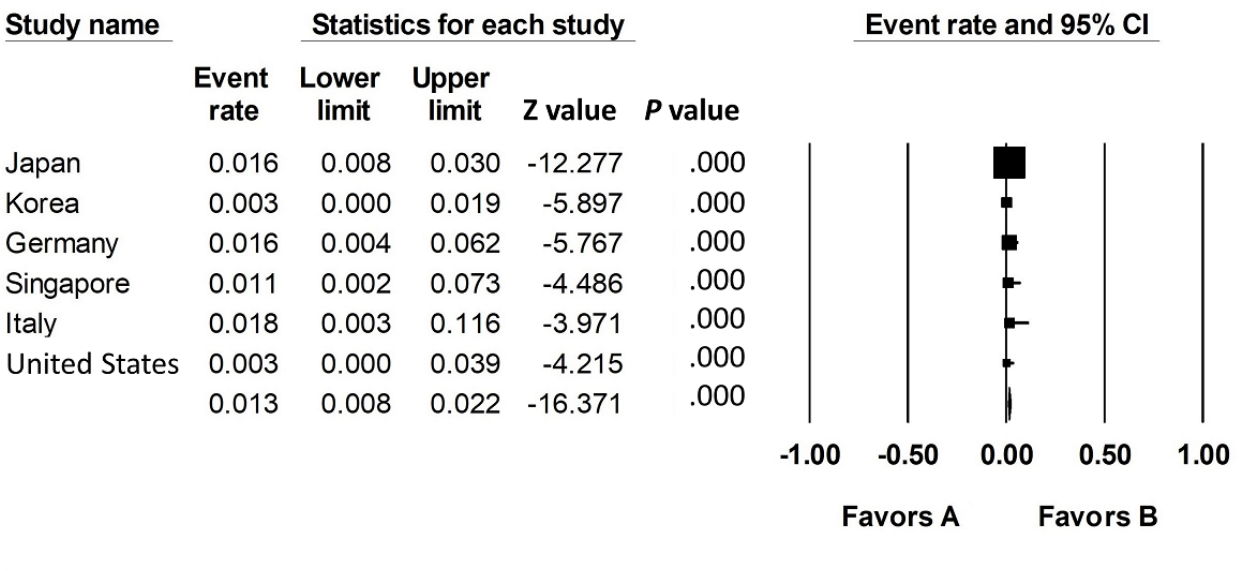
Country	Evacuation date	Confirmed cases (n=14), n	Evacuees (n=1401), n
Japan [5]	N/A[a]	9	566
Korea [6-8]	January 31, 2020	1	368
Germany [9]	February 1, 2020	2	124
Singapore [10-12]	January 30, 2020	1	92
Italy [13]	February 2, 2020	1	56
United States [14]	January 29, 2020	0	195

[a]N/A: not applicable.

Results

Our analysis demonstrates that there is no significant heterogeneity across different population samples (heterogeneity test P=.491). The combined infection rate is 0.013 (95% CI 0.008-0.022) (Figure 1). Based on our results, we estimate the number of infected people in Wuhan, China, to be 143,000 (range 88,000-242,000), which is significantly higher than the estimate proposed by Wu et al [3].

Figure 1. Combined analysis of infection rates of different populations.



Discussion

Our estimate indicates that a large number of infections in Wuhan were not diagnosed. The number of undiagnosed cases in late January and early February is larger than the final diagnosed count reported to date (n=50,333), which has resulted in an overestimated CFR. In addition, our study suggests that the lower CFR (0.51%) estimated by the Centre for Evidence-Based Medicine [15] does not indicate viral variants and loss of virulence. Taken together, increased awareness of the original infection rates in Wuhan, China, is critically important for appropriate public health measures at all levels,

as well as to eliminate panic caused by overestimated mortality rates that may bias health policy actions by the authorities.

Acknowledgments

The study was supported by Institutional Development Funds from the Children's Hospital of Philadelphia to the Center for Applied Genomics, The Children's Hospital of Philadelphia Endowed Chair in Genomic Research to HH.

Authors' Contributions

HQ was responsible for the literature search, study design, data collection, data analysis, data interpretation, and writing; ZJC was involved in data collection; ZD in data interpretation; LT in study design and data interpretation; and HH in study design, data interpretation, and writing.

Conflicts of Interest

None declared.

References

1. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24:e. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
2. Day M. Covid-19: four fifths of cases are asymptomatic, China figures indicate. BMJ 2020 Apr 02;369:m1375. [doi: 10.1136/bmj.m1375] [Medline: 32241884]
3. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet 2020 Feb 29;395(10225):689-697. [doi: 10.1016/S0140-6736(20)30260-9] [Medline: 32014114]
4. Higgins TS, Wu AW, Sharma D, Illing EA, Rubel K, Ting JY, Snot Force Alliance. Correlations of Online Search Engine Trends With Coronavirus Disease (COVID-19) Incidence: Infodemiology Study. JMIR Public Health Surveill 2020 May 21;6(2):e19702 [FREE Full text] [doi: 10.2196/19702] [Medline: 32401211]
5. Ministry of Health, Labor and Welfare. About the present situation of new type coronavirus infectious disease and correspondence of Ministry of Health, Labor and Welfare (February 7, 2019 version). 2019. URL: https://www.mhlw.go.jp/stf/newpage_09396.html [accessed 2020-02-12]
6. KBS World Radio. The second batch of Wuhan evacuation test results were negative, the first batch of 1 person diagnosed with infection. 2020. URL: http://world.kbs.co.kr/service/news_view.htm?lang=c&Seq_Code=66621 [accessed 2020-02-12]
7. Minister of Health and Welfare. New Coronavirus Infectious Disease Central Accident Headquarters Routine New Release. 2020. URL: https://www.mohw.go.kr/eng/nw/nw0101vw.jsp?PAR_MENU_ID=1007&MENU_ID=100701&page=1&CONT_SEQ=352718 [accessed 2020-02-12]
8. Ministry of Health and Welfare. Domestic occurrence of new coronavirus (February 2). 2020. URL: http://www.mohw.go.kr/react/al/sal0301vw.jsp?PAR_MENU_ID=04&MENU_ID=0403&page=3&CONT_SEQ=352645 [accessed 2020-02-12]
9. DW. Coronavirus: German evacuation flight from China carried two infected people. 2020. URL: https://www.dw.com/en/coronavirus-german-evacuation-flight-from-china-carried-two-infected-people/a-52229955 [accessed 2020-02-12]
10. APF. Coronavirus: countries evacuate citizens from China. New Straits Times. 2020. URL: https://www.nst.com.my/world/world/2020/01/561086/coronavirus-countries-evacuate-citizens-china [accessed 2020-02-12]
11. Wei TT. Some Singaporeans with symptoms of virus staying behind in Wuhan even as 92 are evacuated. Straits Times. 2020. URL: https://www.straitstimes.com/singapore/some-singaporeans-with-symptoms-of-virus-staying-behind-in-wuhan-even-as-92-are-evacuated [accessed 2020-02-12]
12. Yong C. Coronavirus: 2 new cases in S'pore, including Certis officer who had served quarantine orders on 2 who tested positive. Straits Times. 2020. URL: https://www.straitstimes.com/singapore/coronavirus-2-new-cases-in-singapore-including-certis-cisco-staff-who-had-served [accessed 2020-02-12]
13. Seckin B. Italy reports third confirmed case of coronavirus. Anadolu Agency. 2020. URL: https://www.aa.com.tr/en/europe/italy-reports-third-confirmed-case-of-coronavirus/1726934 [accessed 2020-02-12]
14. Waldrop T, Yan H. The 1st group of Americans evacuated from Wuhan and quarantined over coronavirus just got released. CNN. 2020. URL: https://edition.cnn.com/2020/02/10/us/coronavirus-american-evacuees-release/index.html [accessed 2020-02-12]
15. Oke J, Heneghan C. Global Covid-19 Case Fatality Rates. Centre for Evidence-Based Medicine. 2020. URL: https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/ [accessed 2020-06-12]

Abbreviations

COVID-19: coronavirus disease

CFR: case fatality rate

Edited by G Eysenbach; submitted 04.06.20; peer-reviewed by W Zhang, P Banik; comments to author 18.07.20; revised version received 18.07.20; accepted 22.07.20; published 14.08.20.

Please cite as:
Qu HQ, Cheng ZJ, Duan Z, Tian L, Hakonarson H
The Infection Rate of COVID-19 in Wuhan, China: Combined Analysis of Population Samples
J Med Internet Res 2020;22(8):e20914
URL: https://www.jmir.org/2020/8/e20914
doi:10.2196/20914
PMID:32707538

Original Paper

Use of the HoloLens2 Mixed Reality Headset for Protecting Health Care Workers During the COVID-19 Pandemic: Prospective, Observational Evaluation

Guy Martin[1,2*], MBBS, PhD; Louis Koizia[3], MBBS, BSc; Angad Kooner[4], MBBS, BSc; John Cafferkey[4], BMBS; Clare Ross[4], MBBS, BSc; Sanjay Purkayastha[1,2], MBBS, MD; Arun Sivananthan[1], MBBS, BSc; Anisha Tanna[5], MBBS, BSc; Philip Pratt[6], PhD; James Kinross[1,2], MBBS, PhD; PanSurg Collaborative[7*]

[1]Department of Surgery and Cancer, Imperial College London, London, United Kingdom

[2]Division of Surgery, Imperial College Healthcare NHS Trust, London, United Kingdom

[3]Cutrale Perioperative and Ageing Group, Imperial College London, London, United Kingdom

[4]Division of Medicine, Imperial College Healthcare NHS Trust, London, United Kingdom

[5]West London Renal and Transplant Centre, Imperial College Healthcare NHS Trust, London, United Kingdom

[6]The Helix Centre, Imperial College London, London, United Kingdom

[7]See Authors' Contributions section

[*]these authors contributed equally

Corresponding Author:
Guy Martin, MBBS, PhD
Department of Surgery and Cancer
Imperial College London
10th Floor, QEQM Building
St Mary's Hospital
London, W2 1NY
United Kingdom
Phone: 44 0203 312 6072
Email: guy.martin@imperial.ac.uk

Abstract

Background: The coronavirus disease (COVID-19) pandemic has led to rapid acceleration in the deployment of new digital technologies to improve both accessibility to and quality of care, and to protect staff. Mixed-reality (MR) technology is the latest iteration of telemedicine innovation; it is a logical next step in the move toward the provision of digitally supported clinical care and medical education. This technology has the potential to revolutionize care both during and after the COVID-19 pandemic.

Objective: This pilot project sought to deploy the HoloLens2 MR device to support the delivery of remote care in COVID-19 hospital environments.

Methods: A prospective, observational, nested cohort evaluation of the HoloLens2 was undertaken across three distinct clinical clusters in a teaching hospital in the United Kingdom. Data pertaining to staff exposure to high-risk COVID-19 environments and personal protective equipment (PPE) use by clinical staff (N=28) were collected, and assessments of acceptability and feasibility were conducted.

Results: The deployment of the HoloLens2 led to a 51.5% reduction in time exposed to harm for staff looking after COVID-19 patients (3.32 vs 1.63 hours/day/staff member; *P*=.002), and an 83.1% reduction in the amount of PPE used (178 vs 30 items/round/day; *P*=.02). This represents 222.98 hours of reduced staff exposure to COVID-19, and 3100 fewer PPE items used each week across the three clusters evaluated. The majority of staff using the device agreed it was easy to set up and comfortable to wear, improved the quality of care and decision making, and led to better teamwork and communication. In total, 89.3% (25/28) of users felt that their clinical team was safer when using the HoloLens2.

Conclusions: New technologies have a role in minimizing exposure to nosocomial infection, optimizing the use of PPE, and enhancing aspects of care. Deploying such technologies at pace requires context-specific information security, infection control, user experience, and workflow integration to be addressed at the outset and led by clinical end-users. The deployment of new telemedicine technology must be supported with objective evidence for its safety and effectiveness to ensure maximum impact.

(J Med Internet Res 2020;22(8):e21486) doi:10.2196/21486

KEYWORDS

COVID-19; mixed reality; telemedicine; protection; acceptability; feasibility; impact; headset; virtual reality; augmented reality; pilot

Introduction

The coronavirus disease (COVID-19) pandemic has overwhelmed even the most developed and well-resourced health systems [1]. Difficult decisions regarding the rationing of personal protective equipment (PPE) for health care workers and even access to care for patients have had to be made [2]. Protecting the health and safety of care workers is a key priority to maintain the quality of care delivered to individual patients and the ability of health systems to deliver care at scale [3]. In Italy, up to 20% of health care workers have become infected with the virus [4], and in the United Kingdom, over 15% of positive tests have been in critical health care workers [5] and over 100 have died [6]—a picture that has been seen globally. A key aspect of this has been severe disruption and shortages in the global supply of PPE due to excess demands and misuse [7]. As such, novel methods that optimize PPE use and protect both health care workers and patients from COVID-19 transmission are urgently needed [8].

Digital innovation has been identified as key to tackling the challenge to staff safety that COVID-19 confers [9]. The pandemic has rapidly accelerated the deployment of new technologies such as telemedicine services [10]. Telemedicine provides a means to deliver care efficiently and leverage access to multiple remote specialists while simultaneously protecting staff and patients from exposure to the virus [11,12]. Mixed-reality (MR) technology offers an immersive experience in which real and virtual elements of an environment dynamically coexist; it is the most recent iteration and extension of telemedicine innovation with the potential to revolutionize clinical care and education through the provision of enhanced functionality and novel content.

The HoloLens2 is an untethered wearable holographic computer that allows bidirectional communication with multiple remote users via video, voice, and MR composites. The technology has been used previously in a variety of clinical and educational scenarios, including perioperative planning, surgical training, anatomical teaching, and 3D telemedicine support [13-17]. The technology offers the potential to increase user immersion and engagement, supplement situational awareness and access to knowledge in real time, and improve performance. It allows users to interact and manipulate spatially registered 3D holographic content within a real environment, and to remotely link with multiple devices and users to allow simultaneous collaborative interaction and working within the visualized environment. MR technology is in its infancy but is the logical next step in the move toward the provision of digitally supported clinical care and education.

The HoloLens2 device has not previously been deployed for the delivery of ward-based secondary care in high-risk environments. This technology-led pilot therefore deployed and evaluated the HoloLens2 for the delivery of remote care across a range of inpatient settings in a UK teaching hospital during the COVID-19 pandemic response. The aim was to assess the practicalities and impact of deploying MR telemedicine technologies for improving staff safety during the pandemic.

Methods

Conduct

The objective of this project was to assess the practicalities and clinical impact of introducing a remote distributed care model supported by the HoloLens2. All technologies sit within a multidomain system that includes its users and environment; it cannot be developed, deployed, or evaluated in isolation [18]. Workflow considerations and human factors must therefore also be considered alongside technical decisions when understanding the practicalities and impact of implementing a digital technology and models of care at pace.

A prospective, observational, nested cohort evaluation of the HoloLens2 as a technology-led quality improvement (QI) project was performed. The device was deployed and evaluated across three distinct clinical clusters: a COVID-19 general medicine ward, a specialist COVID-19 unit providing continuous positive airway pressure support, and finally a specialist unit providing care to COVID-19 patients with renal disease. High-level aggregate outcome data pertaining to staff exposure to high-risk COVID-19 environments and PPE use were collected. Assessments of feasibility and acceptability were undertaken via user experience questionnaires with Likert and free-text responses. Local institutional registration and approval was obtained, and data governance and infection prevention and control procedures were agreed upon prior to the commencement of the project. No additional ethical approval was required as the project was conducted as a technology-led QI project under the supervision of the institutional QI team. Explanatory information was provided to all participants. All data were arranged, structured, and analyzed in Microsoft Excel (V15.22, Microsoft Corporation) and IBM SPSS (V26, IBM Corporation). With regards to statistical analysis, standard descriptive statistics were employed, and two-tailed Student's *t* tests were used to compare differences, with significance set at $P<.05$.

Workflow

Prior to the deployment of the HoloLens2, standard clinical practice in all three areas was to provide clinical care through face-to-face ward rounds comprising 3-8 clinical staff members. This required all to don appropriate PPE and provide care in high-risk environments. For this project clinical teams were provided with standardized training on the basic functions of the HoloLens2 device. Teams were not given specific instructions on how to alter their clinical practice or models of care and were free to use the technology in an optimal way for

their local clinical context to support the transition to a remote distributed care model.

In general, following deployment a single senior member of staff would enter the COVID-19 environment to undertake rounds, with the remainder of the team joining virtually from a COVID-19–protected, nonclinical remote location. All members of the team then played an active role in clinical assessment and decision making through the bidirectional audio-visual functionality of the device, and specifically through the first-person bedside view provided to the remote team members. In addition, relevant imaging and electronic health record (EHR) data were placed directly into the field of vision of the device user, with the aim of improving situational awareness, informing better clinical decision making and further reducing the risk of viral transmission by minimizing the need to interact physically with equipment and technology in high-risk areas. In parallel, members of the remote team would document encounters in the EHR, and undertake electronic ordering and prescribing in real time. Device use and functionality is illustrated in Figure 1.

Figure 1. Images demonstrating the use and functionality of the HoloLens2 device. (A) View of end-user in personal protective equipment wearing the device. (B) View of remote clinical team engaging in clinical round from a safe location. (C) First-person view through the HoloLens2 showing the remote clinical team and relevant imaging placed in the user's field of view as mixed-reality composites (image generated with study group to ensure the protection of patient privacy and data).

HoloLens2

The HoloLens2 is produced and marketed by Microsoft Corporation (Redmond, WA, USA) and is an untethered mixed-reality headset that combines several types of sensors, infrared time-of-flight depth measurement, high-definition cameras, accelerometers, and microphones. It provides a true heads-up display functionality with the ability to place interactive 2D and 3D objects, such as medical imaging or EHR data, within a user's visual field. Simultaneously, it provides live bidirectional communication via video and voice with multiple remote users through the Remote Assist application to enable hands-free multidisciplinary telemedicine at the bedside. Microsoft Dynamic 365 Remote Assist utilizes the architecture of Microsoft Teams—a unified communication and collaboration platform that combines chat, video meetings, file storage, and application integration—which was deployed across the National Health Service (NHS) in England to help providers respond to the COVID-19 pandemic [19]. The deployment of existing well-used and flexible platforms is likely to be beneficial as their technical and security considerations are already understood.

Information Security and Governance

Ensuring appropriate information security and governance is vital when utilizing new digital technologies, especially when sensitive patient data may be shared electronically. In the United Kingdom, NHSX and other relevant bodies recently updated their guidelines on the use of digital technology and sharing of data during the pandemic, recognizing that a pragmatic, risk-based approach needs to be taken and that an effective pandemic response will require new workflows [20]. For the

deployment of the HoloLens2 and Remote Assist, we connected devices to the secure hospital Wi-Fi network; devices were specifically white listed by media access control address (MAC address) and secured by WPA2 preshared key authentication. We utilized mobile device management to automate deployment, provisioning, policy management, application delivery, and updates across all devices. User accounts were protected with strong passwords and multifactor authentication. Institutional approval for data protection, confidentiality, and information sharing was obtained.

Infection Prevention and Control

The aim of this project was to protect staff from infection; therefore, ensuring a standardized process for wearing the device with PPE and for device decontamination was vital. Specific PPE requirements varied according to each clinical setting, local risk assessment, and clinical tasks being undertaken. However, all members entering a high-risk environment are required to wear 4-5 distinct items of PPE, including gown or apron, gloves, hat, mask or respirator, and eye protection. A pragmatic, risk-based approach to wearing and decontaminating the HoloLens2 devices was developed, with a standard method for using the device with different PPE that included the use of surgical caps to protect the device, as well as customized full-face visors for higher risk areas that had a strip cut out to ensure that the sensors and cameras on the headset were not obstructed, while ensuring adequate protection for staff. The cleaning process was aligned with that used for the decontamination of reusable items of PPE such as full-face plastic visors. The development of safe ways to wear and decontaminate the devices was conducted in partnership with our organizational Infection Prevention and Control team prior to commencing the project. An example of donning/doffing and decontamination procedures has been provided in Multimedia Appendix 1.

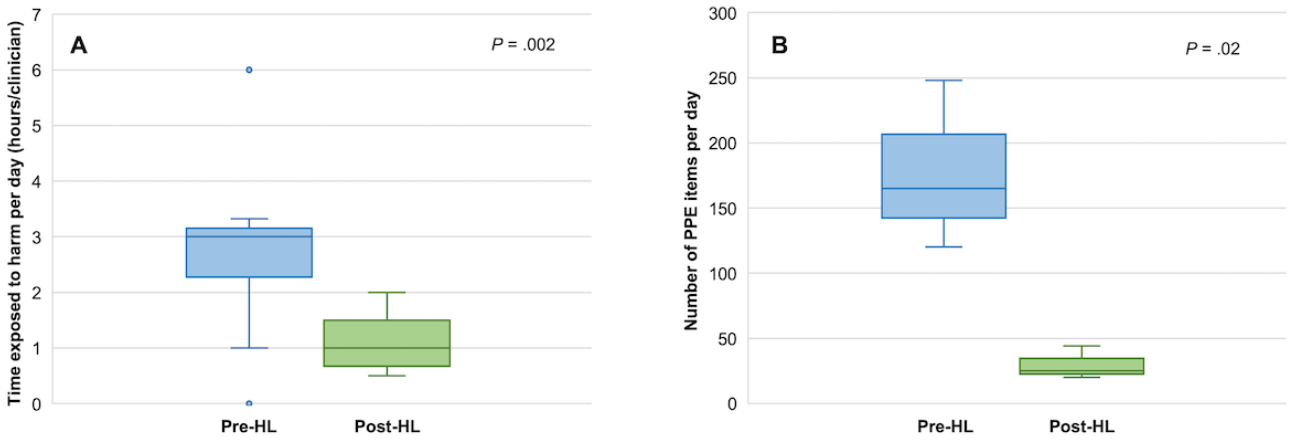
Results

In total, 52 inpatient beds were included across the three clinical clusters, with an aggregate of 51 days of clinical care evaluated during the project. A total of 28 members of clinical staff (n=16, 57.1% male; n=18, 64.3% <35 years of age) completed assessments of acceptability and feasibility, 20 (71.4%) of whom had no experience using the HoloLens2 prior to the project; 23 (82.1%) participants had never used video calling or telemedicine software before the COVID-19 pandemic.

Deployment of the HoloLens2 led to a significant reduction in the mean aggregate duration (hours/day/staff member) that clinical staff were exposed to high-risk COVID-19 environments while delivering clinical care (3.32 hrs vs 1.63 hrs, t=3.21, P=.002) as shown in Figure 2A. The total reduction in exposure ranged from 7.15 hrs to 15.45 hrs per day collectively for each team. This represents 222.98 hrs per week of reduced staff exposure to COVID-19 infection across the three clinical areas evaluated, equivalent to a 51.5% reduction. The overall reduction in exposure to risk was achieved by reducing the number of staff on each round by 2-6 people and by greatly reducing the duration of each round. In addition, by allowing the completion of administrative tasks that were previously undertaken following completion of the round in parallel, overall efficiency was increased and postround workload was reduced:

> *Time was saved by not waiting for other people to change PPE between patients and having to repeat the plan on coming out for bays/side rooms. By the time the ward rounds were complete, the documentation and most of the scans/reviews had been requested.* [Renal medicine consultant]

Figure 2. Aggregate data on staff exposure to risk and personal protective equipment (PPE) use across three clinical areas before and after HoloLens2 (HL) deployment. (A) Reduction in time (hours/day/staff member) exposed to high-risk coronavirus disease (COVID-19) environments (3.32 vs 1.63 hours, t=3.21, P=.002). (B) Reduction in the mean number of PPE items used (178 vs 30, t=3.88, P=.02).



Deployment of the HoloLens2 also led to a significant reduction in the mean number of PPE items used (items/round/day) while delivering clinical care in high-risk COVID-19 environments (178 vs 30, t=3.88, P=.02), as shown in Figure 2B. The total reduction in PPE items used ranged from 100 to 204 per day for each team. This represents a total saving of approximately 3100 items of PPE per week across the three clinical areas evaluated, which is equivalent to an 83.1% reduction. Importantly, 89.3% (25/28) of clinical staff felt that their team was safer when using the HoloLens2 to look after COVID-19

positive patients. No member of staff reported any safety concerns while using the device, and all respondents were happy to use it again:

> *The most important thing is that it is able to protect us from getting infected…* [Junior doctor]

With regards to device functionality, all respondents utilized bidirectional audio-visual communication, 23 (82.1%) used EHR data MR composites, 21 (75.0%) used imaging MR composites, and 16 (57.1%) used interactive MR tools such as 3D object annotation. Overall, 21 (75.0%) users agreed the device was easy to set up and 20 (71.4%) thought it was comfortable to wear:

> *To be honest I forget I am wearing it most of the time.* [General medicine consultant]

In total, 22 (78.6%) staff agreed that the HoloLens2 led to a quicker round, 21 (75.0%) a more efficient round; 22 (78.6%) had a better experience of undertaking care when using the device and 19 (67.9%) felt more engaged with the round and clinical decision making. The majority agreed that the device improved the quality of patient care (n=19, 67.9%) and that it enabled staff to make better clinical decisions (n=17, 60.7%). Only a single respondent (3.6%) reported that it reduced the quality of care, and 2 (7.2%) participants thought it did not help support better decision making. Most agreed that the HoloLens2 improved the quality of communication within the clinical teams (n=20, 71.4%), teamwork (n=23, 82.1%), and clinical situational awareness for staff while reviewing patients (n=19, 67.9%):

> *You can really see a clear picture so we can fully see all the signs, which on a crowded ward round you may not when you are stuck at the back somewhere.* [Junior doctor]

Overall, 21 (75.0%) respondents could see the clear benefit in using the technology, and 15 (53.6%) respondents agreed that it should be used for all ward rounds.

Discussion

Principal Findings

This technology-led pilot project has demonstrated that wearable MR devices may have a role to play in protecting staff and reducing PPE use during a pandemic response. The use of the HoloLens2 led to an 83.1% reduction in PPE use and a 51.5% reduction in the time spent by clinical staff in high-risk areas. Nearly 90% of staff felt that their clinical team was safer when using the HoloLens2 to care for COVID-19–positive patients. This feedback suggests that these material improvements in safety and PPE use do not impact the quality nor consistency of care provided and may even enhance aspects of multiprofessional care that have been hampered due to restrictions and changes in practice resulting from COVID-19. The provision of first-person, hands-free audio and visual communication across a distributed team, together with the ability to introduce relevant health data via MR composites is an important development in telemedicine and has the potential for far wider applicability outside of the immediate response to COVID-19.

Limitations

This project suggests that substantial benefits can be obtained through the wider roll-out of MR-based technology; however, this pilot project is not without its limitations. The technology has been deployed in a digitally mature organization and across clinical areas led by motivated and interested staff. As such, there is a risk of early adopter bias, a deployment timeline that is not universally achievable, and the potential for this enthusiasm to have influenced the reported outcomes and provide an overestimation of its usability and practicality. In addition to this, the deployment within a single organization in a nonblinded, randomized, or controlled fashion may have led to the introduction of further bias and implications for the wider applicability of the findings. Further to the limitations in the methodology of this pragmatic pilot study, some issues with the technology were also highlighted. These included limited battery life, concerns around remote participants adequately hearing those not wearing the device (eg, patients), and stability and connection problems related to network speed and capacity. A key concern articulated by some users was that by providing distributed care, team members do not spend direct time with patients, which in turn may have negative consequences for future interactions and the overall quality of the doctor-patient relationship. Despite these valid concerns, patients themselves recognized the benefits presented by the technology:

> *They are not only saving me… I am not passing anything onto them or their friends.* [Patient]

In order to realize the potential benefits of this technology fully, it is vital that further work is undertaken to better understand all aspects of its impact on care delivery, staff safety, and patient outcomes. COVID-19 has led to a rapid expansion in the use of telemedicine and other digital technologies, but much of this has been understandably done at pace, without robust evaluation or assessment strategies, and often with little evidence for its safety, efficacy, and cost-effectiveness [21,22]. There is a need to demand a standardized process for effective governance, as well as robust and transparent evaluation strategies that encompass all aspects of the technology and how it impacts patients, hospitals, and the staff that use it. In addition, there is a need to develop what is an off-the-shelf solution further so that it can better meet the context-specific aspects of in-hospital care; for example, the production of device-specific protective face shields or the development of clinically focused software that allows more user-friendly blending of health data MR composites. The current ad-hoc use of generic software, while meeting the overall objective, limits future applications and potential impact.

Comparison With Prior Work

The use of electronic or smart PPE has been identified as having the potential to protect staff and conserve resources, while simultaneously providing rapid access to emergency care [11]. However, these have largely focused on mobile or desktop solutions rather than devices which are truly wearable, and often involve ad-hoc technology fixes. The use of hands-free MR technology allows for objective improvements in communication and situational awareness for all members of the team. Enabling remote MR-supported clinical assessments can enhance the

ability of team members to recognize and respond to changes in a patient's condition over and above that offered by more basic voice or video technologies [23]. The provision of first-person, real-time audio and visual information allowed team members to "get a feel" for the patient they are consulting on despite physical distance. This was particularly evident for more junior members of the team, who may previously have not been directly at the bedside during consultations. Further to providing equivalent, if not enhanced, quality of care, obviating the physical presence of the majority of the clinicians in the ward significantly reduced the use of PPE. This is key to rationalize PPE use and may mitigate the significant supply chain disruption and equipment shortages that have been seen globally during the pandemic [24]. In addition to reducing the time clinicians spent exposed to harm and the amount of PPE used, the devices further minimized the risk of disease transmission by removing the need to handle equipment physically, such as computers, in the clinical environment. Even with proper hand sanitization and PPE use, there is still the potential for contamination of portable medical equipment [25], and so reducing physical interaction is key in minimizing risk. Finally, it is estimated that up to 7% of health care workers may have an asymptomatic COVID-19 infection [26], and around 20% of COVID-19 infections in hospital inpatients are thought to be nosocomial [27]; the use of technology to reduce exposure may therefore also act to protect patients as well as staff from disease spread.

When embarking on rapid response projects such as this, it is important to select a relatively mature off-the-shelf technology and only seek to make minor modifications. This ensures that there is background technical knowledge and awareness to help speed up deployment and aid with troubleshooting. In addition, by choosing or modifying existing technologies, issues pertaining to data security and privacy should have already been explored and understood to some extent. Agility is needed in complying with data governance regulations when they themselves may rapidly change in response to a crisis [20]. However, once the initial response is complete and the transition to a mature embedded technology platform begins, an in-depth governance and security assessment must be undertaken to ensure full compliance with regulation and legislation, and to maintain the trust of users and patients. Choosing a mature technology also means that there may be some existing evidence for its efficacy and safety, although this is often not the case with new digital technologies [28]. It is essential to ensure that deployment is context specific as no two clinical areas possess the same workflows, structures, and team practices. Final strategies for deployment and alignment to current workflows need to be flexible and driven by the clinical end-users. A clear evaluation plan must also be developed to ensure that efficacy, safety, and impact are captured and disseminated rapidly and robustly. This will ensure that time and resource are not squandered on failing technologies, and potential benefits and successes are rapidly spread both locally and more widely. Linked to this requirement for robust evidence is a need to examine patient-reported outcomes and measures of impact. It is crucial to ensure that patients perceive the quality and experience of their care to be at least maintained, if not improved with the deployment of new technologies, which is an area that was not explored in detail in this pilot study.

Contamination rates of personal devices in hospital can be over 30% and contribute to the risk of pathogen transmission [29]. The HoloLens2, like any head worn technology, is reusable, and the risk of contamination is high. Therefore, when deployed in high-risk areas, particularly when airborne disease transmission is of greatest concern, effective protocols for cleaning and decontamination need to be developed and implemented. Clear and easy-to-follow protocols for donning and doffing of the equipment must be developed locally; effective training should be delivered to ensure that staff minimize the risk of device contamination and do not put themselves at risk of self-contamination and potential infection when using off the device [30,31]. It is important to include all end-users and subject matter experts in infection control when developing and implementing these procedures to ensure they are evidence-based and that they pragmatically balance risk against ease of implementation and use to ensure optimal adherence and impact [32].

Conclusions

This pilot nested cohort study has shown that new technologies such as the HoloLens2 have a potentially important role to play when delivering care to patients with COVID-19, minimizing staff and patient exposure to nosocomial infection risk, optimizing the use of PPE, and enhancing aspects of care. This technology has empowered a diverse group of clinicians to collaborate more effectively and efficiently, improved the transfer and dissemination of information and knowledge, and allowed care to be delivered more safely with reduced PPE demands. The initial experience of using the HoloLens2 in high-risk clinical areas is promising. What is now required is further development to cement the technology in day-to-day practice and the evolution of bespoke tools and applications that will enhance its capabilities. These developments must be coupled with robust objective evidence for its safety and effectiveness across a range of settings to ensure its impact on staff, patients, and hospitals is fully realized.

Acknowledgments

The research was supported by the UK NIHR (National Institute for Health Research) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. Microsoft Corporation (Redmond, WA, USA) provided in-kind support through the provision of technical assistance and supply of devices. This paper represents independent research. The funders, and specifically Microsoft Corporation, had no involvement in the study design; data collection, analysis, and interpretation; writing of the report; or decision to submit the paper for publication.

Authors' Contributions

GM, LK, AK, JC, CR, SP, AS, AT, PP, and JK all contributed to the paper. GM, PP, and JK designed and led the project. LK, AK, JC, CR, AS, and AT led the clinical deployment of the technology. All authors contributed to the manuscript. JK was responsible for overall supervision and is guarantor for the study. All authors have access to all of the data and can take responsibility for the integrity of the data and accuracy of the data analysis.
The PanSurg collaborative is a diverse group of clinicians and academics from Imperial College London undertaking a range of educational, clinical, and research activities in support of the COVID-19 pandemic response; this project represents part of that work.

Conflicts of Interest

JK is a clinical advisor, and PP the chief scientific officer of Medical iSight, a 3D visualization and surgical guidance software company. At the time of submission, no other authors declare further relationships or activities that could appear to have influenced the submitted work. Microsoft Corporation had no direct influence on how the study was executed or reported; however, the provision of in-kind support through the supply of devices and technical support is a potential conflict to acknowledge.

Multimedia Appendix 1
Suggested protocol for donning/doffing and decontamination of the HoloLens2 device.
[PDF File (Adobe PDF File), 319 KB - jmir_v22i8e21486_app1.pdf]

References

1. Moghadas SM, Shoukat A, Fitzpatrick MC, Wells CR, Sah P, Pandey A, et al. Projecting hospital utilization during the COVID-19 outbreaks in the United States. Proc Natl Acad Sci U S A 2020 Apr 21;117(16):9122-9126 [FREE Full text] [doi: 10.1073/pnas.2004064117] [Medline: 32245814]
2. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. N Engl J Med 2020 May 21;382(21):2049-2055. [doi: 10.1056/NEJMsb2005114] [Medline: 32202722]
3. PanSurg Collaborative. The three vital lessons Italian hospitals have learned in fighting covid-19. Heal Serv J. 2020. URL: https://www.hsj.co.uk/emergency-care/exclusive-the-three-vital-lessons-italian-hospitals-have-learned-in-fighting-covid-19/7027220.article [accessed 2020-05-24]
4. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? The Lancet 2020 Apr;395(10231):1225-1228. [doi: 10.1016/s0140-6736(20)30627-9]
5. Heneghan C, Oke J, Jefferson T. COVID-19 How many healthcare workers are infected? CEBM. 2020. URL: https://www.cebm.net/covid-19/covid-19-how-many-healthcare-workers-are-infected/ [accessed 2020-05-24]
6. Cook T, Kursumovic E, Lennane S. Deaths of NHS staff from covid-19 analysed. HSJ. 2020. URL: https://www.hsj.co.uk/exclusive-deaths-of-nhs-staff-from-covid-19-analysed/7027471.article [accessed 2020-05-24]
7. Shortage of personal protective equipment endangering health workers worldwide. World Health Organization. 2020. URL: https://www.who.int/news-room/detail/03-03-2020-shortage-of-personal-protective-equipment-endangering-health-workers-worldwide [accessed 2020-05-22]
8. The Lancet. COVID-19: protecting health-care workers. The Lancet 2020 Mar 21;395(10228):922 [FREE Full text] [doi: 10.1016/S0140-6736(20)30644-9] [Medline: 32199474]
9. Guest JL, Del Rio C, Sanchez T. The Three Steps Needed to End the COVID-19 Pandemic: Bold Public Health Leadership, Rapid Innovations, and Courageous Political Will. JMIR Public Health Surveill 2020 Apr 06;6(2):e19043 [FREE Full text] [doi: 10.2196/19043] [Medline: 32240972]
10. UK's technological response to the COVID-19 healthcare crisis. Health Tech Digital. 2020. URL: https://www.healthtechdigital.com/uks-technological-response-to-the-covid-19-health-care-crisis/ [accessed 2020-05-18]
11. Turer R, Jones I, Rosenbloom T, Slovis C, Ward M. Electronic personal protective equipment: A strategy to protect emergency department providers in the age of COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):967-971 [FREE Full text] [doi: 10.1093/jamia/ocaa048] [Medline: 32240303]
12. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
13. Pratt P, Ives M, Lawton G, Simmons J, Radev N, Spyropoulou L, et al. Through the HoloLens™ looking glass: augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels. Eur Radiol Exp 2018 Jan;2(1):2 [FREE Full text] [doi: 10.1186/s41747-017-0033-2] [Medline: 29708204]
14. Zuo Y, Jiang T, Dou J, Yu D, Ndaro ZN, Du Y, et al. A Novel Evaluation Model for a Mixed-Reality Surgical Navigation System: Where Microsoft HoloLens Meets the Operating Room. Surg Innov 2020 Apr 10;27(2):193-202. [doi: 10.1177/1553350619893236] [Medline: 31920155]
15. Al Janabi HF, Aydin A, Palaneer S, Macchione N, Al-Jabir A, Khan M, et al. Effectiveness of the HoloLens mixed-reality headset in minimally invasive surgery: a simulation-based feasibility study. Surg Endosc 2020 Mar;34(3):1143-1149 [FREE Full text] [doi: 10.1007/s00464-019-06862-3] [Medline: 31214807]

16. Tepper O, Rudy H, Lefkowitz A, Weimer K, Marks S, Stern C, et al. Mixed Reality with HoloLens. Plastic and Reconstructive Surgery 2017;140(5):1066-1070. [doi: 10.1097/prs.0000000000003802]
17. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ 2017 Nov 17;10(6):549-559. [doi: 10.1002/ase.1696] [Medline: 28419750]
18. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3:i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
19. Messaging tool for NHS to support remote working during coronavirus outbreak. NHS Digital. 2020. URL: https://digital.nhs.uk/news-and-events/latest-news/messaging-tool-for-nhs-to-support-remote-working-during-coronavirus-outbreak [accessed 2020-05-24]
20. COVID-19 information governance advice for staff working in health and care organisations. NHSX. 2020. URL: https://www.nhsx.nhs.uk/covid-19-response/data-and-information-governance/information-governance/covid-19-information-governance-advice-health-and-care-professionals/ [accessed 2020-05-24]
21. Mann D, Chen J, Chunara R, Testa P, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
22. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
23. Callaghan A, Kinsman L, Cooper S, Radomski N. The factors that influence junior doctors' capacity to recognise, respond and manage patient deterioration in an acute ward setting: An integrative review. Aust Crit Care 2017 Jul;30(4):197-209. [doi: 10.1016/j.aucc.2016.09.004] [Medline: 27726906]
24. Rational use of personal protective equipment for coronavirus disease (COVID-19) and considerations during severe shortages. World Health Organization. 2020. URL: https://www.who.int/publications-detail/rational-use-of-personal-protective-equipment-for-coronavirus-disease-(covid-19)-and-considerations-during-severe-shortages [accessed 2020-05-27]
25. Jinadatha C, Villamaria F, Coppin J, Dale C, Williams M, Whitworth R, et al. Interaction of healthcare worker hands and portable medical equipment: a sequence analysis to show potential transmission opportunities. BMC Infect Dis 2017 Dec 28;17(1):800 [FREE Full text] [doi: 10.1186/s12879-017-2895-6] [Medline: 29281998]
26. Treibel T, Manisty C, Burton M, McKnight A, Lambourne J, Augusto J, et al. COVID-19: PCR screening of asymptomatic health-care workers at London hospital. Lancet Elsevier BV 2020;395(10237):1608-1610 [FREE Full text] [doi: 10.1016/s0140-6736(20)31100-4]
27. Evans S, Agnew E, Vynnycky E, Robotham JV. The impact of testing and infection prevention and control strategies on within-hospital transmission dynamics of COVID-19 in English hospitals. medRxiv Cold Spring Harbor Laboratory Press 2020 May 20:e [FREE Full text] [doi: 10.1101/2020.05.12.20095562]
28. Martin G, Khajuria A, Arora S, King D, Ashrafian H, Darzi A. The impact of mobile technology on teamwork and communication in hospitals: a systematic review. J Am Med Inform Assoc 2019 Apr 01;26(4):339-355. [doi: 10.1093/jamia/ocy175] [Medline: 30689893]
29. Haun N, Hooper-Lane C, Safdar N. Healthcare Personnel Attire and Devices as Fomites: A Systematic Review. Infect Control Hosp Epidemiol 2016 Nov;37(11):1367-1373. [doi: 10.1017/ice.2016.192] [Medline: 27609491]
30. Kwon JH, Burnham CD, Reske KA, Liang SY, Hink T, Wallace MA, et al. Assessment of Healthcare Worker Protocol Deviations and Self-Contamination During Personal Protective Equipment Donning and Doffing. Infect Control Hosp Epidemiol 2017 Sep 13;38(9):1077-1083 [FREE Full text] [doi: 10.1017/ice.2017.121] [Medline: 28606192]
31. Okamoto K, Rhee Y, Schoeny M, Lolans K, Cheng J, Reddy S, Centers for Disease ControlPrevention Epicenters Program. Impact of doffing errors on healthcare worker self-contamination when caring for patients on contact precautions. Infect Control Hosp Epidemiol 2019 May 20;40(5):559-565. [doi: 10.1017/ice.2019.33] [Medline: 30890193]
32. Houghton C, Meskell P, Delaney H, Smalle M, Glenton C, Booth A, et al. Barriers and facilitators to healthcare workers' adherence with infection prevention and control (IPC) guidelines for respiratory infectious diseases: a rapid qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Apr 21;4:CD013582. [doi: 10.1002/14651858.CD013582] [Medline: 32315451]

Abbreviations

COVID-19: coronavirus disease
EHR: electronic health record
MAC: media access control
MR: mixed reality
NHS: National Health Service
NIHR: National Institute for Health Research
PPE: personal protective equipment

QI: quality improvement

Edited by G Eysenbach; submitted 16.06.20; peer-reviewed by C Moro, F Incekara; comments to author 08.07.20; revised version received 13.07.20; accepted 27.07.20; published 14.08.20.

Please cite as:
Martin G, Koizia L, Kooner A, Cafferkey J, Ross C, Purkayastha S, Sivananthan A, Tanna A, Pratt P, Kinross J, PanSurg Collaborative Use of the HoloLens2 Mixed Reality Headset for Protecting Health Care Workers During the COVID-19 Pandemic: Prospective, Observational Evaluation
J Med Internet Res 2020;22(8):e21486
URL: http://www.jmir.org/2020/8/e21486/
doi:10.2196/21486
PMID:32730222

Original Paper

Real-Time Forecasting of the COVID-19 Outbreak in Chinese Provinces: Machine Learning Approach Using Novel Digital Data and Estimates From Mechanistic Models

Dianbo Liu[1,2*], PhD; Leonardo Clemente[1,2,3*], MSc; Canelle Poirier[1,2*], PhD; Xiyu Ding[1,4], MSc; Matteo Chinazzi[5], PhD; Jessica Davis[5], BSc; Alessandro Vespignani[5,6], PhD; Mauricio Santillana[1,2,4], PhD

[1]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, United States

[2]Department of Pediatrics, Harvard Medical School, Boston, MA, United States

[3]Tecnologico de Monterrey, Monterrey, Mexico

[4]Harvard TH Chan School of Public Health, Boston, MA, United States

[5]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA, United States

[6]ISI Foundation, Turin, Italy

[*]these authors contributed equally

Corresponding Author:
Mauricio Santillana, PhD
Computational Health Informatics Program
Boston Children's Hospital
300 Longwood Avenue
Landmark 5th Floor East
Boston, MA, 02215
United States
Phone: 1 (617) 919 1795
Email: msantill@g.harvard.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/9/e23996/

Abstract

Background: The inherent difficulty of identifying and monitoring emerging outbreaks caused by novel pathogens can lead to their rapid spread; and if left unchecked, they may become major public health threats to the planet. The ongoing coronavirus disease (COVID-19) outbreak, which has infected over 2,300,000 individuals and caused over 150,000 deaths, is an example of one of these catastrophic events.

Objective: We present a timely and novel methodology that combines disease estimates from mechanistic models and digital traces, via interpretable machine learning methodologies, to reliably forecast COVID-19 activity in Chinese provinces in real time.

Methods: Our method uses the following as inputs: (a) official health reports, (b) COVID-19–related internet search activity, (c) news media activity, and (d) daily forecasts of COVID-19 activity from a metapopulation mechanistic model. Our machine learning methodology uses a clustering technique that enables the exploitation of geospatial synchronicities of COVID-19 activity across Chinese provinces and a data augmentation technique to deal with the small number of historical disease observations characteristic of emerging outbreaks.

Results: Our model is able to produce stable and accurate forecasts 2 days ahead of the current time and outperforms a collection of baseline models in 27 out of 32 Chinese provinces.

Conclusions: Our methodology could be easily extended to other geographies currently affected by COVID-19 to aid decision makers with monitoring and possibly prevention.

(J Med Internet Res 2020;22(8):e20285) doi:10.2196/20285

KEYWORDS

COVID-19; coronavirus; digital epidemiology; modeling; modeling disease outbreaks; emerging outbreak; machine learning; precision public health; machine learning in public health; forecasting; digital data; mechanistic model; hybrid simulation; hybrid model; simulation

Introduction

First detected in Wuhan, China, in December 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection had rapidly spread by late January 2020 to all Chinese provinces and many other countries [1-4]. On January 30, 2020, the World Health Organization (WHO) issued a Public Health Emergency of International Concern (PHEIC) [5-8]; and on March 11th, the WHO declared the coronavirus disease (COVID-19) a pandemic [5]. By April 18, 2020, the virus had affected more than 2,300,000 people and caused the deaths of 150,000 in more than 180 countries [7].

In the last decade, methods that leverage data from internet-based data sources and data from traditional surveillance systems have emerged as a complementary alternative to provide near real-time disease activity estimates (eg, for influenza and dengue) [9-13]. Despite the fact that these methodologies have successfully addressed delays in the availability of health reports as well as case count data quality issues, developing predictive models for an emerging disease outbreak such as COVID-19 is an even more challenging task [14]. There are multiple reasons for this; for example, the availability of epidemiological information for this disease is scarce (there is no historical precedent about the behavior of the disease); the daily/weekly epidemiological reports that become available are frequently revised and corrected retrospectively to account for mistakes in data collection and reporting (a common practice in public health reports); and the presence of a diverse array of uncertainties about disease burden due in part to underreporting of cases [15].

Most efforts to estimate the time evolution of COVID-19 spread and the effect of public health interventions have relied on mechanistic models that parameterize transmission and epidemiological characteristics to produce forecasts of disease activity [16,17]. In contrast, only a limited number of studies have investigated ways to track COVID-19 activity, leveraging internet search data [1,13,18], and few to the best of our knowledge have combined internet-based data sources and mechanistic estimates to forecast COVID-19 activity [19].

We present a novel hybrid methodology that combines mechanistic and machine learning methodologies to successfully forecast COVID-19 in real time at the province level in China [20,21]. We used a data-driven approach to incorporate inputs from (a) official health reports from Chinese Center for Disease Control and Prevention (China CDC), (b) COVID-19–related internet search activity from Baidu, (c) news media activity reported by Media Cloud, and (d) daily forecasts of COVID-19 activity from the simulation epidemiological model GLEAM (global epidemic and mobility), a metapopulation mechanistic model [16]. Inspired by a methodology previously used to successfully forecast seasonal influenza in the United States at the state level [11] and previous methods to monitor emerging outbreaks [22,23], our method is capable of reliably forecasting COVID-19 activity even when limited historical disease activity observations are available. From a methodological perspective, the novelty in our approach comes from a clustering technique that enables the exploitation of geospatial synchronicities of COVID-19 activity across Chinese provinces and a data augmentation technique to mitigate the scarcity of historical data for model training.

Methods

Experimental Design

Our method was designed for forecasting COVID-19 2 days ahead of the current time. We used as inputs the following data sources: COVID-19 activity reports from China CDC; internet search frequencies from Baidu; a number of related news reports from 311 media sources, as reported by the Media Cloud platform; and COVID-19 daily forecasts from a metapopulation mechanistic model. Our machine learning methodology also used a clustering and data augmentation technique. We provide details about data sources and statistical methods in the following sections.

Data Sources

Daily Reports of COVID-19

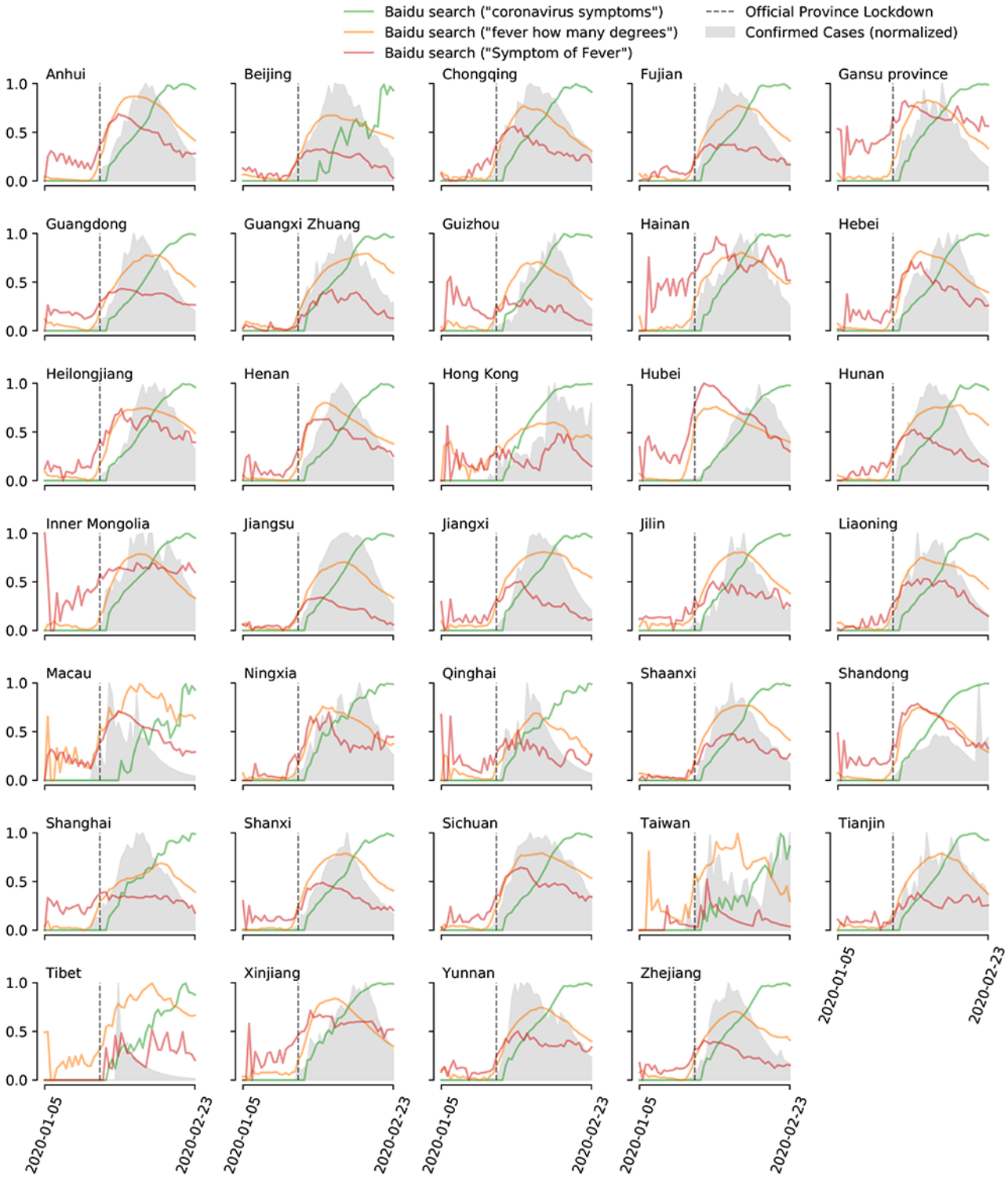
Case counts of COVID-19 were obtained from China CDC. These data are curated and publicly available via the Models of Infectious Disease Agent Study (MIDAS) association [24]. All data were collected on the original date they became available. Indeed, case counts released by China CDC can be revised, up to several weeks later. In this study, we only used unrevised data, which is the real case scenario to produce real-time estimates. The reports, available for all the provinces, include various activity trends such as new diagnosed cases, new suspected cases, and new reported deaths. For our study, we selected the number of confirmed cases as the epidemiological target and collected activity reports from January 10, 2020, to February 21, 2020.

Baidu Internet Search Activity: Data Exclusion

We collected the daily search fraction for three different COVID-19–related search terms in Mandarin (“COVID-19 symptoms” [“新冠肺炎症状”], “how many degree is fever” [“多少度算发烧”], and “symptoms of fever” [“发烧症状”]). These terms were selected based on their correlation and potential association with case counts of COVID-19 [25] and collected individually for each province from January 1, 2020, to February 21, 2020. Our decision to use internet activity as a source of information is based on the hypothesis that search frequencies from COVID-19–related keywords reflect, to an extent, the number of people presenting symptoms related to COVID-19 before their arrival at a clinic. Given Baidu imposes limits to data access for researchers, we were unable to conduct

a broad analysis on a wide range of keywords. A visualization of the Baidu search term time series can be seen in Figure 1.

Figure 1. Visualization of the evolution of coronavirus disease (COVID-19) cases and Baidu search trends. The evolution of COVID-19 cases is represented in gray and Baidu search trends in green and orange. All-time series have been smoothed for visualization purposes.



News Reports

An online open-source platform called Media Cloud, which allows the tracking and analysis of media for any topic of interest through the matching of keywords, was used. We obtained volumes of the number of news articles available over time from a collection of 311 Chinese media websites using the keywords "coronavirus," "COVID-19," "2019-nCoV," "pneumonia," "fever," "cough," and the name of each province to generate province-specific news activity trends. Media data from January 1, 2020, to February 21, 2020 were collected and used as additional source information.

Global Epidemic and Mobility Model

The global epidemic and mobility model, GLEAM, is an individual-based, stochastic, and spatial epidemic model [26-28]

that has been used to simulate the early stages of the COVID-19 epidemic in mainland China and across the world [16]. GLEAM is based on a metapopulation approach in which the world population is divided into subpopulations centered around major transportation hubs (usually airports). Over 3000 subpopulations in about 200 different countries and territories are included in the model. The subpopulations are connected by short-range commuting and long-range travel networks that determine the flow of individuals traveling daily among them. Short-range mobility patterns (eg, daily commuting) are derived from data collected from the National Statistical Offices of 30 countries on five continents [26]. In addition, for the COVID-19 epidemic, mobility variations in mainland China are further calibrated using deidentified and aggregated domestic population movement data as derived from Baidu Location-Based Services. The airline transportation data consider daily origin-destination traffic flows obtained from the Official Aviation Guide and the International Air Transport Association databases (updated in 2019), and accounting for travel restrictions in 2020. Within each subpopulation, the human-to-human transmission of COVID-19 is modeled using a compartmental representation of the disease where each individual can occupy one of the following four states: susceptible (S), latent (L), infectious (I), and removed (R). Susceptible individuals can acquire the virus through contacts with individuals in the infectious state, and become latent, meaning they are infected but cannot transmit the infection yet. Latent individuals progress to the infectious stage with a rate inversely proportional to the latent period. Infectious individuals progress into the removed stage with a rate inversely proportional to the infectious period. Removed individuals represent those who can no longer infect others, meaning they were isolated, hospitalized, died, or have recovered.

The model produces an ensemble of possible epidemic scenarios providing epidemic indicators, such as the number of newly generated infections and deaths in each subpopulation. The model is initialized by a starting date of the epidemic between November 15, 2019, and December 1, 2019, with 20 to 40 cases caused by zoonotic exposure [29-32]. The transmission dynamic is calibrated by using an Approximate Bayesian Computation approach to estimate the posterior distribution of the basic reproductive number R_0 that uses as evidence the detection of infections imported from China at international locations across the world [33-37]. A sensitivity analysis has been performed on the initial conditions of the model considering different values for the mean latency period (range 3-6 days), the mean infectious period (range 2-8 days), the generation time (range 6-11 days), and the initial number of zoonotic cases (range 20-80). The calibrated model is then used to generate the out-of-sample ensemble of stochastic epidemic evolutions across mainland China.

Statistical Analysis

Aggregation of Daily Reports

To enhance signal and reduce noise, we aggregated case count, search volumes, and media article count for each $\delta t = 2$ days window.

As COVID-19 is an emerging outbreak, the amount of epidemiological information, either official or unofficial, is low, and thus, limits our capacity to build predictive models. To maximize usage of data, we applied the strategies below.

Clustering

We clustered the 32 provinces into several groups and trained a model for each group. Clustering and model retraining processes were repeated on every single new prediction date. To determine the similarities in outbreak patterns across Chinese provinces, we calculated the pairwise correlation matrix for confirmed COVID-19 cases by using all historical data available. Then, based on similarity matrix, provinces were clustered by using complete linkage hierarchical clustering, which is an agglomerative hierarchical clustering method, creating clusters based on most dissimilar pairs [38]. The number of clusters K was determined by choosing the K, thereby maximizing the Calinski-Harabasz index [39]. Our clustering method gained higher stability when more data points were available for clustering [40]. More details of the clustering method are presented in Multimedia Appendix 1 [41-43].

Data Augmentation

We conducted data augmentation by using a bootstrap method to resample each data point of the training data set. We made 100 bootstrap samples for each data point to which we added a random Gaussian noise with a mean of 0 (SD 0.01). Due to the stochasticity of both the clustering algorithm and the model training processing, on each prediction day, we run the whole clustering-training process 20 times and take an average of the outputs as our final prediction. Our multistep approach may introduce stochasticity in three different steps: (a) the clustering process, (b) the data augmentation process, and (c) the regression algorithm. To ensure robustness of our prediction results, the whole process (from clustering to out-of-sample prediction) on each prediction date was repeated at least 20 times and the ensemble (via an averaging approach) predictions were reported as the final prediction. We chose to use an empirical approach to explore whether the number of computational experiments were sufficient to lead to a stable performance. In order to achieve this, we conducted ensemble prediction experiments using realizations from 1 to 50 prediction efforts. We documented the performance of these ensemble predictions using root mean square error (RMSE) and correlation in (Multimedia Appendix 2, Table S1). The performance of the ARGONet + GLEAM method plateaued after about 10-15 realizations as seen on this table. Therefore, we concluded that 20 realizations of our algorithm was an adequate number to ensure robustness and stability of the prediction while not imposing too much computational burden.

Predictive Model

For our prediction task, we fitted a LASSO (least absolute shrinkage and selection operator) multivariable regularized linear model for every data set generated from our clustering and augmentation steps at time t.

The LASSO technique minimizes the mean squared error between observations and predictions subject to a L1 norm constraint (more details of this method are provided in

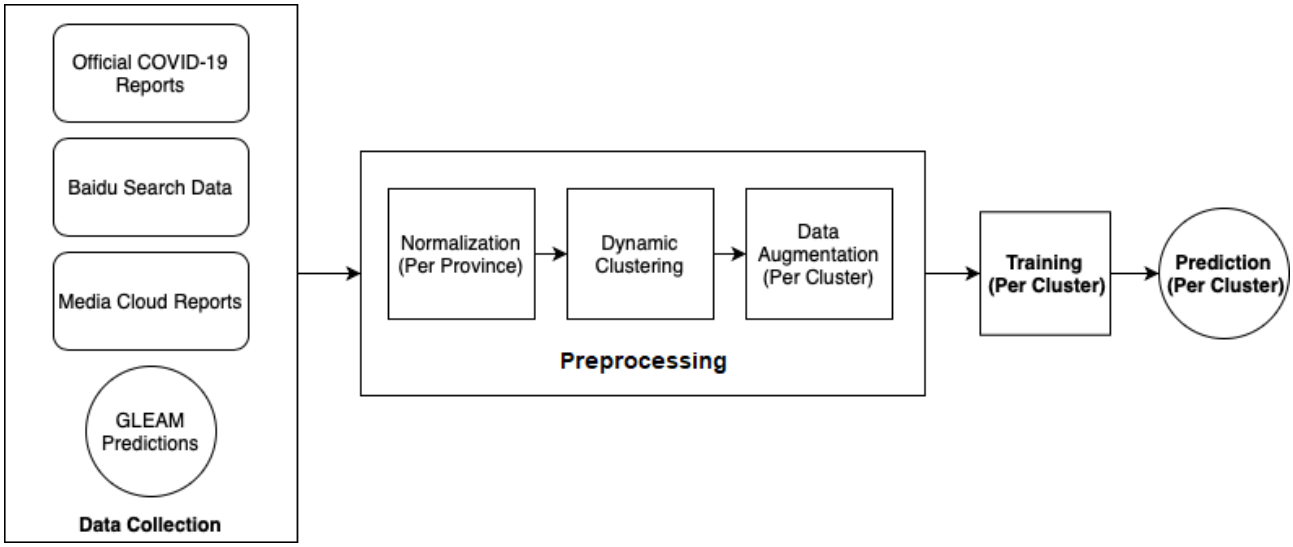
Multimedia Appendix 1). The number of new confirmed COVID-19 cases for the next bi-day can be then expressed as:

where $y_{T+\delta t}$ is the estimate at date $T + \delta t$; $\delta t = 2$ days; y_T is the number of cases at date T; S_T is the search volume at date T; M_T is the number of media articles at date T; D_T is the number of deaths at date T; C_T is the number of cumulative cases at date T; and $_{T+\delta t}$ is the normally distributed error term.

Models were dynamically recalibrated, similar to the method presented by Santillana et al [44] and Lu et al [11]. Our method, ARGONet + GLEAM, was implemented in an R 3.5.3 environment with a glmnet 3.0-2 library.

A summary of our method can be seen in Figure 2.

Figure 2. Summary of the methods used to obtain our coronavirus disease (COVID-19) estimates. GLEAM: global epidemic and mobility.



Performance of Model and Relevance of Predictors

Two different metrics were used to measure the performance of ARGONet + GLEAM: (1) the RMSE and (2) the Pearson correlation. To assess the predictive power of our methodology, we compared our performance against the following models:

1. Persistence rule (baseline): a rule-based model that uses the new case count at date T as an estimate of the prediction for $T+\delta t$ so that $yT+\delta t = yT$
2. Autoregressive (AR): a simple AR model built on COVID-19 cases that occurred in the previous three AR lags (2-day reports) (see Multimedia Appendix 1 for more information on this model)
3. ARGONet: an alternate version of our methodology that does not include any mechanistic information but including clustering and data augmentation approaches.

As linear models are used in this study, the relevance of predictors in predicting new cases can be defined thanks to the associated factor of each term in the trained model. As all data were normalized using the z-score (strictly within the training data sets) during training and prediction, the associated factor can be approximately understood as how many standard deviations the predicted new cases $y_T + \delta t$ will change if 1 standard deviation changes in the predictor.

Data Sharing

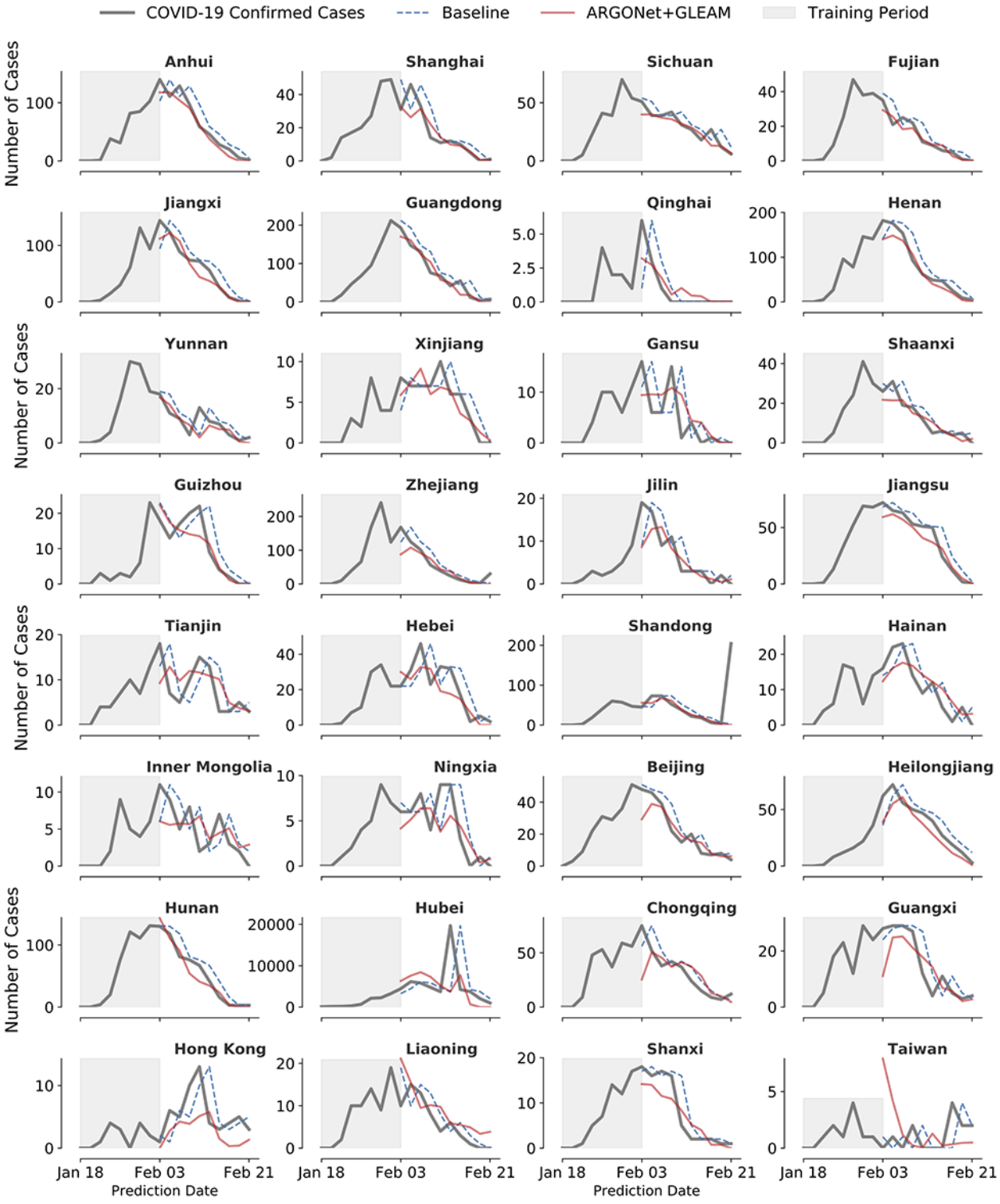
All codes and data will be made available via the Harvard dataverse.

Results

We produced 2-day-ahead (strictly out-of-sample) and real-time COVID-19 forecasts for 32 Chinese provinces for the time period spanning February 3, 2020, to February 21, 2020. A visual representation of our out-of-sample model forecasts is shown in Figure 3 along with the subsequently observed COVID-19 cases, as reported by China CDC.

Our results show that ARGONet + GLEAM outperforms the persistence model in 27 out of 32 Chinese provinces. Even in provinces where ARGONet + GLEAM failed to produce improvements to the baseline model, our model produced reasonable disease estimates as seen in Figure 3. These provinces include Shanxi, Liaoning, Taiwan, Hong Kong, and Guangxi (the latter three with very different administration, and likely health care, systems compared to the rest of the provinces).

Figure 3. Graphical visualization of the estimates obtained by ARGONet + GLEAM. The number of new confirmed cases for coronavirus disease (COVID-19), as reported by China CDC (solid black), along with ARGONet + GLEAM (solid red) 2-day, ahead-of-time estimates between February 3, 2020, to February 21, 2020. As a comparison, the dotted blue line represents the persistence model.

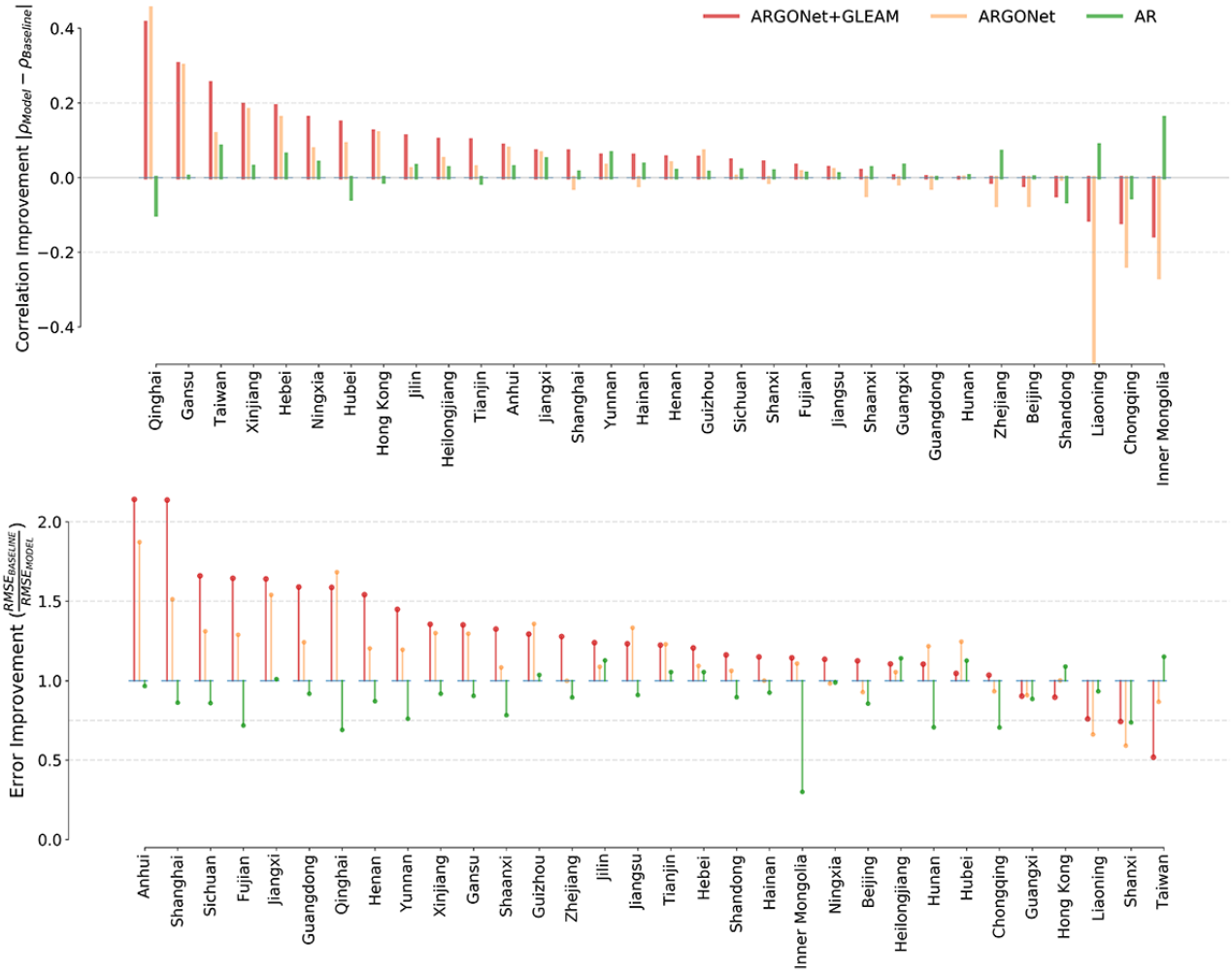


Experimental Design AR Model

We analyzed the performance of models built using only local, province-level epidemiological data as input. We generated an AR model for each province, built on COVID-19 cases that occurred in the previous three AR lags (ie, the previous three 2-day reports), and compared our estimates with the baseline. Our results, presented in Figure 4 (also see Tables S2 and S3 in Multimedia Appendix 2 for a detailed description of our model results), labeled AR, show that the AR model's predictive power was overall inferior to baseline performance, with exception to Jilin, Tianjin, Hebei, Hubei, and Heilongjiang. Subsequently, we incorporated local disease-related internet search information from Baidu and news alert data from Media Cloud as inputs to build ARGO-type models [9]. These

ARGO-type models showed marginal predictive power improvements when compared with AR models and only outperformed the baseline in seven provinces.

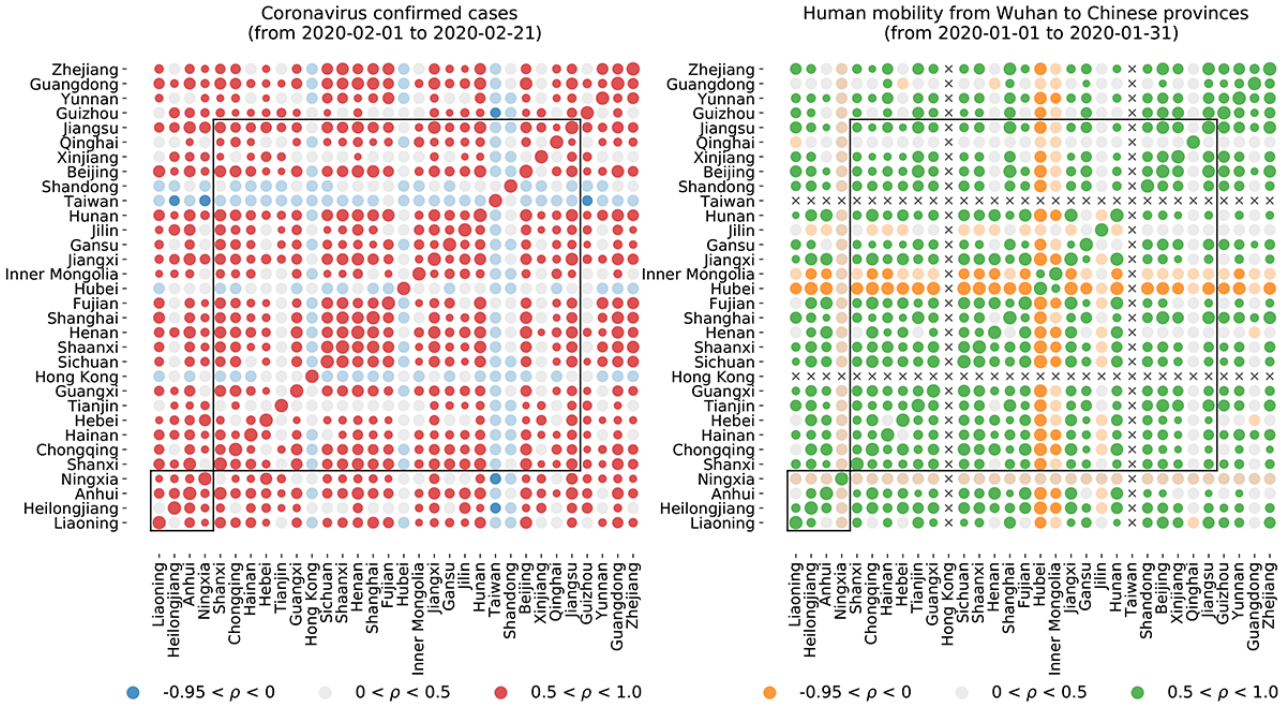
Figure 4. Graphical visualization of the models' performances. Comparison of the improvement in terms of root mean square error (RMSE) (top) and Pearson correlation (bottom) for each model used in the study. To facilitate comparison between model scores in each province in terms of RMSE, we normalized the RMSE score of each model by the baseline's RMSE and visualized its inverse value. In this way, scores above one imply an improvement (RMSE reduction), whereas a score below one implies the model had a bigger RMSE in comparison to the baseline. In the case of correlation, we plotted the difference between the absolute values between each model's correlation and the baseline. Each panel is ordered, from left to right, based on the metric performance of ARGONet + GLEAM (solid red). AR: autoregressive.



Dynamic Clustering of Chinese Provinces

Based on prior work on influenza activity prediction [11], we added historical COVID-19 activity information for all Chinese provinces to the input of our local models. We calculated the pairwise correlation matrix for confirmed COVID-19 cases between all Chinese provinces, between February 1 and February 21, 2020 (Figure 5). Our results showed that most of the provinces experienced similar epidemic trends. To build our (clustered) predictive models, we combined the data available from several provinces with similar trends (in terms of correlation, which was strictly calculated within our training period at the time-step of prediction). The clustering modeling approach, which incorporated internet-based data sources as the ARGO-type models, produced forecasts that led to error reductions for 17 out of 32 provinces compared to the persistence model and improved correlation values in 20 out of 32 provinces.

Figure 5. Visualization of the pairwise correlation matrices of confirmed cases and human mobility from Wuhan to each Chinese province. During the period of January, we can see a similar trend of mobility for a big cluster of provinces as well as a similar trend of number of confirmed cases for the period of February.



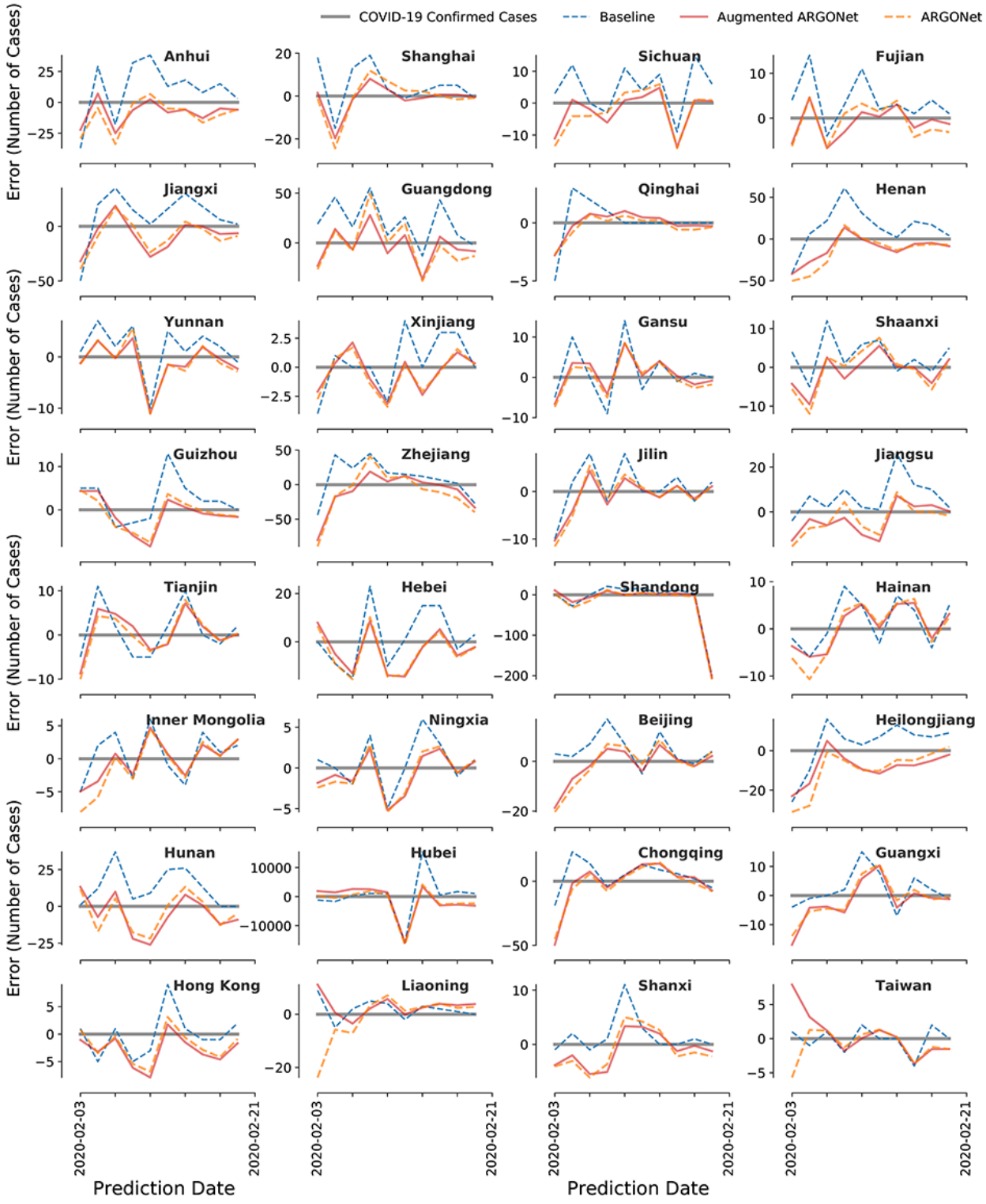
Data Augmentation

As an additional way to increase the number of observations in the training set of each cluster, we implemented a data augmentation technique. This process consisted of generating new observations via a Bootstrap method and addition of random Gaussian noise to every randomly selected observation.

ARGONet Model

The results of incorporating both clustering and augmentation techniques can be seen in Figure 4 and a visualization of the errors can be seen in Figure 6. For simplicity, we labeled these predictions ARGONet, even though this implementation of ARGONet is an enhanced version specifically designed for emerging outbreaks where data are scarce. In terms of RMSE, our results show that ARGONet's predictive power was able to outperform AR and the persistence model in 25 of the 32 Chinese provinces. In terms of correlation, ARGONet outperformed the baseline (persistence) model in 18 provinces.

Figure 6. Visualization of the errors. Graphical visualization of the out-of-sample coronavirus disease (COVID-19) error (ŷ–y) between February 3, 2020, and February 21, 2020.



ARGONet + GLEAM Model

We included forecasts produced by mechanistic model as an additional input in our models (prior to the clustering and augmentation steps). The results of incorporating these estimates can be seen in Figure 4 with the name of ARGONet + GLEAM and a visualization of the errors can be seen in Figure 6. Our results show that the inclusion of mechanistic model estimates improved ARGONet's predictive power across most provinces. ARGONet + GLEAM led to error reductions in 27 out of 32 provinces compared to the baseline. In terms of correlation, it improved in 26 out of 32 provinces. Provinces like Qinghai, Hunan, and Jiangxi showed the biggest improvement, whereas Taiwan, Hong Kong, Shanxi, and Liaoning did not display error reductions.

Visualization of the Results

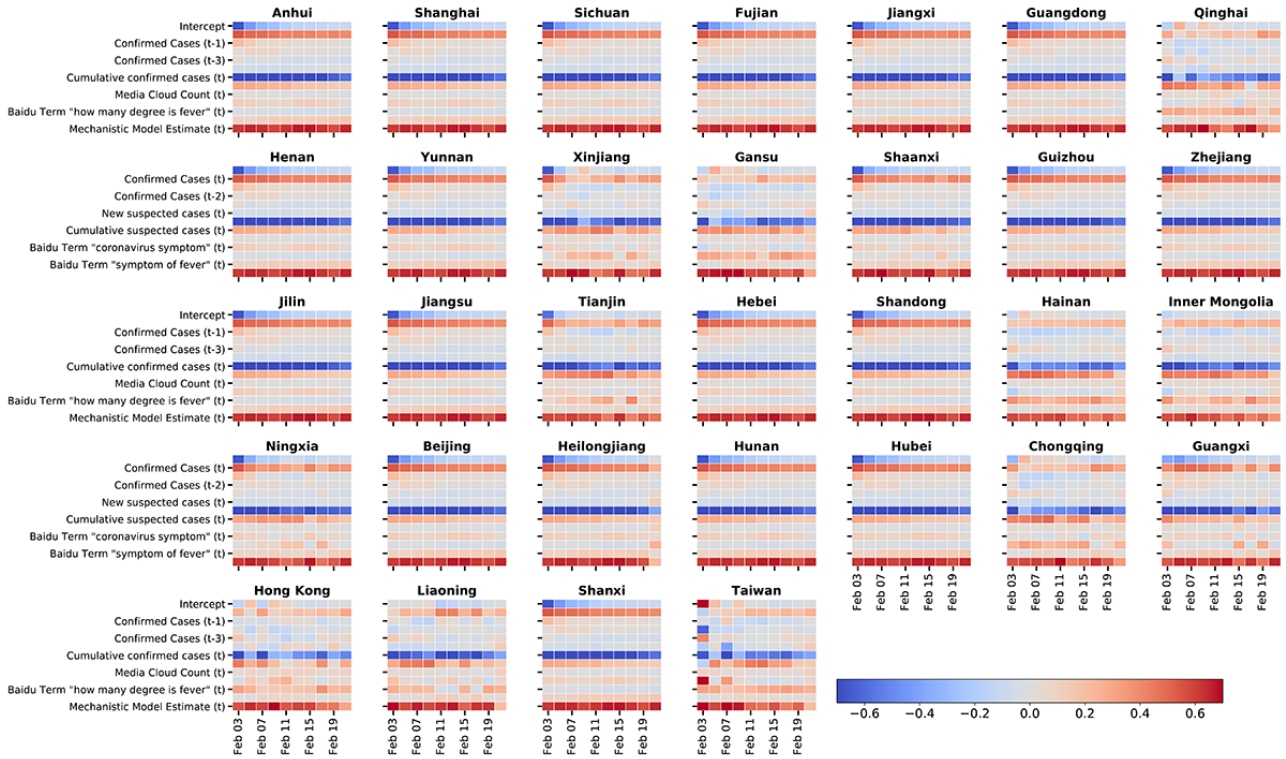
As an alternative way to visualize ARGONet + GLEAM's predictive performance, we plotted a map with Chinese provinces (Figure 7), color coded based on the improvement shown in Figure 4. From a geographical perspective, the provinces where ARGONet + GLEAM had the most improvement (Anhui, Jiangxi, Fujian, Sichuan, and Guangdong) were located in south central China. Shanxi, Liaoning, Taiwan, Hong Kong, and Guangxi are the provinces where our models were not able to reduce the error compared to the baseline. While ARGONet + GLEAM's performance in these provinces was not superior to the baseline, its predictions were within a reasonable range, as seen in Figures 2 and 3. We were not able to perform any analysis on Tibet, one of the largest provinces in China, and Macau given their low count of detected COVID-19 cases.

Figure 7. Geographical visualization of the relative improvement of ARGONet + GLEAM compared to the baseline. Chinese provinces that show an increase in performance relative to the baseline are shaded green, while provinces that did not perform better than our baseline are shaded purple. Provinces with the highest improvement (Anhui, Shanghai, Sichuan, Fujian, Jiangxi, Guangdong, and Qinghai) and underperformance (Taiwan, Shanxi, Liaoning, Hong Kong, and Guangxi) are identified by a red dot over the province.

Analysis of the Importance of the Sources of Information Over Time

To minimize the prediction errors in our estimates, the dynamic design of our methodology utilizes different sources of information as needed over time. This means that for each province (or group of provinces within a cluster), we can quantify the predictive power of different features used in our models as time evolves. Our analysis, visualized in Figure 8, shows that historical COVID-19 confirmed cases and suspected cases were consistently relevant sources of information over most of the study period. Internet-based search terms from Baidu were also frequently used. Daily news counts were used by our models in a selected number of provinces. However, for many of these provinces, the importance of media article counts decrease over time. Estimates from mechanistic models contributed to our model prediction, especially in early February 2020.

Figure 8. Graphical visualization of the relevance of data sources. Time evolution of the value (averaged over the 20 experiments) of the linear coefficients for the features used in our methodology, visualized per province. Every heatmap includes the same number of features (rows) and is organized in the same order.



Discussion

Principal Findings

We presented a methodology capable of producing meaningful and reliable short-term (2 days ahead) forecasts of COVID-19 activity, at the province level in China, by combining information from reports from China CDC, internet search trends, news article trends, and information from mechanistic models. Our approach is capable of overcoming multiple challenges characteristic of emerging outbreaks caused by novel pathogens. These challenges include the lack of historical disease activity information to calibrate models, the low volume of case count data, and the inherent delay in gaining access to data. Methodologically speaking, our method maximizes the use of a limited number of observations as the outbreak unfolded by (a) choosing an appropriate aggregation time-window (2 days) to improve the signal-to-noise ratio, (b) leveraging synchronicities in the spatiotemporal trends in COVID-19 across provinces to produce cluster-specific models of prediction, and (c) using data augmentation methods to increase stability in the training of our models.

Previous methods, such as the ARGONet model [11,45], have been shown to make accurate real-time prediction at the state level in the United States for seasonal infectious diseases such as influenza. In addition, Chinazzi et al [16] showed that it was possible to estimate the evolution of an emerging outbreak using a mechanistic model. Nevertheless, as far as we know, reliable real-time methodologies to forecast new case counts for an emerging disease outbreak remained an unsolved problem. In this study, we showed that a dynamically trained machine learning model can accurately produce real-time estimates for COVID-19 outbreaks.

In terms of prediction error, our proposed methodology, ARGONet + GLEAM, was able to outperform the persistence model in 27 out of 32 provinces. While our method does not show prediction error improvements in Guangxi, Liaoning, Shanxi, Taiwan, and Hong Kong, our forecasts are still within range in all provinces except for Taiwan, where very few cases were reported during the time period of this study. It is important to note that Taiwan, Hong Kong, and Guangxi have different administrative (and likely health care) systems compared with the other provinces. This could explain the differences in COVID-19 trends in these regions and could help explain why our models do not seem to add value to the persistence model. Features studies should investigate if incorporating disease activity estimates from other mechanistic models, likely designed with different assumptions and mathematical formulations, could lead to further improvements.

We were unable to identify an accurate (daily) parametrization of changes in human mobility due to the widespread local lockdowns during the period of our study (February 3-21, 2020), and thus, we did not include this data source as a potential predictor. Future studies may incorporate (high temporal resolution) human mobility data as a modulator of transmission and predictor of disease activity. When looking at the entire time period of this study, however, we observed that the data-driven clustering of provinces used in our approach and based on COVID-19 activity appears to have similarities with the clustering one would obtain from using human mobility data made available by Baidu (Figure 5). This result aligns with

the conclusion of other available studies that found that the time evolution of the COVID-19 outbreak in China was significantly influenced by changes in human mobility (consequence of public health interventions) [16,17,46], and associated with the percentage of people traveling from Wuhan in the early stages.

Limitations

One limitation of our study is that during the test time period of our methods a consistent decrease in COVID-19 cases (due to strong public health interventions) was observed and thus our methods could not be tested for their ability to identify the epidemic peaks across provinces. The brevity of the COVID-19 epidemic outbreak in Chinese provinces was the limiting factor for this as the observations that corresponded to the growth phase of the outbreak were used for training purposes. Future model implementations in other locations where the growth phase has spanned longer time periods, like New York, United States, should investigate the ability of our models to properly identify peaks.

Conclusions

Our findings suggest that it is possible to use very limited amounts of data from multiple data sources to conduct real-time forecasting in the early stage of an emerging outbreak. We believe that our method, ARGONet + GLEAM, could prove to be useful for public health officials to monitor (and perhaps prevent) the spread of the virus [8,11,25,47]. As the SARS-CoV-2 virus continues to spread around the world, extensions of our methods could be implemented to provide timely and reliable disease activity estimates to decision makers.

Acknowledgments

We thank Dr Wei Luo for his assistance and guidance on the interpretation of mobility data for Chinese provinces. MC and AV acknowledge support from Google Cloud Healthcare and Life Sciences Solutions via the GCP research credits program.

CP, AV, and MS were partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM130668. MC and AV report grants from Metabiota Inc, outside the submitted work. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

DL, LC, CP, AV, and MS conceived and designed the study. DL, LC, CP, XD, and MC collected the different data sources. MC, JD, and AV produced predictions using the GLEAM modeling platform. DL, LC, and CP implemented the ARGONet + GLEAM methodology. DL, LC, CP, and MS analyzed the results. DL, LC, CP, and MS wrote the first draft of the manuscript. All authors contributed to and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobility data correlation.
[DOCX File , 18 KB - jmir_v22i8e20285_app1.docx]

Multimedia Appendix 2
Supplementary tables.
[DOCX File , 24 KB - jmir_v22i8e20285_app2.docx]

References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207. [doi: 10.1056/nejmoa2001316]
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017]
3. Chan J, Yuan S, Kok K, To K, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The Lancet 2020 Feb;395(10223):514-523 [FREE Full text] [doi: 10.1016/S0140-6736(20)30154-9]
4. Gilbert M, Pullano G, Pinotti F, Valdano E, Poletto C, Boelle P. Preparedness and vulnerability of African countries against introductions of 2019-nCoV. medRxiv 2020:e [FREE Full text] [doi: 10.1101/2020.02.05.20020792]
5. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). World Health Organization. 2020. URL: https://www.who.int/news-room/detail/

30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov) [accessed 2020-02-18]
6. Du Z, Wang L, Cauchemez S, Xu X, Wang X, Cowling BJ. Risk of 2019 novel coronavirus importations throughout China prior to the Wuhan quarantine. medRxiv 2020:e [FREE Full text] [doi: 10.1101/2020.01.28.20019299]
7. Sun K, Chen J, Viboud C. Early epidemiological analysis of the 2019-nCoV outbreak based on a crowdsourced data. medRxiv 2020:e [FREE Full text] [doi: 10.1101/2020.01.31.20019935]
8. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. The Lancet 2020 Feb;395(10223):470-473. [doi: 10.1016/s0140-6736(20)30185-9]
9. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. Proc Natl Acad Sci U S A 2015 Nov 24;112(47):14473-14478 [FREE Full text] [doi: 10.1073/pnas.1515373112] [Medline: 26553980]
10. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. PLoS Comput Biol 2015 Oct 29;11(10):e1004513 [FREE Full text] [doi: 10.1371/journal.pcbi.1004513] [Medline: 26513245]
11. Lu FS, Hattab MW, Clemente CL, Biggerstaff M, Santillana M. Improved state-level influenza nowcasting in the United States leveraging Internet-based data and network approaches. Nat Commun 2019 Jan 11;10(1):147 [FREE Full text] [doi: 10.1038/s41467-018-08082-0] [Medline: 30635558]
12. Cleaton JM, Viboud C, Simonsen L, Hurtado AM, Chowell G. Characterizing Ebola Transmission Patterns Based on Internet News Reports. Clin Infect Dis 2016 Jan 01;62(1):24-31 [FREE Full text] [doi: 10.1093/cid/civ748] [Medline: 26338786]
13. Li C, Chen L, Chen X, Zhang M, Pang C, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Eurosurveillance 2020;25(10):e [FREE Full text] [doi: 10.2807/1560-7917.es.2020.25.10.2000199]
14. Lipsitch M, Santillana M. Enhancing Situational Awareness to Prevent Infectious Disease Outbreaks from Becoming Catastrophic. Curr Top Microbiol Immunol 2019;424:59-74. [doi: 10.1007/82_2019_172] [Medline: 31292726]
15. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 2020 May 01;368(6490):489-493 [FREE Full text] [doi: 10.1126/science.abb3221] [Medline: 32179701]
16. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
17. Lai S, Ruktanonchai N, Zhou L, Prosper O, Luo W, Floyd J. Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. medRxiv 2020:e [FREE Full text] [doi: 10.1101/2020.03.03.20029843]
18. Jung S, Akhmetzhanov A, Hayashi K, Linton N, Yang Y, Yuan B, et al. Real-Time Estimation of the Risk of Death from Novel Coronavirus (COVID-19) Infection: Inference Using Exported Cases. J Clin Med 2020 Feb 14;9(2):9 [FREE Full text] [doi: 10.3390/jcm9020523] [Medline: 32075152]
19. Gore R, Lynch C. Virginia County COVID-19 daily case total forecaster. VMASC. 2020. URL: https://vmasc.shinyapps.io/va-county-covid-forecast/ [accessed 2020-07-16]
20. Eldabi T. Hybrid simulation: Historical lessons, present challenges and futures. : IEEE; 2016 Presented at: WSC '16: Proceedings of the 2016 Winter Simulation Conference; 11-14 Dec 2016; Washington, DC, USA URL: https://ieeexplore.ieee.org/document/7822192 [doi: 10.1109/WSC.2016.7822192]
21. Giabbanelli P. Solving challenges at the interface of simulation and big data using machine learning. : IEEE; 2019 Presented at: 2019 Winter Simulation Conference (WSC); 8-11 Dec 2019; National Harbor, MD, USA, USA p. 572-583 URL: https://ieeexplore.ieee.org/document/9004755 [doi: 10.1109/wsc40007.2019.9004755]
22. Aiken E, McGough S, Majumder M, Wachtel G, Nguyen A, Viboud C. Real-time Estimation of Disease Activity in Emerging Outbreaks using Internet Search Information. medRxiv 2019:e [FREE Full text] [doi: 10.1101/19010470v1]
23. McGough SF, Brownstein JS, Hawkins JB, Santillana M. Forecasting Zika Incidence in the 2016 Latin America Outbreak Combining Traditional Disease Surveillance with Search, Social Media, and News Report Data. PLoS Negl Trop Dis 2017 Jan 13;11(1):e0005295 [FREE Full text] [doi: 10.1371/journal.pntd.0005295] [Medline: 28085877]
24. Models of Infectious Disease Agent Study Association. midas-network / COVID-19. GitHub. URL: https://github.com/midas-network/COVID-19/tree/master/data/cases [accessed 2020-08-07]
25. Phan LT, Nguyen TV, Luong QC, Nguyen TV, Nguyen HT, Le HQ, et al. Importation and Human-to-Human Transmission of a Novel Coronavirus in Vietnam. N Engl J Med 2020 Feb 27;382(9):872-874. [doi: 10.1056/nejmc2001272]
26. Balcan D, Gonçalves B, Hu H, Ramasco JJ, Colizza V, Vespignani A. Modeling the spatial spread of infectious diseases: the GLobal Epidemic and Mobility computational model. J Comput Sci 2010 Aug 01;1(3):132-145 [FREE Full text] [doi: 10.1016/j.jocs.2010.07.002] [Medline: 21415939]
27. Gomes MFC, Pastore Y Piontti A, Rossi L, Chao D, Longini I, Halloran ME, et al. Assessing the international spreading risk associated with the 2014 west african ebola outbreak. PLoS Curr 2014 Sep 02;6:e [FREE Full text] [doi: 10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5] [Medline: 25642360]

28. Zhang Q, Sun K, Chinazzi M, Pastore Y Piontti A, Dean NE, Rojas DP, et al. Spread of Zika virus in the Americas. Proc Natl Acad Sci U S A 2017 May 30;114(22):E4334-E4343 [FREE Full text] [doi: 10.1073/pnas.1620161114] [Medline: 28442561]
29. Rambaut A. Phylogenetic analysis of 23 nCoV-2019 genomes, 2020-01-23. Virological.org. 2020. URL: https://virological.org/t/phylogenetic-analysis-of-23-ncov-2019-genomes-2020-01-23/335 [accessed 2020-02-05]
30. Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly CA, Riley S, et al. Report 3: Transmissibility of 2019-nCoV. Imperial College London COVID-19 Response Team. 2020 Jan. URL: https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/news--wuhan-coronavirus [accessed 2020-02-05]
31. Andersen K. Estimates of the clock and TMRCA for 2019-nCoV based on 27 genomes. Virological.org. 2020. URL: http://virological.org/t/clock-and-tmrca-based-on-27-genomes/347 [accessed 2020-02-05]
32. Bedford T, Neher R, Hadfield J, Hodcroft E, Ilcisin M, Muller N. Genomic analysis of nCoV spread. Situation report 2020-01-23. Nextstrain. 2020. URL: https://nextstrain.org/narratives/ncov/sit-rep/2020-01-23 [accessed 2020-02-05]
33. Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. Approximate Bayesian computation. PLoS Comput Biol 2013 Jan 10;9(1):e1002803 [FREE Full text] [doi: 10.1371/journal.pcbi.1002803] [Medline: 23341757]
34. Niehus R, De Salazar PM, Taylor A, Lipsitch M. Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers. medRxiv 2020 Feb 18:e [FREE Full text] [doi: 10.1101/2020.02.13.20022707] [Medline: 32511442]
35. De Salazar PM, Niehus R, Taylor A, Buckee C, Lipsitch M. Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases. medRxiv 2020 Feb 11:e [FREE Full text] [doi: 10.1101/2020.02.04.20020495] [Medline: 32511458]
36. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. The Lancet Digital Health 2020 Apr;2(4):e201-e208. [doi: 10.1016/s2589-7500(20)30026-1]
37. Pinotti F, Di Domenico L, Ortega E, Mancastroppa M, Pullano G, Valdano E. Lessons learnt from 288 COVID-19 international cases: importations over time, effect of interventions, underdetection of imported cases. medRxiv 2020:e [FREE Full text] [doi: 10.1101/2020.02.24.20027326]
38. Defays D. An efficient algorithm for a complete link method. The Computer Journal 1977 Apr 01;20(4):364-366. [doi: 10.1093/comjnl/20.4.364]
39. Calinski T, Harabasz J. A dendrite method for cluster analysis. Comm in Stats 1974;3(1):1-27. [doi: 10.1080/03610927408827101]
40. Henelius A, Puolamäki K, Boström H, Papapetrou P. Clustering with Confidence: Finding Clusters with Statistical Guarantees. arXiv 2016:e [FREE Full text]
41. Sasirekha K, Baby P. Agglomerative Hierarchical Clustering Algorithm - A Review. International Journal of Scientific and Research Publications 2013;3(3):1-3 [FREE Full text] [doi: 10.1007/978-1-4419-9863-7_100033]
42. Seal HL. Studies in the History of Probability and Statistics. XV The historical development of the Gauss linear model. Biometrika 1967;54(1-2):1-24. [doi: 10.1093/biomet/54.1-2.1]
43. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological) 1996;58(1):267-288. [doi: 10.1111/j.2517-6161.1996.tb02080.x]
44. Santillana M, Nsoesie E, Mekaru S, Scales D, Brownstein J. Using clinicians' search query data to monitor influenza epidemics. Clin Infect Dis 2014 Nov 15;59(10):1446-1450 [FREE Full text] [doi: 10.1093/cid/ciu647] [Medline: 25115873]
45. Poirier C, Hswen Y, Bouzille G, Cuggia M, Lavenu A, Brownstein J. Influenza forecasting for the French regions by using EHR web and climatic data sources with an ensemble approach ARGONet. medRxiv 2019:e [FREE Full text] [doi: 10.1101/19009795]
46. Kraemer MUG, Yang C, Gutierrez B, Wu C, Klein B, Pigott DM, Open COVID-19 Data Working Group, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 2020 May 01;368(6490):493-497 [FREE Full text] [doi: 10.1126/science.abb4218] [Medline: 32213647]
47. Wu J, Leung K, Leung G. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet 2020 Feb;395(10225):689-697. [doi: 10.1016/s0140-6736(20)30260-9]

Abbreviations

AR: autoregressive
China CDC: Chinese Center for Disease Control and Prevention
COVID-19: coronavirus disease
GLEAM: global epidemic and mobility
I: infectious
L: latent
LASSO: least absolute shrinkage and selection operator

MIDAS: Models of Infectious Disease Agent Study
PHEIC: Public Health Emergency of International Concern
R: removed
RMSE: root mean square error
S: susceptible
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach; submitted 14.05.20; peer-reviewed by P Giabbanelli, G Fagherazzi; comments to author 13.07.20; revised version received 24.07.20; accepted 24.07.20; published 17.08.20.

Please cite as:
Liu D, Clemente L, Poirier C, Ding X, Chinazzi M, Davis J, Vespignani A, Santillana M
Real-Time Forecasting of the COVID-19 Outbreak in Chinese Provinces: Machine Learning Approach Using Novel Digital Data and Estimates From Mechanistic Models
J Med Internet Res 2020;22(8):e20285
URL: http://www.jmir.org/2020/8/e20285/
doi:10.2196/20285
PMID:32730217

Original Paper

COVID-19 Mortality Underreporting in Brazil: Analysis of Data From Government Internet Portals

Lena Veiga e Silva[1,2], MSc; Maria Da Penha de Andrade Abi Harb[1], MSc; Aurea Milene Teixeira Barbosa dos Santos[1], MSc; Carlos André de Mattos Teixeira[1], BSc; Vitor Hugo Macedo Gomes[1], BSc; Evelin Helena Silva Cardoso[1], MSc; Marcelino S da Silva[1], PhD; N L Vijaykumar[3,4], PhD; Solon Venâncio Carvalho[4], PhD; André Ponce de Leon Ferreira de Carvalho[5], PhD; Carlos Renato Lisboa Frances[1], PhD

[1]Federal University of Pará, Belém, Brazil
[2]University of Amazon, Belém, Brazil
[3]Federal University of São Paulo, São José dos Campos, Brazil
[4]National Institute for Space Research, São José dos Campos, Brazil
[5]University of São Paulo, São Carlos, Brazil

Corresponding Author:
Lena Veiga e Silva, MSc
Federal University of Pará
R Augusto Corrêa, 01
Guamá
Belém, 66073-040
Brazil
Phone: 55 91 3201 7634
Email: lenaveiga@gmail.com

Abstract

Background: In Brazil, a substantial number of coronavirus disease (COVID-19) cases and deaths have been reported. It has become the second most affected country worldwide, as of June 9, 2020. Official Brazilian government sources present contradictory data on the impact of the disease; thus, it is possible that the actual number of infected individuals and deaths in Brazil is far larger than those officially reported. It is very likely that the actual spread of the disease has been underestimated.

Objective: This study investigates the underreporting of cases and deaths related to COVID-19 in the most affected cities in Brazil, based on public data available from official Brazilian government internet portals, to identify the actual impact of the pandemic.

Methods: We used data from historical deaths due to respiratory problems and other natural causes from two public portals: DATASUS (Department of Informatics of the Unified Healthcare System) (2010-2018) and the Brazilian Transparency Portal of Civil Registry (2019-2020). These data were used to build time-series models (modular regressions) to predict the expected mortality patterns for 2020. The forecasts were used to estimate the possible number of deaths that were incorrectly registered during the pandemic and posted on government internet portals in the most affected cities in the country.

Results: Our model found a significant difference between the real and expected values. The number of deaths due to severe acute respiratory syndrome (SARS) was considerably higher in all cities, with increases between 493% and 5820%. This sudden increase may be associated with errors in reporting. An average underreporting of 40.68% (range 25.9%-62.7%) is estimated for COVID-19–related deaths.

Conclusions: The significant rates of underreporting of deaths analyzed in our study demonstrate that officially released numbers are much lower than actual numbers, making it impossible for the authorities to implement a more effective pandemic response. Based on analyses carried out using different fatality rates, it can be inferred that Brazil's epidemic is worsening, and the actual number of infectees could already be between 1 to 5.4 million.

(J Med Internet Res 2020;22(8):e21413) doi:10.2196/21413

KEYWORDS

Brazil; COVID-19; mortality; underreporting; respiratory system diseases; public health; pandemic; time series; forecasting

Introduction

Background

On December 31, 2019, the World Health Organization (WHO) received a report from China about cases of pneumonia of unknown etiology in Wuhan, Hubei Province. By January 7, 2020, Chinese scientists isolated the virus, identifying it as a novel coronavirus and initially referred to it as 2019-nCoV (later named severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2]) [1-3]. The virus, which causes coronavirus disease (COVID-19) [4], ended up spreading to other countries and, by late January 2020, the WHO declared it an Public Health Emergency of International Concern; the outbreak was declared a pandemic on March 11, 2020.

The global impact of the virus has been of great concern and has overburdened public health systems worldwide. It can be considered the first true global epidemic of this magnitude in the digital era [5]. COVID-19 is an acute respiratory disease, often severe, which may become fatal to those who are infected [1]. The disease occurs when one comes into contact with contaminated secretions, in particular, large respiratory droplets, as well as when in contact with contaminated surfaces [3]. It disseminates rapidly, compromising the health of a large number of people, and consequently overwhelms health care infrastructure and resources. Decision makers must act immediately to minimize the effects of the disease and flatten the contagion curve to control both spread and fatalities.

As the disease propagates, the burden to health care systems increases, despite a large number of asymptomatic cases. Studies in China show that 62% of COVID-19 transmissions occur as a result of asymptomatic and presymptomatic individuals [6]. Thus, there is a high chance that the actual number of infectees is far larger than that officially announced. Moreover, it is very likely that the actual proliferation of the disease is being underestimated, with a very high number of underreported cases.

The Pandemic in Brazil

Outside the Asian continent, the disease was initially concentrated in Western Europe and North America. In a short period of time, however, it expanded to other parts of the world like Africa and Latin America [7]. Brazil's first case and death were announced on February 26th and March 17th, respectively. Since then, the disease has been spreading rapidly, devastating almost all regions of the country; at present, Brazil has the fourth highest number of deaths and the second highest number of confirmed infections [8]. According to the coronavirus website of Brazilian Ministry of Health [9], there were more than 700,000 confirmed cases and almost 40,000 deaths, as of June 9, 2020.

The country's difficult situation is magnified due to social inequalities. According to the Brazilian Institute of Geography and Statistics (IBGE) [10], Brazil has a population of approximately 204.5 million people, of which 85% are <59 years of age. The country has 65 million (31.8%) people living in poor or extreme conditions of poverty (eg, precarious living, lack of basic sanitation, reduced access to health care, etc). It has recorded an unemployment rate of 12.2% in the first quarter of 2020 [10]. Public measures tailored to these populations are necessary. On a positive note, Brazil has a government-funded Unified Healthcare System (Sistema Único de Saúde, SUS) that is responsible for 70% of the population [11].

Brazil has 27 states divided territorially into five major regions: North, Northeast, Midwest, Southeast, and South, with specific climatic, social, and economic characteristics. According to the IBGE [12], the North region has the lowest demographic density, with 4.72 inhabitants/km^2 and a Human Development Index (HDI) of 0.683. The Southeast region is more developed and the most populous, with approximately 92 inhabitants/km^2, and accounts for 55.2% of the national gross domestic product (GDP) (HDI=0.784). There is greater social inequality in the Northeast region (HDI=0.608).

A proper estimation of underreported or wrongly reported cases is necessary for a better understanding of the actual epidemic scenario; this will allow for necessary and effective measures to be undertaken by the authorities. In Brazil, underreporting is due to the low rate of testing per 1 million inhabitants. Additionally, there is significant delay in the reporting of test results [13]. During the first weeks of the COVID-19 outbreak, Brazil had tested all suspected cases as well as those that had been in contact with a confirmed case. However, low availability of RT-PCR (reverse transcription polymerase chain reaction) tests forced the Ministry of Health to recommend testing for only serious cases [9]. This approach was also extended to those belonging to high-risk groups (eg, health care professionals).

Different grades of testing and reporting are observed in other countries [14] so it is difficult to understand what the actual situation in Brazil and its states looks like. According to WorldoMeter [15], 1,182,581 tests have been conducted in Brazil so far, a rate of 5566 tests per 1 million inhabitants, which is much lower than that other countries like Spain (86,921 tests per 1 million inhabitants), Portugal (78,030 tests per 1 million inhabitants), and the United States (53,156 tests per 1 million inhabitants).

This undersampling leads to a high degree of underreported cases, which affects estimates of the actual fatality rate of the disease [7]. Therefore, it is of fundamental importance to uncover the degree to which underreporting has occurred in order to define and establish public health policies related to pandemic response.

It has been suggested that the reproduction number (R) must be less than 1 in order to reduce the number of infected cases [7]. However, although several Brazilian states have adopted isolation, social distancing, and even lockdown measures, noncompliance is an issue.

Official Brazilian Government Internet Portals

With the increasing spread of SARS-CoV-2 in Brazil, there has been a considerable growth in the population's interest for information about the disease. According to Google Trends [16], web queries for the term "Coronavirus" increased substantially in Brazil, reaching its peaks on March 15th and 21st. The most searched terms included "cases of coronavirus," "deaths coronavirus," "coronavirus symptoms," and

"coronavirus update." During this period, access to news about the virus increased by more than 5000% when compared with the previous period. Additionally, tweets related to the novel coronavirus were among those that were most commented on; in Brazil, topics such as chloroquine, Minister of Health, quarantine, and treatment of coronavirus were the most sought after on Twitter [17].

To manage this increase in interest, several official internet portals were created by the Brazilian municipal, state, and federal bodies for dissemination, monitoring, and guidance. However, the data presented by these public internet portals are contradictory and inaccurate. Some of the data released highly underreport the true number of cases, leading to false perceptions that the contagion is under control.The population must trust the data provided to them in order to accept proposed recommendations [18].

We believe that by aggregating officially available information into a single internet portal, removing contradictions, and using reliable sources, we can gather support from the Brazilian populace to follow WHO-recommended guidelines, thus reducing the contagion rate in Brazil. This portal is under development as part of the work presented in this paper and will enable policy and decision makers to base their assessments on scientific evidences and guide citizens in adopting recommended measures and behaviors (eg, social distancing, frequent hand sanitizing, and more attention to hygiene issues).

This Study

The work described in this paper conducts an investigation into underreported deaths with respect to COVID-19 based on historical mortality data due to respiratory problems and other natural causes. These data are publicly available on the internet through the two main portals of the Brazilian government: the Mortality Information System (SIM) of DATASUS (Department of Informatics of the Unified Healthcare System) [19] and the Brazilian Transparency Portal of Civil Registry [20]. The aim is to systematize the contradictory information in these portals to provide a more representative picture of the pandemic and estimate the possible number of death reports that were incorrectly recorded. These data were used to build time-series models (modular regressions) with the ability to predict the expected mortality rate for 2020. This was done to assess whether significant disagreement is present between the real and expected number of deaths for this period. By estimating the actual number of COVID-19–related deaths, it is possible to determine the number of infected people from officially published fatality rates.

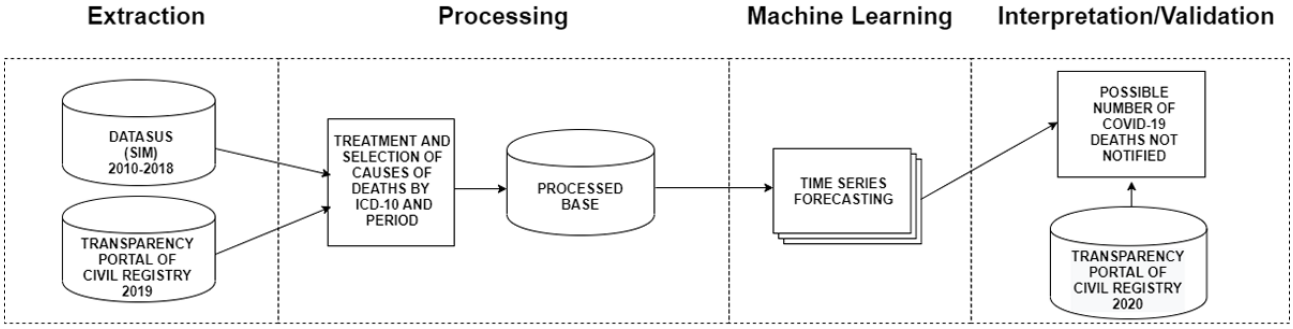
In this study, we used as case studies the capital cities of three regions that were most affected by the pandemic: North (Belém and Manaus), Northeast (Fortaleza and Recife), and Southeast (São Paulo and Rio de Janeiro). The resulting mortality underreporting scenario will be considered for the entire country as these cities represent around 47% of the total deaths in Brazil as of June 9, 2020 [9].

Methods

Overview

We followed the Knowledge Discovery in Databases workflow to extract new and relevant data to enable decision making (Figure 1). Two public databases with nationally consolidated data were consulted: DATASUS and the Brazilian Transparency Portal of Civil Registry. In the analysis, these steps were followed: data extraction, data processing, machine learning, and data interpretation and validation. Health care specialists aided in some of these steps.

Figure 1. Methodology diagram adapted from Fayyad et al [21]. DATASUS: Department of Informatics of the Unified Healthcare System; SIM: Mortality Information System; ICD-10:International Statistical Classification of Diseases and Related Health Problems–10th Revision; COVID-19: coronavirus disease.



Data Extraction

Data were collected from two government sources accessible for public use. The registers present in both databases follow the international standards set by the WHO.

Part of the data collected for this research was extracted from DATASUS (SIM) [19]. It is a system from which one can access regular information on mortality rates in Brazil to assist public health management sectors [19]. Data were extracted for the 2010-2018 period for all capital cities of Brazilian states. It is important to clarify that SIM is updated annually; hence, 2019 was not considered since the data is not available yet. Each entry in the SIM database is highly detailed, concisely presenting all the information contained in the death certificate.

Another source was the Brazilian Transparency Portal of Civil Registry [20]. It comprises deaths registered due to COVID-19 (confirmed or suspected) and respiratory diseases, such as severe acute respiratory syndrome (SARS), pneumonia, and respiratory failure. The civil registry data website is based on death certificates sent by the registry offices countrywide for deaths that take place in hospitals, residences, public roads, etc. Data

were collected for the January 1 to June 1, 2020 period, as well as the same period for the year 2019. For the years 2019 and 2020, the civil registry portal records another category—deaths from other causes (when these were unrelated to COVID-19 but related to respiratory problems). This last category was also considered in this study.

The Brazilian civil registry portal presents the data duly notarized by the civil registry offices and follows a series of legal timelines established by the Brazilian Constitution—a family has 24 hours after the death of a member to notify the registry office, and in turn the registry office has up to 5 days to duly register the death; within 8 days the Information Center of Civil Registry receives the report, which is published by the civil registry portal. Therefore, there may be a delay of 14-15 days for the portal to publish a record.

In addition to the large delay in the Transparency Portal of Civil Registry death reports, it is important to highlight that the update frequency might be different for each city. For certain regions, the delays are even longer. In general, the data for capital cities are updated more frequently. For this reason, although the data were collected on June 1st, the analysis will be conducted using data made available up to May 21st. By adopting this procedure, we can mitigate the effect of late notifications in the analysis.

Data Processing

Data were preprocessed by removing missing and duplicated information to improve quality, so that more significant results can be presented. This removal of data was not substantial, and the entire data set was stored in a single database.

The time series of deaths due to the previously mentioned diseases were from DATASUS (SIM) and were duly processed to be concatenated with those from the Transparency Portal of Civil Registry. Following the conditions used by the civil registry portal, each occurrence of death was classified according to the International Statistical Classification of Diseases and Related Health Problems (ICD) [22] and based on the last, underlying, and immediate cause of death present in the death certificate. The fields used in the database for date of death and ICD are mandatory. The nested classification conditions are summarized in Table 1.

In order to classify each record of data from DATASUS (SIM) based on the listed conditions, it was necessary to identify the ICDs [22]. Thus, the corresponding IDs for the causes of deaths from the civil registry portal are shown in Table 2. Health care specialists contributed to identifying and classifying the ICDs.

In order to merge the databases, data referring only to death records for capital cities were extracted from DATASUS (SIM). These records were then aggregated on a daily basis. Therefore, both the databases are now compatible with respect to their indices and columns, making it possible to concatenate the data and merge into a single data set, which was then used to conduct this study.

Table 1. Conditions established by the Transparency Portal of Civil Registry to classify deaths.

Order	Condition
1	If there is any mention of COVID-19[a] in the death certificate, suspected or confirmed, it was considered a death attributed to COVID-19.
2	If there is any mention of severe acute respiratory syndrome (SARS), it was considered the cause of death.
3	If there is any mention of pneumonia, it was considered the cause of death.
4	If respiratory failure is listed as the only cause, it was considered the cause of death.
5	If the certificate does not mention any of the above conditions, the cause of death was considered as “other”.

[a]COVID-19: coronavirus disease.

Table 2. International Statistical Classification of Diseases and Related Health Problems–10th Revision (ICD-10) classification adopted by the Transparency Portal of Civil Registry.

Disease	ICD-10 classification
Severe acute respiratory syndrome (SARS)	I260, U04, J22, J100, J110
Pneumonia	J12, J13, J14, J15, J16, J180, J181, J182, J188, J189, B953, B960, B961
Respiratory failure	J96

Time-Series Forecasting Model

The models used for time-series prediction were adjusted to predict the expected number of deaths for 2020 based on a historical series from 2010 to 2018 for six capital cities. In order to conduct the experiment, training based on the modular regression model FbProphet [23] was employed. The resulting decomposed time-series model is shown in the following equation:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad \textbf{(1)}$$

where, according to the model by Harvey and Peters [24], *g(t)* represents a function of tendency used to capture nonperiodic changes in a historical series; *s(t)* refers to periodic seasonality, representing the annual, monthly, and weekly recurring behavior; and *h(t)* represents the effects of holidays on the data. The component ε_t is used to represent peculiar changes not included in the model.

The main component of equation 1, *g(t)*, is used to represent the trend model. Equation 2 refers to this component when used in forecasting problems that exhibit a linear trend with change points:

$$g(t) = (k + a(t)T\delta)t + (m + a(t)T\gamma) \quad \textbf{(2)}$$

where is the growth rate, δ is a vector containing adjustments to the growth rate, is used as an offset parameter, and γ is used as an adjustment vector for the parameter . The vector *a(t)* is used to define the change points, allowing the growth rate to be adjusted accordingly.

As previously mentioned, component *s(t)* of equation 1 is used to represent the seasonal influences and recurring behaviors present in the time series. Those seasonal effects rely on a Fourier series representation (equation 3). It is possible to adjust the parameter *P*, represented in days, in order to obtain the desired seasonality (eg, *P*=7 for weekly seasonality).

In order to fit the model to the data, the time-series forecasting is treated as a curve-fitting problem, taking the data seasonalities and holiday effects into consideration [23]. The framework uses an implementation of the Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm, referenced by Zhu et al [25], in order to find a maximum *a posteriori* estimate.

Data Interpretation and Validation

For this analysis, we used data on COVID-19–related deaths of the six capital cities with the highest number of deaths recorded by the civil registry website: Belém (capital of Pará), Fortaleza (capital of Ceará), Manaus (capital of Amazonas), Recife (capital of Pernambuco), Rio de Janeiro (capital of Rio de Janeiro), and São Paulo (capital of São Paulo).

Once the processing workflow and data cleaning are completed, it is possible to devise a system to predict trends in deaths caused by respiratory issues, as well as to predict the expected behavior of diseases for 2020. Based on the number of deaths per year for each disease for the capital cities under consideration, an estimate of deaths was calculated for normal conditions (ie, no pandemic). Thus, the difference between the number of expected cases for 2020 and recorded cases for 2020 was determined. Next, this extrapolation was added to the deaths reported for COVID-19, allowing us to estimate the actual number of deaths due to the pandemic. With this analysis, the actual cause of sudden increase in deaths, not only due to respiratory issues but also other deaths, could be estimated.

Results

We conducted an exploratory analysis of the data to evaluate patterns in the number of deaths during the pandemic. Subsequently, we employed a time-series model to estimate the number of incorrectly reported figures.

Exploratory Data Analysis

The historical series of deaths for 2010-2018 (extracted from SIM [19]), 2019, and 2020 (extracted from civil registry portal [20]) for a same period for all the mentioned years were considered. We observed an increase of 965% (from 75.8 to 732) with respect to the average number of registered deaths due to SARS and respiratory failure per year for Manaus, one of the most affected capitals (Figure 2). Due to a high disagreement from the historical series of deaths for the mentioned period that coincides with the pandemic period, it is necessary to investigate the cause of this large difference.

Recife, Belém, Fortaleza, São Paulo, and Rio de Janeiro also presented a significant increase in the number of deaths in 2020. Figure 2 illustrates the disagreement between the number of deaths that occurred between the 13th and 19th weeks of the epidemic in 2020 with respect to the average of the historical series for the same period in previous years for both the diseases—respiratory failure and SARS—that presented a large variation. It is possible to observe distinct behaviors in the discrepancy in records for each city. In Recife, the substantial increase in SARS cases draws a great deal of attention, while Manaus presented a considerable increase for all causes of death. Despite the increase being more significant for SARS and respiratory failure, we observed occasional discrepancies in regard to pneumonia and deaths due to other causes. The mean number of deaths and standard deviations, along with the percentage of increase with respect to the average of the historical series for these diseases, are presented in Table 3.

As previously mentioned, we observed a major discrepancy for SARS-related deaths for all cities. A sudden increase of 6991% (from 9.8 to 685) for SARS in Recife, for example, might be associated with errors in reporting. SARS, first detected in China in November 2002, is caused by a type of coronavirus called severe acute respiratory syndrome coronavirus (SARS-CoV), with symptoms similar to COVID-19, causing a severe respiratory viral infection [26]. Thus, it is possible that the similarities between the diseases can compromise the accuracy of death records.

Figure 2. Increases in the number of deaths due to respiratory failure and severe acute respiratory syndrome (SARS).

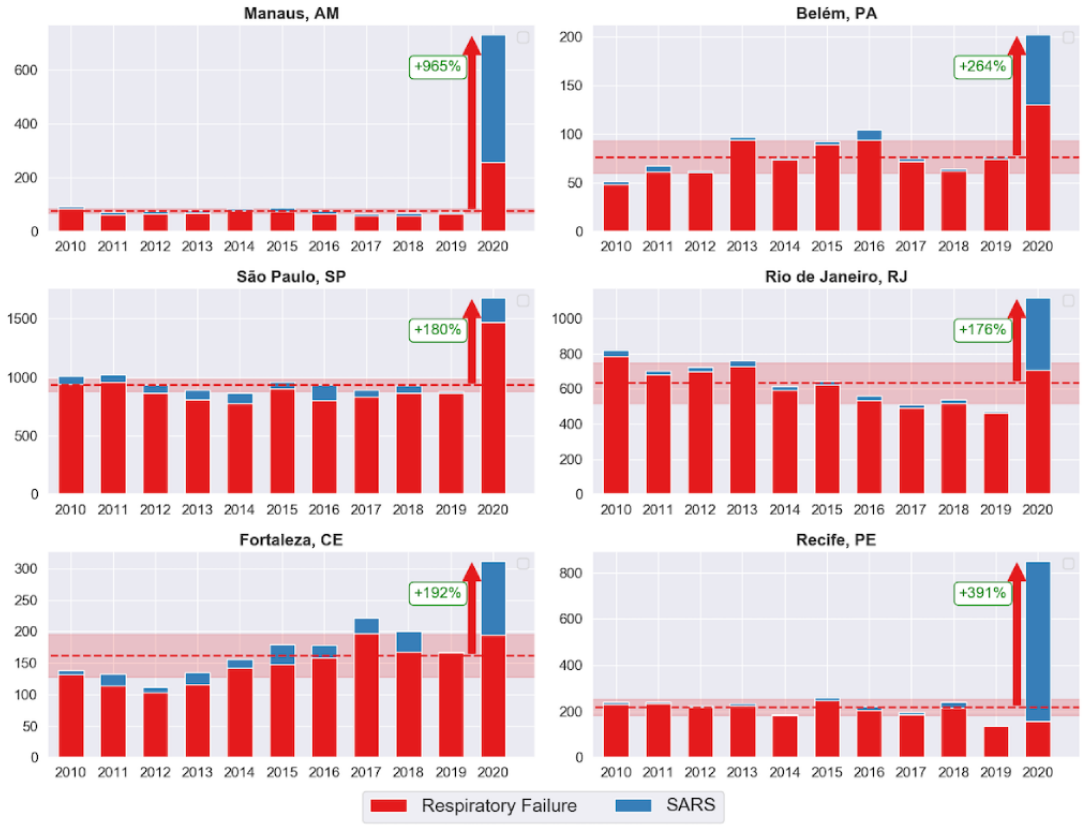


Table 3. Mean (SD) for the historical series and percent increase/decrease of deaths caused by respiratory failure, pneumonia, severe acute respiratory syndrome (SARS), and other causes.

City	Respiratory failure	Pneumonia	SARS	Other causes
Belém				
Mean (SD)	72.7 (15.62)	180.6 (42.09)	3.6 (2.54)	491.1 (72.5)
Increase/decrease (%)	+78	+37	+1900	–1
Fortaleza				
Mean (SD)	144.3 (29.09)	442.2 (147.90)	17.9 (10.12)	1474.6 (161.92)
Increase/decrease (%)	+35	–1	+553	–11
Manaus				
Mean (SD)	66.8 (9.01)	259.4 (41.80)	9.0 (2.16)	1162.6 (122.70)
Increase/decrease (%)	+283	+192	+5188	+69
Recife				
Mean (SD)	207.5 (32.2)	307.1 (39.3)	9.8 (6.98)	1963.2 (305.40)
Increase/decrease (%)	–24	–43	+6991	–25
Rio de Janeiro				
Mean (SD)	611.3 (108.85)	1501.1 (166.94)	22.7 (7.64)	6065.1 (495.47)
Increase/decrease (%)	+15	+16	+1701	–5
São Paulo				
Mean (SD)	861.8 (59.12)	2933.3 (247.11)	70.4 (30.82)	8418.1 (571.08)
Increase/decrease (%)	+70	–2	+192	+6

Time-Series Prediction

The exploratory analysis identified values that were much higher than the average of the historical series for registered deaths during the pandemic period. For this reason, in this section we further analyze the results obtained from the time-series models developed to compare the expected trend (predicted) and the actual trend.

We trained the time-series models with data from January 2010 to May 2019. The model was adjusted to individually predict the behavior of each of the three diseases and deaths over other causes in each.

To compute the error metrics, each model was initially trained using 7 years of data. A cross-validation process was then conducted for the remaining data for every 90-day cutoff at a 470-day horizon. Table 4 shows the absolute errors for the validation set predictions.

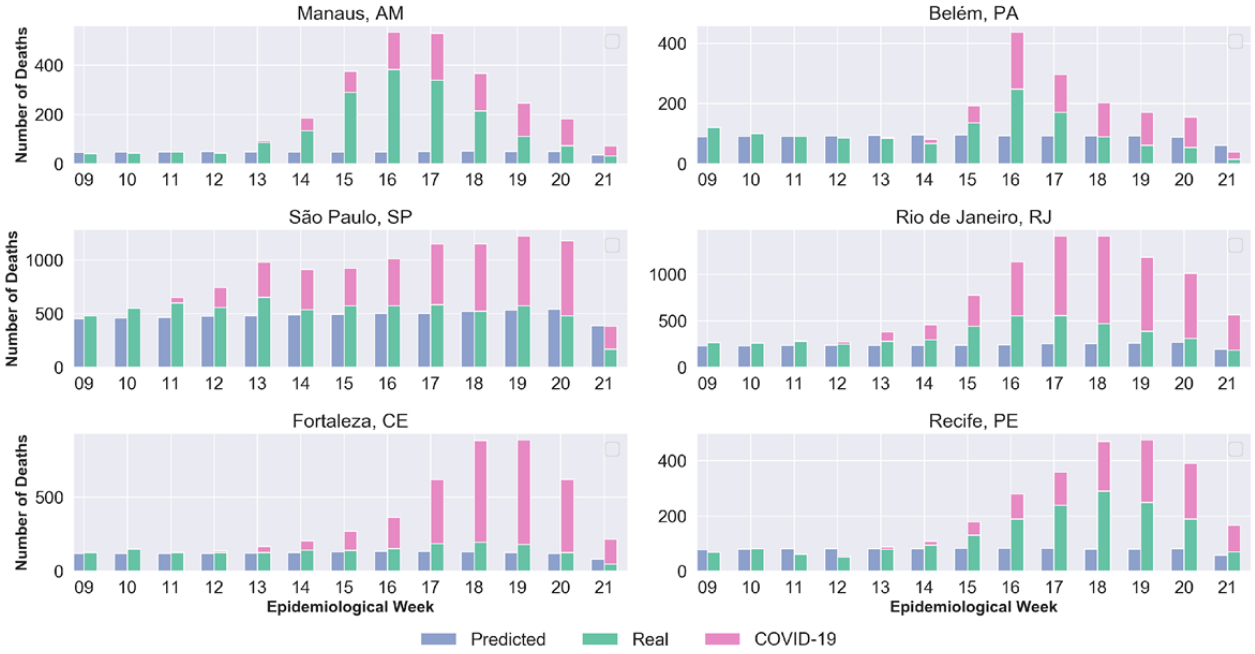
The models were then used to predict data up to May 21, 2020, to be compared with the actual data presenting the observed anomalies. Figure 3 compares the number of registered deaths (actual) from civil registry website, including deaths due to COVID-19, and the predicted deaths returned by the time-series models. The results are grouped by epidemiological weeks and considers data from the 9th week until the 21st week of 2020. Our results demonstrated that each city presented a different trend with respect to the peak periods for disease activity within the considered timeframe. Therefore, analysis must be performed considering their specific periods.

Table 4. Mean absolute error (MAE) and mean absolute percentage error (MAPE).

City	Respiratory failure	Pneumonia	SARS[a]	Other causes
Belém				
MAE	1.61	2.64	0.40	5.15
MAPE	9.6	11.6	33.5	8.3
Fortaleza				
MAE	1.81	2.34	0.57	6.59
MAPE	11.4	2.6	37.0	10.1
Manaus				
MAE	0.75	1.88	0.34	4.47
MAPE	14.0	10.0	28.4	8.3
Recife				
MAE	1.91	2.34	0.59	7.45
MAPE	12.3	7.8	40.0	6.0
Rio de Janeiro				
MAE	2.77	4.99	0.50	13.38
MAPE	6.7	6.8	25.8	5.2
São Paulo				
MAE	3.34	7.78	0.96	12.27
MAPE	2.4	3.1	36.4	2.3

[a]SARS: severe acute respiratory syndrome.

Figure 3. Predicted and actual deaths per epidemiological week related to respiratory diseases. COVID-19: coronavirus disease.



Taking into account the peak periods for each city, predicted figures are smaller than the actual values in terms of the days with a high number of deaths due to respiratory and other causes. The estimates of errors in death reports for each disease, per city, are shown in Figure 4. The number at the end of each bar represents an estimate, in absolute numbers, of the number of cases that deviate from the expected pattern, and most probably were incorrectly recorded.

Each city, with its own particularities (Figure 4), has its causes of death recorded differently. Table 5 presents the considered periods for each city and the difference between the number of reported cases and the number of predicted cases both quantitatively and percentagewise. The last column shows the total difference in the number of deaths for the period not covered in the historical series.

The predicted values show different increases for the investigated cities. For São Paulo, where the first COVID-19 death confirmed by the Brazilian government occurred in the 11th week, the increase was 24.4% (from 7238 to 9004). For the other cities, the following increases were observed: 144.7% (from 1274 to 3117) for Manaus, 128.9% (from 575 to 1317) for Recife, 99.6% (from 485 to 968) for Belém, 41.2% (from 1279 to 1806) for Fortaleza, and 39.9% (from 3475 to 4863) for Rio de Janeiro. These percentages refer to the increase in death records that didn't reference COVID-19. Thus, one can see a significant increase in the number of deaths during the epidemic period that attributed to causes that deviate from the expected pattern.

The discrepancy is clearly very large, in terms of percentage values, with respect to the reports on deaths due to diseases considered in this research and other causes, especially SARS, which reported an increase of around 5820% (from 8.04 to 476) in Manaus and 2880% (from 23.32 to 695) in Recife.

Figure 4. Estimated number of deaths wrongfully attributed to respiratory system diseases for the considered periods. SARS: severe acute respiratory syndrome.

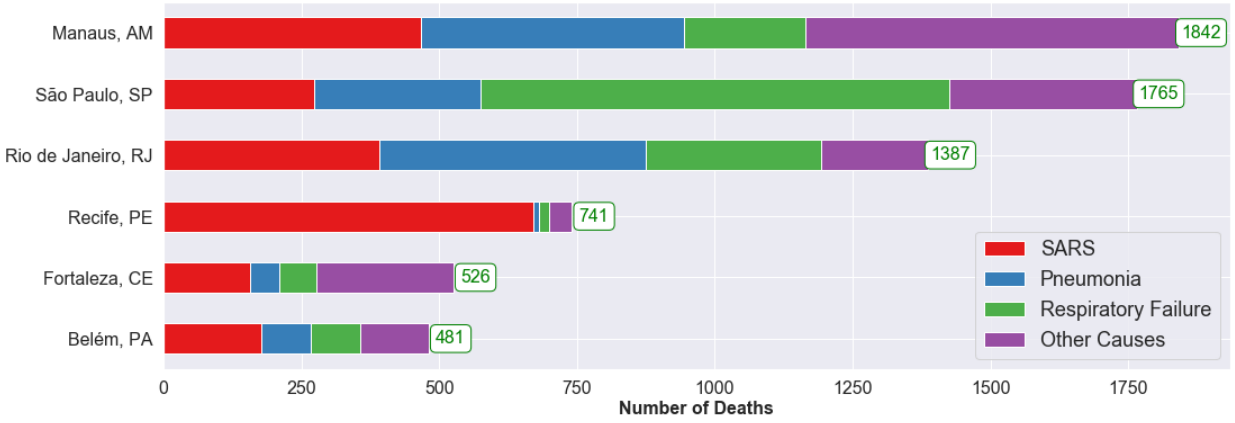


Table 5. Difference (Δ) between real and predicted values.

Cities (epidemiological weeks)	Δ Deaths				Δ Total deaths
	Respiratory failure	Pneumonia	SARS[a]	Other causes	
Belém (15th to 17th)					
Difference	88	90	178	125	481
Increase (%)	127.24	60	1715.56	49.41	99.61
Fortaleza (14th to 18th)					
Difference	69	52	157	248	526
Increase (%)	52.69	23.47	815.20	27.33	41.17
Manaus (13th to 19th)					
Difference	220	477	467	678	1842
Increase (%)	611.11	196.12	5820.40	68.75	144.74
Recife (14th to 19th)					
Difference	18	11	671	41	741
Increase (%)	25	35.48	2880.27	9.13	128.92
Rio de Janeiro (13th to 19th)					
Difference	319	483	391	194	1387
Increase (%)	96.28	42.49	2284.8	9.7	39.96
São Paulo (10th to 17th)					
Difference	851	301	274	339	1765
Increase (%)	91.15	24.31	493.53	6.77	24.40

Discussion

Principal Findings

It is reasonable to assume that the values presented in Table 5 were incorrectly reported, concealing the actual number of deaths due to the pandemic. The reporting bias for COVID-19 (relating to respiratory diseases) may have occurred due to delays in releasing the results, lack of tests, or even errors in identifying the disease. It is important to stress that even other causes of deaths increased significantly during the pandemic period (eg, an increase of 68% [from 677.92 to 1664] in Manaus). This study attributes some of these deaths to COVID-19 as well.

Therefore, the extrapolated (period not covered in the historical series) values of the number of deaths were attributed to the underreporting of the pandemic. Table 6 shows the estimates of the percentage of underreporting of COVID-19–related deaths for each city compared to the official number of deaths up to May 21, 2020.

For the cities of this case study, an average underreporting of 40.7% is estimated for deaths related to COVID-19. The values vary between 25.9% to 62.7%, with emphasis on Manaus, which had the highest number of deaths underreported (62.7%), and Recife, with almost 50%. Fortaleza had the lowest number, with 25.9% of underreporting, in spite of its count being substantial.

Table 6. Underreported deaths due to coronavirus disease (COVID-19).

City	Population, N (PNAD[a])	Extrapolated number of predicted deaths	Official number of deaths[b]	Total number of estimated deaths	Number of deaths per 1 million inhabitants	Underreported deaths (%)
Belém	1,492,745	481	952	1433	959.98	33.57
Fortaleza	2,669,342	526	1503	2029	760.11	25.92
Manaus	2,182,763	1842	1094	2936	1345.08	62.74
Recife	1,645,727	739	747	1486	902.94	49.73
Rio de Janeiro	6,718,903	1387	2376	3763	560.06	36.86
São Paulo	12,252,023	1765	3238	5003	408.34	35.28

[a]PNAD: Pesquisa Nacional por Amostra de Domicílios (National Household Sample Survey).

[b]As of May 21, 2020,

The National Household Sample Survey (Pesquisa Nacional por Amostra de Domicílios, PNAD) of the IBGE compiles data based on the socioeconomic characteristics of the Brazilian population [12]. By analyzing the number of deaths and population counts from the PNAD (Table 6), one can see the differences in underreporting and number of deaths per 1 million inhabitants for each city. The differences are also found in Table 5; there are several disagreements for underreporting bias for COVID-19. The differences may have occurred due to the distinct socioeconomic characteristics of each city, such as demographic density, HDI, population age group, access to health care, and number of intensive care unit (ICU) beds available, etc.

São Paulo, for example, ended up with the least number of deaths in terms of percentage (per population) and the least total difference (percentagewise) in deaths for the period not covered in the historical series (Table 5). Moreover, São Paulo has the highest HDI (0.8) in Brazil. It has a one of the highest numbers of ICU beds in the country—22.3 ICU beds per 100,000 inhabitants [27], which is much higher than necessary. On the other hand, Manaus, one of the most affected cities in Brazil, showed the highest difference in records for the extrapolated period not covered in the historical series (Table 5) and the highest number of deaths (population wise) as well as underreporting of deaths. Manaus has the lowest HDI (0.73) among the six capital cities and 9.63 ICU beds per 100,000 inhabitants, the smallest number among the considered cities.

In a recent study, EPICOVID19-BR, carried out by the Federal University of Pelotas (UFPel) [28], researchers interviewed and tested (for SARS-CoV-2) a group of people selected by lottery in the cities identified as the most affected in the country. The objective was to estimate the number of infectees for each city. The first stage considered 133 cities from all Brazilian states and took place between May 14-21, 2020. In this study, the authors reported the following percentage values of infection: Belém (15.10%), Fortaleza (8.7%), Manaus (12.5%), Recife (3.2%), Rio de Janeiro (2.2%), and São Paulo (3.1%).

In the context of EPICOVID19-BR, fatality rates were estimated using the total deaths predicted, along with the official figures of infections and the number of infections estimated by UFPel [28]. The discrepancy between the official number of the fatality rates—Belém (0.64%), Fortaleza (1.37%), Manaus (1.08%), Recife (2.82%), Rio de Janeiro (1.62%), and São Paulo (2.22%)—becomes evident as there is much difference between official figures and counts reported by EPICOVID19-BR. These rates are compatible with those found in several studies [7,29,30]. Therefore, it is estimated that mortality values range from 0.64% (Belém) to 2.82% (Recife), and is much more reliable with respect to officially published counts. Emphasis must be given to the results presented by UFPel (CI 4.8%), which confirms the hypothesis that there is a substantial underreporting not only in the number of deaths but also and especially in the number of infections published by official government bodies.

Another relevant study, from Imperial College [7], estimated the COVID-19 impact in Brazilian states from February 25, 2020 to May 6, 2020, using a hierarchical Bayesian model. This model estimates the number of infections, deaths, and reproduction. These fatality rates are estimated to be much more optimistic than those from UFPel. The following fatality rates were calculated: Belém (Pará: 0.9%), Fortaleza (Ceará: 1.1%), Manaus (Amazon: 0.8%), Recife (Pernambuco: 1.1%), Rio de Janeiro (Rio de Janeiro: 0.8%), and São Paulo (São Paulo: 0.7%).

From the several fatality rates investigated (up to the time this study was conducted), and considering the main countries affected by the pandemic and number of predicted deaths in our research, it is possible to estimate the number of infected cases and consequently estimate the percentage of underreporting of infected cases. Table 7 presents estimations of the numbers of those that were infected in each city considering different fatality rates and also shows the estimated percentage of underreporting of infected cases.

Depending on how high or low the fatality ratio is, there is variation in the number of infected cases. For example, as seen in Table 7, the number of cases for São Paulo is estimated to be almost 76,000, considering the highest fatality ratio (Brazil, 6.6%), or approximately 715,000 when considering the lowest fatality ratio (Imperial College, 0.7%).

Based on these differing fatality rates, underreported infection numbers may be monumental. For example, underreporting of infected cases in Manaus (using the fatality ratio from the Imperial College study [7]) and Belém (using the fatality ratio from the EPICOVID19-BR study [28]) may reach 2880% and 2837%, respectively. Such scenarios show, in both the cities, a count that is 30 times the number of confirmed cases. For other capital cities, the numbers may be up to 11 (Recife), 12 (Fortaleza), 17 (São Paulo), and almost 25 times (Rio de Janeiro).

There were 739,503 confirmed cases and 38,406 official deaths, as of June 9, 2020 [9]. If we consider the average percentage of 40.7% for underreporting of deaths as shown in this study, Brazil would have around 64,746 deaths related to COVID-19. Considering the lowest and highest percentage of underreporting presented by the cities studied (Table 6), it would have around 51,846 (25.9%) and 103,071(62.7%) deaths, respectively, thus, estimating a much higher number of deaths than those officially reported.

Table 7. Estimated number of infection cases and percentage of cases underreported considering differing estimations in fatality rate.

Cities (predicted number of deaths)	Official count[a]	Fatality rate					
		UFPel[b] [26]	Imperial College [7]	China (1.38%) [27]	Brazil (6.6%)	United States (6%)	Global (6.5%)
Belém (n=1433)	7675						
Infections, n		225,404	159,222	103,841	21,712	23,883	22,046
Underreported (%)		2837	1975	1253	183	211	187
Fortaleza (n=2029)	1864						
Infections, n		232,233	184,455	147,029	30,742	33,817	31,215
Underreported (%)		1146	889	689	65	81	67
Manaus (n=2936)	12,317						
Infections, n		272,845	367,000	212,754	44,484	48,933	45,170
Underreported (%)		2115	2880	1627	261	297	267
Recife (n=1486)	11,584						
Infections, n		52,663	135,091	107,681	22,515	24,767	22,861
Underreported (%)		355	1066	830	94	114	97
Rio de Janeiro (n=3763)	18,743						
Infections, n		147,816	470,375	272,681	57,015	62,717	57,892
Underreported (%)		689	2410	1355	204	235	209
São Paulo (n=5003)	41,451						
Infections, n		379,813	714,714	362,536	75,803	83,383	76,969
Underreported (%)		816	1624	775	83	101	86

[a]As of May 21, 2020.

[b]UFPel: Federal University of Pelotas.

Regarding the number of those infected by the pandemic, based on the value previously calculated for the number of total deaths (40.7%, 64,746 deaths), it can be inferred that Brazil's count of infection ranges between 981,013 and 5,395,571 (considering respectively the highest and lowest lethality rate, 6.6% and 1.2%, respectively [7]). Hence, it is reasonable to assume that Brazil either is, or may become in the near future, the new epicenter of the COVID-19 pandemic, surpassing the United States, which of June 9, 2020, has the highest number of infected persons (n=1,933,560) [8].

When comparing both countries, the United States currently performs more tests for the disease than any other country in the world [31]. According to WorldoMeter [15], the United States has conducted 22,624,758 tests—70,799 tests per 1 million inhabitants. These numbers are well ahead of Brazil, which so far has conducted a total of 1,182,581 tests—5566 tests per 1 million inhabitants. Thus, with the testing coverage in the United States being much larger, the actual impact of the pandemic can be more realistically analyzed in that country and, therefore, in comparison to Brazil, more effective actions can be carried out to control the disease.

It is also worth considering the tendency to flatten the evolution curve of COVID-19, which represents the reduction in the number of daily new cases. We compared the evolution of weekly confirmed cases from United States and Brazil, up to June 9th. The reduction in the number of occurrences in the United States indicates that the curve is flattening. In contrast, the number of weekly confirmed cases in Brazil is still increasing. This ascending curve indicates that the pandemic is still growing, tending to surpass the official number of infected Americans in the near future when considering the official numbers. If we consider the highest lethality rates presented in this work, the actual number of infected Brazilian citizens would have already surpassed that of the United States.

Conclusions

The significant rates of underreporting of deaths presented in our research indicate that the counts released by the official Brazilian internet portals are much lower than the actual numbers, making it impossible for the authorities to take more effective action. This is also confusing to citizens, who have demonstrated failure to comply with social isolation measures. Therefore, a public access portal is being developed in order to disseminate more realistic and reliable data on the pandemic, in order to undo the contradictions of official data, guide the population, formulate policies, and estimate the R factor more efficiently.

Our results suggest a growing pandemic and reveal a wide heterogeneity in the outbreak of the epidemic in the cities considered in this case study, suggesting a greater number of underreporting in deaths and infected cases in some cities. This demonstrates differing levels of the outbreak stage, more advanced in some cities compared to others. However, in no city do the results indicate that herd immunity is close to being achieved. In addition, the underreporting of deaths is not stationary over time and may increase during the pandemic period.

The number of deaths due to SARS was considerably higher than the expected number for all six cities, indicating that a large number of deaths related to COVID-19 were possibly mistakenly recorded as SARS. It is assumed that this is due to lack of confirmation and delays in testing or confusion in diagnosis, since COVID-19 is a new disease. Furthermore, delays in disclosing test results also impact the effect and reach of the pandemic. Therefore, it is of paramount importance to increase testing in order to reduce underreporting and encourage rapid dissemination of test results to allow for a closer view of the real COVID-19 situation in Brazil.

Acknowledgments

The authors would like to thank CEPID CeMEAI and FAPESP (process 2013/07375-0) for supporting this work.

Conflicts of Interest

None declared.

References

1. Guo Y, Cao Q, Hong Z, Tan Y, Chen S, Jin H, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. Mil Med Res 2020 Mar 13;7(1):11 [FREE Full text] [doi: 10.1186/s40779-020-00240-0] [Medline: 32169119]
2. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet 2020 Feb;395(10224):565-574. [doi: 10.1016/S0140-6736(20)30251-8]
3. Phelan AL, Katz R, Gostin LO. The Novel Coronavirus Originating in Wuhan, China: Challenges for Global Health Governance. JAMA 2020 Jan 30;323(8):709-710 [FREE Full text] [doi: 10.1001/jama.2020.1097] [Medline: 31999307]
4. Coronavirus disease (COVID-19) situation reports - 51. World Health Organization. 2020 Mar 11. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10 [accessed 2020-05-20]
5. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
6. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 2020 May 01;368(6490):489-493 [FREE Full text] [doi: 10.1126/science.abb3221] [Medline: 32179701]
7. Mellan T, Hoeltgebaum H, Mishra S. Report 21: Estimating COVID-19 cases and reproduction number in Brazil. Imperial College COVID-19 Response Team. 2020 May 8. URL: https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-05-08-COVID19-Report-21.pdf [accessed 2020-05-15]
8. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2020-05-03]
9. Coronavírus Brasil COVID-19. Ministério da Saúde. 2020. URL: https://covid.saude.gov.br [accessed 2020-06-01]
10. Instituto Brasileiro de Geografia e Estatística. 2020. URL: https://www.ibge.gov.br/en/home-eng.html [accessed 2020-05-25]
11. Dana S, Simas A, Filardi B, Rodriguez R, Valiengo L, Neto J. Brazilian Modeling of COVID-19 (BRAM-COD): a Bayesian Monte Carlo approach for COVID-19 spread in a limited data set context. medRxiv 2020:e [FREE Full text] [doi: 10.1101/2020.04.29.20081174]
12. Sistema agregador de informações do IBGE sobre os municípios e estados do Brasil (IBGE's aggregated information system on Brazilian municipalities and states). Instituto Brasileiro de Geografia e Estatística Cidades. 2019. URL: https://cidades.ibge.gov.br/ [accessed 2020-06-03]
13. Análise subnotificação (Underreporting analysis). COVID-19 Brasil. 2020 Apr 11. URL: https://ciis.fmrp.usp.br/covid19/analise-subnotificacao/ [accessed 2020-05-14]
14. Our World in Data. 2020 May 19. URL: https://ourworldindata.org/covid-testing-us-uk-korea-italy [accessed 2020-05-25]
15. COVID-19 Coronavirus Pandemic. WorldoMeters. 2020. URL: https://www.worldometers.info/coronavirus [accessed 2020-05-25]
16. Coronavirus. Google Trends. 2020. URL: https://trends.google.com.br/trends/story/US_cu_4Rjdh3ABAABMHM_en_pt-BR [accessed 2020-07-22]
17. Brazil - Trending Topics. Twitter Trending. 2020. URL: https://www.twitter-trending.com/brazil/en [accessed 2020-07-22]

18. Sesagiri Raamkumar A, Tan SG, Wee HL. Measuring the Outreach Efforts of Public Health Authorities and the Public Response on Facebook During the COVID-19 Pandemic in Early 2020: Cross-Country Comparison. J Med Internet Res 2020 May 19;22(5):e19334 [FREE Full text] [doi: 10.2196/19334] [Medline: 32401219]
19. DATASUS - Ministério da Saúde. 2020. URL: http://www2.datasus.gov.br/DATASUS/index.php?area=0901 [accessed 2020-05-05]
20. Especial COVID-19. Portal da Transparência do Registro Civil. 2020. URL: https://transparencia.registrocivil.org.br/especial-covid [accessed 2020-05-23]
21. Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R. Advances in knowledge discovery & data mining (1st edition). Menlo Park, California: American Association for Artificial Intelligence; 1996.
22. World Health Organization. ICD-10: international statistical classification of diseases and related health problems: tenth revision. 2004. URL: https://apps.who.int/iris/bitstream/handle/10665/42980/9241546530_eng.pdf [accessed 2020-05-12]
23. Taylor A, Letham B. Forecasting at Scale. PeerJ Preprints 2017:e [FREE Full text] [doi: 10.7287/peerj.preprints.3190v1]
24. Harvey AC, Peters S. Estimation procedures for structural time series models. J Forecast 1990 Mar;9(2):89-108. [doi: 10.1002/for.3980090203]
25. Zhu C, Byrd R, Lu P, Nocedal J. BFGS-B: a limited memory FORTRAN code for solving bound constrained optimization problems. Evanston, IL: EECS Department, Northwestern University; 1994.
26. Peiris JSM, Guan Y, Yuen KY. Severe acute respiratory syndrome. Nat Med 2004 Dec 30;10(12 Suppl):S88-S97 [FREE Full text] [doi: 10.1038/nm1143] [Medline: 15577937]
27. Sistema com informações sobre Leitos de UTI, do SUS e da rede privada, associados a quantidade de habitantes e ao número de casos confirmados e de óbitos por COVID-19 (System with information on ICU, SUS and private network beds, associated with the number of inhabitants and the number of confirmed cases and deaths due to COVID-19). Instituto Brasileiro de Geografia e Estatística e Fundação Oswaldo Cruz. 2020. URL: https://leitos-ibgedgc.hub.arcgis.com/ [accessed 2020-06-03]
28. COVID-19 no Brasil: várias epidemias num só país Primeira fase do EPICOVID19 reforça preocupação com a região Norte (COVID-19 in Brazil: several epidemics in one country first phase of EPICOVID19 reinforces concern with the Northern region). EPICOVID19. 2020 May 25. URL: https://wp.ufpel.edu.br/covid19/files/2020/05/EPICOVID19BR-release-fase-1-Portugues.pdf [accessed 2020-05-26]
29. Russell T, Hellewell J, Abbott S, Golding N, Gibbs H, Jarvis CI, et al. Using a delay-adjusted case fatality ratio to estimate under-reporting. CMMID Repository. 2020. URL: https://cmmid.github.io/topics/covid19/global_cfr_estimates.html [accessed 2020-06-07]
30. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. The Lancet 2020 Mar 12;20(7):P773 [FREE Full text] [doi: 10.1016/S1473-3099(20)30195-X] [Medline: 26745243]
31. How Does Testing in the U.S. Compare to Other Countries? Johns Hopkins University of Medicine. 2020 Jun 11. URL: https://coronavirus.jhu.edu/testing/international-comparison [accessed 2020-06-09]

Abbreviations

COVID-19: coronavirus disease
DATASUS: Department of Informatics of the Unified Healthcare System
GDP: gross domestic product
HDI: Human Development Index
IBGE: Brazilian Institute of Geography and Statistics
ICD-10: International Statistical Classification of Diseases and Related Health Problems–10th Revision
PNAD: Pesquisa Nacional por Amostra de Domicílios (National Household Sample Survey)
R: reproduction number
RT-PCR: reverse transcription polymerase chain reaction
SARS: severe acute respiratory syndrome
SARS-CoV: severe acute respiratory syndrome coronavirus
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
SIM: Mortality Information System
SUS: Sistema Único de Saúde (Unified Healthcare System)
UFPel: Federal University of Pelotas
WHO: World Health Organization

Edited by G Eysenbach, G Fagherazzi; submitted 14.06.20; peer-reviewed by B Chen, A Rovetta; comments to author 18.07.20; revised version received 25.07.20; accepted 26.07.20; published 18.08.20.

Please cite as:
Veiga e Silva L, de Andrade Abi Harb MDP, Teixeira Barbosa dos Santos AM, de Mattos Teixeira CA, Macedo Gomes VH, Silva Cardoso EH, S da Silva M, Vijaykumar NL, Venâncio Carvalho S, Ponce de Leon Ferreira de Carvalho A, Lisboa Frances CR
COVID-19 Mortality Underreporting in Brazil: Analysis of Data From Government Internet Portals
J Med Internet Res 2020;22(8):e21413
URL: http://www.jmir.org/2020/8/e21413/
doi:10.2196/21413
PMID:32730219

Original Paper

Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence

Man Hung[1,2,3,4,5,6,7], PhD, MED, MSTAT, MSIS, MBA; Evelyn Lauren[8], BS; Eric S Hon[9]; Wendy C Birmingham[10], PhD; Julie Xu[11], BS; Sharon Su[1], BS; Shirley D Hon[12,13,14], MS; Jungweon Park[1], MS; Peter Dang[1], BS; Martin S Lipsky[1], MD

[1]College of Dental Medicine, Roseman University of Health Sciences, South Jordan, UT, United States
[2]Department of Orthopaedics, University of Utah, Salt Lake City, UT, United States
[3]George E Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT, United States
[4]Department of Occupational Therapy & Occupational Science, Towson University, Towson, MD, United States
[5]David Eccles School of Business, University of Utah, Salt Lake City, UT, United States
[6]Department of Educational Psychology, University of Utah, Salt Lake City, UT, United States
[7]Division of Public Health, University of Utah, Salt Lake City, UT, United States
[8]Department of Biostatistics, Boston University, Boston, MA, United States
[9]Department of Economics, University of Chicago, Chicago, IL, United States
[10]Department of Psychology, Brigham Young University, Provo, UT, United States
[11]College of Nursing, University of Utah, Salt Lake City, UT, United States
[12]Department of Electrical & Computer Engineering, University of Utah, Salt Lake City, UT, United States
[13]School of Computing, University of Utah, Salt Lake City, UT, United States
[14]International Business Machines Corporation, Poughkeepsie, NY, United States

Corresponding Author:
Man Hung, PhD, MED, MSTAT, MSIS, MBA
College of Dental Medicine
Roseman University of Health Sciences
10894 South River Front Parkway
South Jordan, UT, 84095-3538
United States
Phone: 1 801 878 1270
Email: mhung@roseman.edu

Abstract

Background: The coronavirus disease (COVID-19) pandemic led to substantial public discussion. Understanding these discussions can help institutions, governments, and individuals navigate the pandemic.

Objective: The aim of this study is to analyze discussions on Twitter related to COVID-19 and to investigate the sentiments toward COVID-19.

Methods: This study applied machine learning methods in the field of artificial intelligence to analyze data collected from Twitter. Using tweets originating exclusively in the United States and written in English during the 1-month period from March 20 to April 19, 2020, the study examined COVID-19–related discussions. Social network and sentiment analyses were also conducted to determine the social network of dominant topics and whether the tweets expressed positive, neutral, or negative sentiments. Geographic analysis of the tweets was also conducted.

Results: There were a total of 14,180,603 likes, 863,411 replies, 3,087,812 retweets, and 641,381 mentions in tweets during the study timeframe. Out of 902,138 tweets analyzed, sentiment analysis classified 434,254 (48.2%) tweets as having a positive sentiment, 187,042 (20.7%) as neutral, and 280,842 (31.1%) as negative. The study identified 5 dominant themes among COVID-19–related tweets: health care environment, emotional support, business economy, social change, and psychological stress. Alaska, Wyoming, New Mexico, Pennsylvania, and Florida were the states expressing the most negative sentiment while Vermont, North Dakota, Utah, Colorado, Tennessee, and North Carolina conveyed the most positive sentiment.

Conclusions: This study identified 5 prevalent themes of COVID-19 discussion with sentiments ranging from positive to negative. These themes and sentiments can clarify the public's response to COVID-19 and help officials navigate the pandemic.

(J Med Internet Res 2020;22(8):e22590) doi:10.2196/22590

KEYWORDS

COVID-19; coronavirus; sentiment; social network; Twitter; infodemiology; infodemic; pandemic; crisis; public health; business economy; artificial intelligence

Introduction

The outbreak of coronavirus disease (COVID-19) upended people's lives worldwide. COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a novel human pathogen that virologists believe emerged from bats and eventually jumped to humans via an intermediary host [1]. Clinical manifestations range from mild or no symptoms to more severe illness that may result in pulmonary failure and even death [2].

On March 11, 2020, the World Health Organization (WHO) declared COVID-19 a pandemic [3]. By June 23, the WHO reported 8,993,659 confirmed COVID-19 cases globally, and 469,587 deaths [4], and the Centers for Disease Control and Prevention (CDC) reported more than 2 million confirmed cases in the United States and more than 120,000 deaths [5]. These numbers illustrate how swiftly an emerging infection can spread.

For a novel virus without an available vaccine or highly effective antiviral drug therapy, community mitigation represents one strategy to slow the rate of infections. Community mitigation for COVID-19 consists of physical distancing including closing schools, bars, restaurants, movie theatres, and encouraging businesses to have their employees work from home. Large public gatherings such as festivals, graduations, and sporting events are discouraged or banned. The economic impact of mitigation devastated numerous businesses, while in the United States alone, over 40 million people filed initial unemployment claims [6].

Mitigation can also incorporate stay-at-home orders except for managing essential needs and for workers with an essential job. The isolation associated with mitigation is linked to stress, depression, fear and denial, exacerbation, and posttraumatic stress disorder (PTSD) [7-9]. Extended social isolation can exacerbate existing mental health problems, anxiety, and angry feelings. People in isolation may have also lost social support from families and friends. Resentment and resistance to these changes in daily life is becoming increasingly evident [10].

To inform personal decisions about health issues, individuals often use the media as a source of up-to-date information [11]. The intensity of information may make this particularly true for the COVID-19 pandemic. Despite daily information, major questions remain about viral spread, postrecovery immunity and drug therapy [12]. To interpret what may seem as information overload, many individuals turn to social media for clarification. There they can find an abundance of pandemic-related discussion about the economy, school closure, lack of medical supplies and personnel, and social distancing.

Unfortunately, media messaging may not always align with science and misinformation, baseless claims, and rumors can spread quickly. For example, commentary that SARS-CoV-2 originated as a Chinese conspiracy increased xenophobic sentiment toward Asian Americans [13]. The impact and speed of the COVID-19 pandemic mean that understanding public perception and how it affects behavior is critically important. Failing to do so creates both time and opportunity costs.

In contrast to traditional news reporting, which often takes weeks, social media messages are available in virtually real time [14]. These sources offer an opportunity for earlier insights into the public's reaction to the pandemic. Among social media sites, Twitter is the most popular form of social media used for health care information [15]. Previous studies indicate that Twitter can yield important public health information including tracking infectious disease outbreaks, natural disasters, drug use, and more [16].

Despite the importance of understanding the public reaction to COVID-19, gaps in the understanding of COVID-19–related themes remain. To address this gap, this study conducted a social network analysis of Twitter to examine social media discussions related to COVID-19 and to investigate social sentiments toward COVID-19–related themes. Study goals were twofold: to provide clarity about online COVID-19–related discussion themes and to examine sentiments associated with COVID-19. Findings from this study can shed light on unnoticed sentiments and trends related to the COVID-19 pandemic. The results should help guide federal and state agencies, business entities, schools, health care facilities, and individuals as they navigate the pandemic.

Methods

Data Source

Twitter is a microblogging and social network platform where users post and interact with messages called "tweets." With 166 million daily users [17], Twitter is a valuable data source for social media discussion related to national and global events. This study collected data from the Twitter website by applying machine learning (ML) methods used in the field of artificial intelligence. To be representative of the population, this study examined tweets originating from the United States during the 1-month period from March 20 to April 19, 2020. The study excluded tweets written in languages other than English or with geolocation outside of the United States. A modified Delphi method was used to identify potential keywords for the Twitter search. Specifically, one author reviewed the literature to identify potential key words. These keywords were then circulated among the other authors for feedback and to solicit

additional terms. After two cycles, consensus was obtained for the 13 keywords (Table 1) used to search Twitter posts related to COVID-19. Data extracted from Twitter consisted of the following: date of post, username, tweet content, likes count, replies count, retweets count, place, and mentions. The collected tweet set did not include the content of retweets and quoted tweets.

Table 1. Keywords for Twitter post search (N=1,001,380).

Keyword	Frequency, n
Coronavirus	250,849
Covid	340,522
COVID-19	108,035
SARS-CoV-2	670
Stay home	47,772
covid19	134,773
lockdown	46,452
shelter in place	9967
coronavirus truth	1694
outbreak	16,045
pandemic	135,879
quarantine	325,770
social distancing	65,725
hoax	14,703
be kind	4071
health heroes	88
ppe	48,710
isolation	22,459
homeschooling	3271
school cancelled	50
online teaching	475

Data Analyses

This study applied natural language processing, a form of ML, to process the tweets. To increase precision and to facilitate content analysis of the tweets, background noise such as URLs, hashtags, stop words, and tweets with less than three characters were removed. Lemmatization, a process of reducing the inflectional forms of words to a common root or a single term [18], was applied to the tweets as part of data cleaning. Topic modeling using Latent Dirichlet Allocation (LDA) [19] was used to extract the hidden semantic structures in the tweet posts. The LDA is an unsupervised ML method suitable for performing topic modeling. It groups common words into multiple topics and works well with short or long texts. The study employed several sets of topic modeling, with each set containing 5 to 10 topics, with the authors selecting the topic sets that looked more sensible and interpretable. Following the selection of a set of topics, the authors reviewed the top 10 words from each topic and by consensus developed a theme for each of the topics. Sentiment analysis using Valence Aware Dictionary and sEntiment Reasoner (VADER) determined whether the tweet posts expressed positive, neutral, or negative sentiments, as well as the degree of sentiments (also known as compound score or sentiment score). Sentiment scores were calculated for each theme, ranging from –1 to 1, with –1 representing the most negative sentiment and 1 representing the most positive sentiment. VADER, a sentiment analysis tool based on lexicons of sentiment-related words, allows automatic classification of each word in the lexicon as positive, neutral, or negative. Positive sentiment was categorized by having sentiment scores ≥0.05; neutral sentiment was categorized by sentiment scores between –0.05 and 0.05; and negative sentiment was defined by having sentiment scores ≤–0.05. A random sample of 300 tweets were manually coded as having positive, neutral, or negative sentiments by the investigators, and checked against the machine's output of sentiment classifications. Sensitivity and specificity were then calculated to evaluate the quality of the work done by the machine. The distribution of the sentiment and user social network connectivity were examined across themes. Centrality measures assessed importance, influence, and significance of the social network themes. Further analyses examined average sentiments across different states in the United States. Python (Version 3.8.2) [20] and R (Version 3.6.2; R Foundation for Statistical Computing) were used to collect and process the data as well as to conduct the data analyses.

To enable research reproducibility and ensure completely transparent methodologies and analyses, all of the computer code for data collection, data analyses, and figure generation are provided in Multimedia Appendix 1. Readers interested in replicating this study or conducting a similar study can reference these computer codes. Due to the large amount of data that needs to be processed and analyzed, using a supercomputer or other high-performance computing resources is recommended.

Results

During the 1-month data collection period, a total of 1,001,380 tweets were retrieved from 334,438 unique Twitter users, representing 12,203 cities within the 50 states in the United States and the District of Columbia. Figure 1 displays the number of tweets related to COVID-19 from March 20 to April 19, 2020. There was a gradual decline in the number of tweets over time. There was a total of 14,180,603 likes, 863,411 replies, 3,087,812 retweets, and 641,381 mentions. After accounting for background noise and performing lemmatization, there was a total of 902,138 tweets remaining, in which sentiment analysis classified 434,254 (48.2%) tweets as having positive COVID sentiment, 187,042 (20.7%) as having neutral COVID sentiment, and 280,842 (31.1%) as having negative COVID sentiment (Figure 2). Overall, positive tweets outweighed negative tweets with a ratio of 1.55 to 1. The most positive sentiment words consisted of “today,” “love,” “work,” “great,” “time,” “thank,” “think,” “right,” and “know,” whereas negative sentiment words related to “people,” “Trump,” “think,” “right,” “time,” “need,” “virus,” and “shit” (Figure 3 and Table 2). Examples of tweets expressing positive, neutral, and negative sentiments are displayed in Table 3. Sensitivity of the positive sentiments was 89.3% and specificity was 77.3%.

Figure 1. Number of tweets related to coronavirus disease (COVID-19) from March 20, 2020, to April 19, 2020.

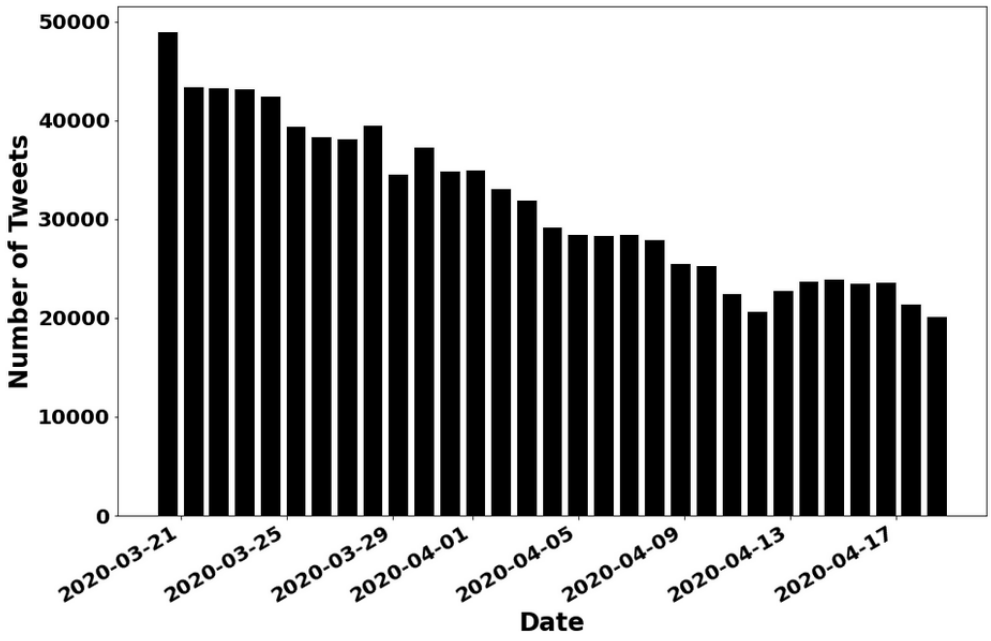


Figure 2. Frequency distribution of dominant topic tweets across sentiment types.

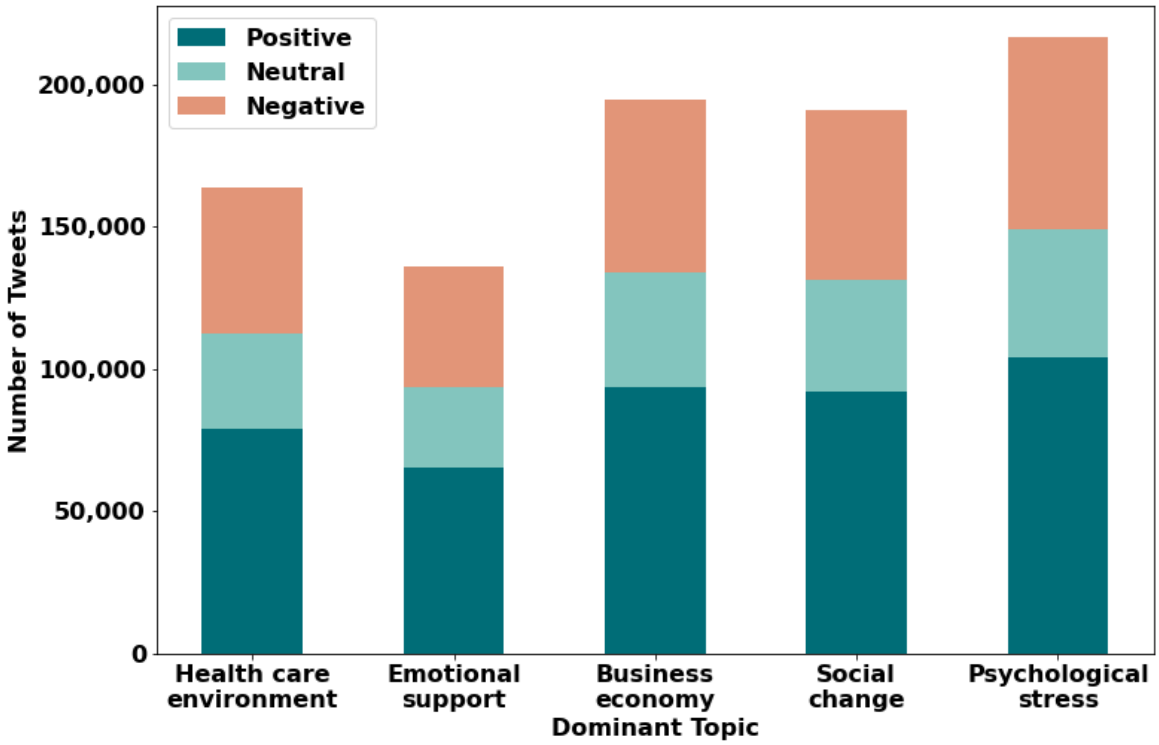


Figure 3. Word clouds showing the most frequently used word stems across Twitter users' post descriptions related to coronavirus disease (COVID-19). The upper left image is a word cloud formed from all tweets, while the upper right image is formed from tweets of positive sentiment. The lower left image is formed from tweets of neutral sentiment, while the lower right image is formed from tweets of negative sentiment.

Table 2. The most frequently used word stems across Twitter users' post descriptions related to coronavirus disease (COVID-19) by sentiment.

All tweets	Positive sentiment	Negative sentiment	Neutral sentiment
today	thank	people	time
time	today	time	today
right	time	trump	need
even	love	even	people
know	people	know	going
thank	need	right	week
think	know	need	know
stay home	right	think	right
going	work	shit	still
work	even	going	think
still	still	today	work
week	good	still	life
need	going	virus	even
love	think	week	thing
look	great	work	trump
year	friend	crisis	first
life	week	make	year
well	well	want	make
want	stay home	year	really
good	want	thing	stay home
said	life	really	everyone
virus	hope	said	come
people	look	fuck	getting
many	trump	life	made
friend	help	stay home	take
great	year	many	self
family	make	everyone	home
trump	family	state	back
come	better	stop	state
made	best	country	update
really	stay safe	come	virus
make	virus	world	family
mean	many	hospital	live
show	said	much	said
shit	made	look	gonna
self	getting	damn	many
hope	thing	mean	quarantinelife
thought	live	already	started
world	everyone	month	mean
first	world	well	call
another	really	president	much
live	show	china	show

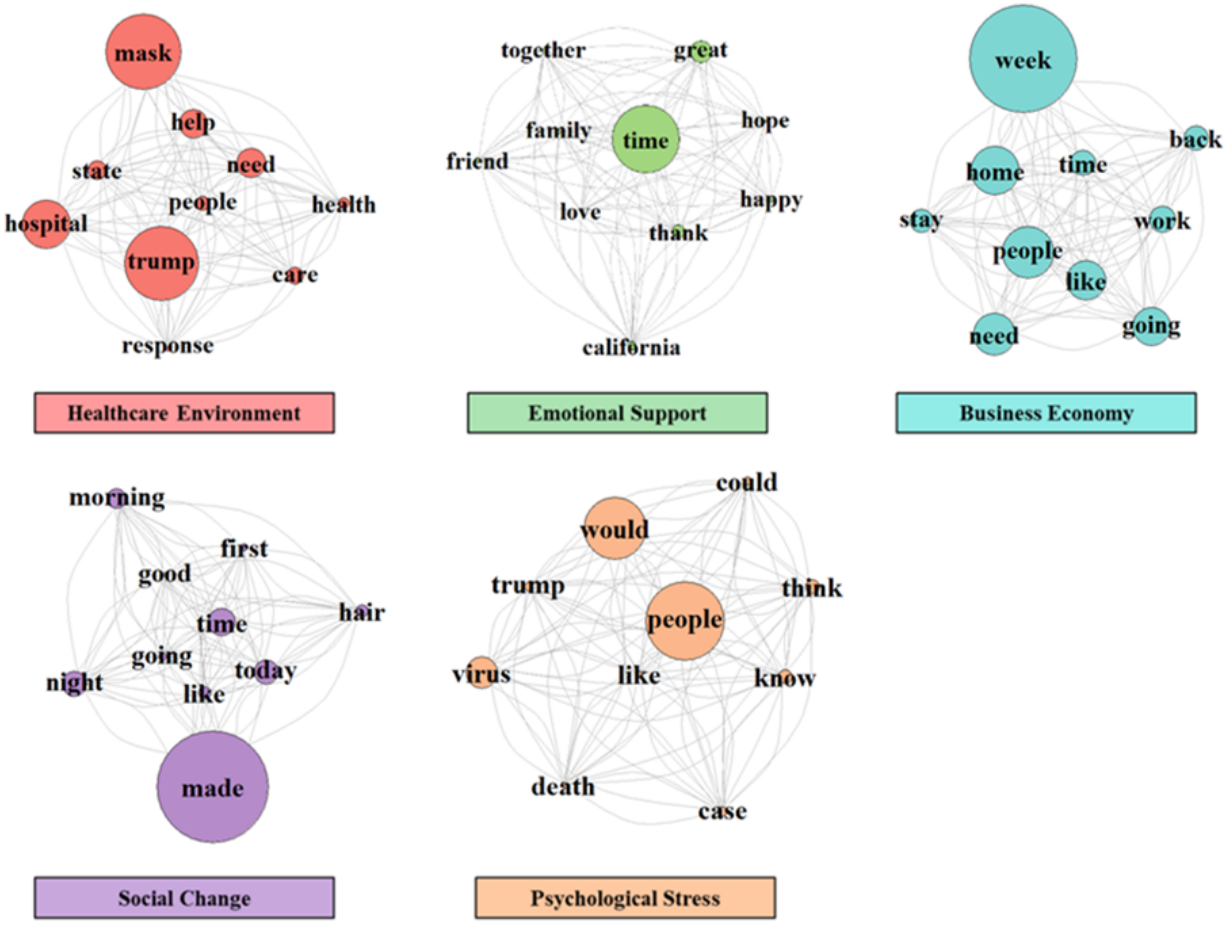
All tweets	Positive sentiment	Negative sentiment	Neutral sentiment
call	come	family	house
thing	self	take	look
already	feel	made	thought
everyone	first	getting	working
much	much	live	check
getting	back	first	done
feel	every	another	keep
little	another	nothing	another

Table 3. Examples of tweets expressing positive, neutral, and negative sentiments about coronavirus disease (COVID-19).

Positive sentiments	Neutral sentiments	Negative sentiments
• *great zoom call this morning. helping a handfull of clients and business associates. seems like a monday. quarantine won't stop business from happening. it just might look a little different.* • *this quarantine is doing me good on the writing* • *finding food pleasures in a time of #covid19 crisis* • *on the bright side I'll come out of this quarantine with gorgeous and glowing skin* • *thank you to all of the amazing nurses who are putting themselves on the frontline every day, being of great service to those who are in need during this pandemic. thank you for all you do* • *woke up to excellent news! i knew a person who was battling covid-19, on a ventilator in a different state. she is my age. she was discharged from the hospital yesterday and is now home! hooray!*	• *this is why social distancing is so important* • *day 9 of quarantine: i bought legos.... they should be here by friday* • *what i have learned from quarantine is that people like to do push ups and take shots* • *what if someone has #coronavirus with no symptom, can they donate blood?* • *i keep a list of all the people i have come in contact and the places i've been to since the social distancing and shelter at home started* • *praying my friend recovers from covid19*	• *so sick of covid-19, corona virus !!! tired of social distancing. tired of it all.* • *the only thing covid 19 will accomplish is turning a bunch of medical professionals into functioning alcoholics* • *had my first quarantine related super hardcore anxiety attack today* • *the saddest moment during the covid-19 pandemic was when muni stopped running the 38r* • *i should add that dad died before tests were available, and while his doctor said he believed it was covid-19, he could not definitely say that it was so. but this, too, is a failure of governance, given that we knew it was a threat in January* • *this is possibly the most insensitive (and couldn't possibly be true) ads i've seen in awhile. taking a victory lap over a financial crisis (because of a viral pandemic which is costing people much worse things than their finances) is gross.*

Topic modeling identified 5 salient topics that dominated Twitter discussions of COVID-19 and each of the 5 topics was labeled with a theme: health care environment [21], emotional support, business economy, social change, and psychological stress. Figure 4 displays 5 social network graphs, each corresponding to 1 of the 5 themes. Each social network graph shows the top 10 most frequently used words in tweets corresponding to a specific theme. The words are referred to as nodes or social actors when describing social network graphs, in which the size of the node represents the frequency of a certain word showing up. The lines between the words are referred to as links or actions, and these show the relationship between nodes. "Trump," "mask," and "hospital" dominated the discussion of health care environment. "Time" dominated the discussion of emotional support. "Week," "people," "home," "work," and "need" dominated business economy. For social change, "made," "today," and "time" dominated. In psychological stress, "people," "would," and "virus" dominated. Of note, Figure 4 reveals that among all 5 topics, the closeness centrality measure is the highest for emotional support, indicating that emotional support is the topic that is likely activated in each of the topic discussions.

Figure 4. Social network graphs of the dominant topics about coronavirus disease (COVID-19), with the top 10 associated words per topic. The size of the node is proportional to the weight of the edges.

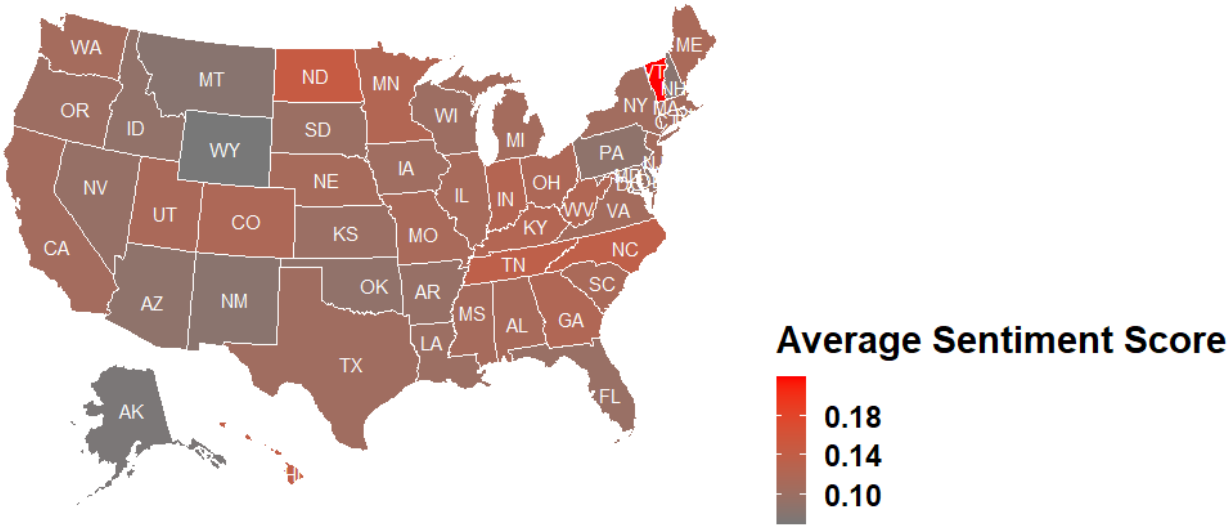


Social network centrality measures of the top 10 words on major COVID-19 themes.

Themes	Degree	Betweenness	Closeness	Eigenvector
Healthcare Environment	18	4.0	0.001885	0.5443
Emotional Support	18	3.6	0.009339	0.5834
Business Economy	18	1.3	0.000421	0.6495
Social Change	18	5.0	0.002602	0.5315
Psychological Stress	18	2.2	0.000656	0.5790

Figure 5 displays a heat map of the average sentiment score seen in each state in the United States. The darker the color of a certain state, the more negative the sentiment. Conversely, the lighter the color, the more positive the sentiment. Among the 50 states, Alaska, Wyoming, New Mexico, Pennsylvania, and Florida showed the most negative sentiment. Tweets from Vermont, North Dakota, Utah, Colorado, Tennessee, and North Carolina had the most positive sentiment in general.

Figure 5. Heat map of the average sentiment score by state in the United States. A larger number represents a more positive sentiment score.



Discussion

Overview

An unprecedented situation such as the COVID-19 outbreak makes it important to rapidly define the zeitgeist as new issues arise in a difficult time. This study found that overall, positive sentiment outweighed negative sentiment. Negative sentiment about COVID-19 was more prevalent in sparsely populated states with lower infection rates. Common negative sentiment tweet key words included “Trump,” “crisis,” “need,” “people,” “time,” “virus,” and “right.” Positive sentiment included words of encouragement and key phrases calling for the population to come together. These words included “thank,” “love,” “today,” “good,” and “friend.” The 5 most common themes were health care environment, business economy, emotional support, social change, and psychological stress, which represent the biggest concerns for the public.

Sentiments

Sentiment analysis is useful to capture public perception of an event. Overall, sentiments expressed in tweets about the pandemic were more likely to be positive, implying that the public remained hopeful in the face of an unprecedented public health crisis. Positive sentiment keywords commonly expressed gratitude for frontline workers and community efforts to support vulnerable members of the community, but a few keywords conveyed negative sentiments toward those on the frontline. Some words drew a similarity to frontline workers and soldiers fighting a war. Another form of positive sentiment was the encouragement of infection prevention to maintain public health standards, such as the “Stay Home” trend. Overall, positive sentiments outweighed negative sentiments. The high proportion of positive sentiments suggests that some people might have underestimated the severity of COVID-19 during the early period of the pandemic. One cautionary note is that tweets generated in states experiencing lower rates of infection tended to be most positive, suggesting that states more directly impacted by the pandemic were more likely to be negative. Strategies targeted to high-impact areas may be needed to keep the public engaged and hopeful about their futures.

The negative sentiment keywords suggest that tweets may be a way to vent negative feelings about consequences imposed by COVID-19 restrictions. Negative keywords commonly featured “know” and “think,” words that pertain to information and information sharing. Information sharing can play into the public’s risk perception, which can be affected by an individual’s trust in authorities and (in)ability to recognize misinformation [22]. A recent study found that most Americans trust the director of the CDC or National Institutes of Health to lead the COVID-19 response over Congress [23]. This study also found that the public generally supports infection prevention measures in the United States.

Twitter Discussion Themes Related to COVID-19

Health Care Environment

The discussions around the health care environment overlapped with politics. Vital supplies such as personal protective equipment and intensive care resources were linked to the need for government support. Health care workers discussed safety on the frontlines as an issue which negatively impacted their mental and physical health [24]. These findings tended to correlate to health care environments in COVID-19 hotspots affected by massive redistributions of resources [25].

Business Economy

A particularly stressful part of the shutdown has been its impact on the US and global economy [26]. Millions of people lost jobs, and unemployment numbers rose rapidly. The term “week” was a significant topic and was linked with conversations about “work”: the words “going,” “back,” and “need” were mentioned. Many people were unemployed or worked from home, and a common topic was “home,” “work,” and “stay” interlinked with the term “like.” These findings indicate that while discussions overwhelmingly centered around the week and work concerns, not all conversations related to loss of work. Some tweets indicated liking working from home. This finding suggests that strategies to reopen the economy that embrace working from home may both be popular and also help to reduce SARS-CoV-2 spread.

Emotional Support

Support from one's network of family and friends during periods of high stress can help reduce its harmful effects [27,28]. However, social isolation and distancing can preclude receiving the support needed. Our results showed that "time" was an overwhelming topic of discussion and clearly linked to "family," "friend," "together," and "hope." It may be that despite isolation, individuals find comfort and support through social media connections with their family and friends, while remaining hopeful they will soon have time together.

Social Change

The pandemic and its associated societal upheaval appeared to leave individuals in a state of uncertainty. Overwhelmingly, discussions focused on the word "made," with links to "today," "night," and "morning." Changes made to individuals' lives included daily activity restrictions, which could occur in different way across the entire day. Some changes could be "good," or individuals may "like" some changes. An interesting finding was "hair." Shutdown orders for nonessential businesses included hair salons, and these restrictions likely impact individuals' self-perception of their appearance and the way they maintain their personal grooming.

Psychological Stress

Psychological stress can be acute or chronic and both physiologically and psychologically detrimental [29,30]. Acute stress can become chronic if one is repeatedly exposed to stressful events. The current pandemic took what could have been an acute stress (eg, having the flu) and transformed it into a chronic stressful situation of worldwide sickness and death, and economic disruption. Our study found discussions of psychological stress overlapped with politics. Most discussions centered around "people," highly overlapping with "virus," "case," and "death." "Would," also featured prominently, along with "could" and "know," suggesting that individuals were concerned with the impact that the virus might have on people they know. Discussion also connected "Trump" with "people," "death," and "virus," suggesting that other discussions were focused on the role of the president in leading people to fight the virus and minimize death.

It is unclear why the number of tweets declined during the study period and followed a power law distribution as shown in Figure 1. One possible explanation is that most people tend to have more questions and discussions regarding a phenomenon when it is novel, but the discussions may slow down as time progresses. It is also noted that the behavior of extremely rare events such as stock market crashes and large natural disasters seem to follow the power law distribution [31]. Future research is needed to explore this fascinating pattern and understand the driving force behind the distribution.

The first case of COVID-19 in the United States was reported on January 19, 2020 [32]. By mid-March, all 50 states and 4 United States territories had reported cases. Of the 12,757 COVID-19–related reported deaths as of April 7, approximately half of all deaths were from New York and New Jersey with case-fatality ratios being lowest in Utah [26]. The more positive sentiment of Utah may be in part due to the lower incidences of reported cases and its having the lowest ratio of COVID-19 fatalities in April.

Potential Impact

Our results demonstrate that applying ML methods to mine pandemic-related tweets can yield useful data for agencies, local leaders, and health providers. For example, this study found that sentiments differed by region and geocaching tweets can allow localities to leverage data to match strategies and communications to community needs. When properly analyzed, digital data such as tweets can add to real-time epidemiologic data [33], allowing a more comprehensive and instantaneous evaluation of the pandemic situation. This is important since traditional public health data may take 1 to 2 weeks to become available. By virtue of the sheer volume, Twitter data might also help to identify or track rare event occurrences such as the multisystem inflammatory syndrome associated with COVID-19 in children [16].

In addition, Twitter offers an inexpensive and efficient platform to evaluate the effectiveness of public health communications [34], and to target public health campaigns on the dominant topics of Twitter discussion. For example, tweet analysis regarding mask wearing and hand hygiene can assess messaging. Applying ML to tweets can also provide insight into how the public interprets mixed messages regarding therapies such as hydroxychloroquine.

As the likelihood of a new coronavirus vaccine increases, one concern is that despite the established value of vaccines, only about half the public might elect to take a coronavirus vaccine [35,36]. Even a clinically proven vaccine depends on a high level of acceptance [37] and unsubstantiated concerns about negative side effects might overshadow the benefits of coronavirus immunization. Twitter offers an opportunity to follow vaccine acceptance and to tailor responses to those who oppose vaccination. The local risk of vaccine-preventable diseases can rise when there is a geographic aggregation of persons refusing vaccination and expressing more negative sentiments. Twitter analysis provides a potentially powerful and inexpensive tool for public health officials to identity geographic clusters for interventions and to evaluate their effectiveness.

Limitations

One limitation is that Twitter represents community interaction and its user profiles contain little demographic data, rendering an analysis of Twitter user demographic subgroups meaningless. An analysis of subgroups might yield more insights. Further, Twitter users do not fully represent the United States population, since only 15% of adults use Twitter, and younger adults aged 18 to 29 years old and minorities tend to be more active in Twitter discussions than the general population [38]. Additionally, active and passive Twitter users are more prevalent than moderate users [38]. With such potentially nonuniform sampling distribution and nonrepresentative demographic distribution of the United States population, the findings of specific sentiments can be biased [39], thus cautious interpretation of the findings is needed. However, the number of Twitter users over age 65 continues to increase, reducing the

level of age-related bias. Additionally, the overrepresentation of minorities may be a strength in terms of assessing health disparities.

The real-time posting of tweets is both a strength and a weakness. A strength is that it captures what is happening at the time, but a weakness is that tweet content can evolve very quickly [40], thus requiring constant monitoring of posts. In addition, the use of Twitter is not uniform across time or geography. Padilla et al [41] noted that Thursdays and Saturdays have slightly higher sentiment scores, but this study did not factor such differences into our analyses. Nevertheless, our results illustrate the insights that monitoring tweets can provide to a health-related event. Using ML to assess tweets is a potential weakness since it may not perform as well as human curation [42]. However, a strength is that ML processes a vast amount of data much faster than human methods. Finally, while social media may not capture the sentiment of those less vocal, a tweet analysis can provide insight into the type of information they process.

Conclusions

This study identified 5 overarching themes related to COVID-19: health care environment, emotional support, business economy, social change, and psychological stress. “Trump,” “mask,” and “hospital” dominated the tweets of health care environment. “Week,” “people,” “home,” “work,” and “need” dominated business economy. In psychological stress, “people,” “would,” and “virus” dominated the discussion. Overall, positive tweets outweighed negative tweets. The sentiments can clarify the public response to COVID-19 and help guide government officials, private entities, and the public with information as they navigate the pandemic.

Acknowledgments

The authors would like to thank the Clinical Outcomes Research and Education at Roseman University of Health Sciences College of Dental Medicine for supporting this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Computer code for data collection, data analyses, and figure generation.
[DOCX File , 57 KB - jmir_v22i8e22590_app1.docx]

References

1. Zu ZY, Jiang MD, Xu PP, Chen W, Ni QQ, Lu GM, et al. Coronavirus Disease 2019 (COVID-19): A Perspective from China. Radiology 2020 Aug;296(2):E15-E25 [FREE Full text] [doi: 10.1148/radiol.2020200490] [Medline: 32083985]
2. Cascella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R. Features, Evaluation and Treatment Coronavirus (COVID-19). Treasure Island, FL, USA: StatPearls; 2020.
3. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. 2020. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-06-24]
4. Coronavirus disease (COVID-19) situation reports. World Health Organization. 2019. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-06-24]
5. Coronavirus Disease 2019 (COVID-19). Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html [accessed 2020-06-24]
6. U.S. Jobless Claims Pass 40 Million: Live Business Updates. The New York Times. 2020. URL: https://www.nytimes.com/2020/05/28/business/unemployment-stock-market-coronavirus.html [accessed 2020-05-28]
7. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The Lancet 2020 Mar;395(10227):912-920. [doi: 10.1016/s0140-6736(20)30460-8]
8. Cava MA, Fay KE, Beanlands HJ, McCay EA, Wignall R. Risk perception and compliance with quarantine during the SARS outbreak. J Nurs Scholarsh 2005;37(4):343-347. [doi: 10.1111/j.1547-5069.2005.00059.x] [Medline: 16396407]
9. Hawryluck L, Gold WL, Robinson S, Pogorski S, Galea S, Styra R. SARS control and psychological effects of quarantine, Toronto, Canada. Emerg Infect Dis 2004 Jul;10(7):1206-1212 [FREE Full text] [doi: 10.3201/eid1007.030703] [Medline: 15324539]
10. Studdert DM, Hall MA. Disease Control, Civil Liberties, and Mass Testing — Calibrating Restrictions during the Covid-19 Pandemic. N Engl J Med 2020 Jul 09;383(2):102-104. [doi: 10.1056/nejmp2007637]
11. Ball-Rokeach S, DeFleur M. A Dependency Model of Mass-Media Effects. Communication Research 2016 Jun 30;3(1):3-21. [doi: 10.1177/009365027600300101]
12. Ye G, Pan Z, Pan Y, Deng Q, Chen L, Li J, et al. Clinical characteristics of severe acute respiratory syndrome coronavirus 2 reactivation. J Infect 2020 May;80(5):e14-e17 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.001] [Medline: 32171867]

13. Wang P, Anderson N, Pan Y, Poon L, Charlton C, Zelyas N, et al. The SARS-CoV-2 Outbreak: Diagnosis, Infection Prevention, and Public Perception. Clin Chem 2020 Mar 10:2020 [FREE Full text] [doi: 10.1093/clinchem/hvaa080] [Medline: 32154877]
14. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg 2012 Jan;86(1):39-45 [FREE Full text] [doi: 10.4269/ajtmh.2012.11-0597] [Medline: 22232449]
15. Pershad Y, Hangge P, Albadawi H, Oklu R. Social Medicine: Twitter in Healthcare. J Clin Med 2018 May 28;7(6):121 [FREE Full text] [doi: 10.3390/jcm7060121] [Medline: 29843360]
16. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May 04;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
17. Wong Q, Skillings J. Twitter's user growth soars amid coronavirus, but uncertainty remains. Cnet. 2020 Apr 30. URL: https://www.cnet.com/news/twitters-user-growth-soars-amid-coronavirus-but-uncertainty-remains/ [accessed 2020-08-13]
18. Liu H, Christiansen T, Baumgartner WA, Verspoor K. BioLemmatizer: a lemmatization tool for morphological processing of biomedical text. J Biomed Semantics 2012 Apr 01;3(1):3 [FREE Full text] [doi: 10.1186/2041-1480-3-3] [Medline: 22464129]
19. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, Latin American Network of Coronavirus Disease 2019-COVID-19 Research (LANCOVID-19). Electronic address: https://www.lancovid.org. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. Travel Med Infect Dis 2020 Mar;34:101623 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101623] [Medline: 32179124]
20. Python Software Foundation. Python. URL: https://www.python.org/ [accessed 2020-07-08]
21. Dhanuthai K, Rojanawatsirivej S, Thosaporn W, Kintarak S, Subarnbhesaj A, Darling M, et al. Oral cancer: A multicenter study. Med Oral Patol Oral Cir Bucal 2018 Jan 01;23(1):e23-e29. [doi: 10.4317/medoral.21999] [Medline: 29274153]
22. Cori L, Bianchi F, Cadum E, Anthonj C. Risk Perception and COVID-19. Int J Environ Res Public Health 2020 Apr 29;17(9):2020 [FREE Full text] [doi: 10.3390/ijerph17093114] [Medline: 32365710]
23. McFadden SM, Malik AA, Aguolu OG, Willebrand KS, Omer SB. Perceptions of the adult US population regarding the novel coronavirus outbreak. PLoS One 2020 Apr 17;15(4):e0231808 [FREE Full text] [doi: 10.1371/journal.pone.0231808] [Medline: 32302370]
24. Delgado D, Wyss Quintana F, Perez G, Sosa Liprandi A, Ponte-Negretti C, Mendoza I, et al. Personal Safety during the COVID-19 Pandemic: Realities and Perspectives of Healthcare Workers in Latin America. Int J Environ Res Public Health 2020 Apr 18;17(8):2798 [FREE Full text] [doi: 10.3390/ijerph17082798] [Medline: 32325718]
25. Ranney ML, Griffeth V, Jha AK. Critical Supply Shortages — The Need for Ventilators and Personal Protective Equipment during the Covid-19 Pandemic. N Engl J Med 2020 Apr 30;382(18):e41. [doi: 10.1056/nejmp2006141]
26. Congressional Research Service. Global Economic Effects of COVID-19. 2020. URL: https://fas.org/sgp/crs/row/R46270.pdf [accessed 2020-07-02]
27. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]
28. Cobb S. Presidential Address-1976. Social support as a moderator of life stress. Psychosom Med 1976;38(5):300-314. [doi: 10.1097/00006842-197609000-00003] [Medline: 981490]
29. Serido J, Almeida DM, Wethington E. Chronic Stressors and Daily Hassles: Unique and Interactive Relationships with Psychological Distress. J Health Soc Behav 2016 Jun 23;45(1):17-33. [doi: 10.1177/002214650404500102]
30. Kiecolt-Glaser JK, Glaser R, Gravenstein S, Malarkey WB, Sheridan J. Chronic stress alters the immune response to influenza virus vaccine in older adults. Proc Natl Acad Sci U S A 1996 Apr 02;93(7):3043-3047 [FREE Full text] [doi: 10.1073/pnas.93.7.3043] [Medline: 8610165]
31. Power law. Wikipedia. 2020. URL: https://en.wikipedia.org/wiki/Power_law [accessed 2020-08-13]
32. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First Case of 2019 Novel Coronavirus in the United States. N Engl J Med 2020 Mar 05;382(10):929-936. [doi: 10.1056/nejmoa2001191]
33. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoS Comput Biol 2012 Jul 26;8(7):e1002616 [FREE Full text] [doi: 10.1371/journal.pcbi.1002616] [Medline: 22844241]
34. Gesualdo F, Stilo G, Agricola E, Gonfiantini MV, Pandolfi E, Velardi P, et al. Influenza-like illness surveillance on Twitter through automated learning of naïve language. PLoS One 2013 Dec 4;8(12):e82489 [FREE Full text] [doi: 10.1371/journal.pone.0082489] [Medline: 24324799]
35. Cornwall W. Just 50% of Americans plan to get a COVID-19 vaccine. Here's how to win over the rest. Science 2020 Jul 01;29:A [FREE Full text] [doi: 10.1126/science.abd6018]
36. Only 57 percent of Americans say they would get a COVID-19 vaccine. Medicalxpress. 2020 Jul 10. URL: https://medicalxpress.com/news/2020-07-percent-americans-covid-vaccine.html [accessed 2020-08-13]
37. Omer SB, Salmon DA, Orenstein WA, deHart MP, Halsey N. Vaccine Refusal, Mandatory Immunization, and the Risks of Vaccine-Preventable Diseases. N Engl J Med 2009 May 07;360(19):1981-1988. [doi: 10.1056/nejmsa0806477]

38. Smith A, Brenner J. Twitter Use 2012. Pew Research Center. 2012 Jul 29. URL: https://www.pewresearch.org/internet/2012/05/31/twitter-use-2012/ [accessed 2020-07-29]
39. Gore RJ, Diallo S, Padilla J. You Are What You Tweet: Connecting the Geographic Variation in America's Obesity Rate to Twitter Content. PLoS One 2015 Sep 2;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
40. Duggan M, Ellision NB, Lampe C, Lenhart A, Madden M. Social Media Update 2014. Pew Research Center. 2014. URL: https://www.pewresearch.org/internet/2015/01/09/social-media-update-2014/ [accessed 2020-07-08]
41. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018 Jun 14;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
42. Hawkins JB, Brownstein JS, Tuli G, Runels T, Broecker K, Nsoesie EO, et al. Measuring patient-perceived quality of care in US hospitals using Twitter. BMJ Qual Saf 2016 Jun 13;25(6):404-413 [FREE Full text] [doi: 10.1136/bmjqs-2015-004309] [Medline: 26464518]

Abbreviations

COVID-19: coronavirus disease
CDC: Centers for Disease Control and Prevention
LDA: Latent Dirichlet Allocation
ML: machine learning
PTSD: posttraumatic stress disorder
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
VADER: Valence Aware Dictionary and sEntiment Reasoner
WHO: World Health Organization

Edited by G Eysenbach; submitted 17.07.20; peer-reviewed by JM Chen, R Gore; comments to author 26.07.20; revised version received 30.07.20; accepted 03.08.20; published 18.08.20.

Please cite as:
Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, Hon SD, Park J, Dang P, Lipsky MS
Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence
J Med Internet Res 2020;22(8):e22590
URL: http://www.jmir.org/2020/8/e22590/
doi:10.2196/22590
PMID:32750001

Review

Diagnostic Value of Imaging Modalities for COVID-19: Scoping Review

Rowa Aljondi[1*], PhD; Salem Alghamdi[1*], PhD

Department of Applied Radiologic Technology, College of Applied Medical Sciences, University of Jeddah, Jeddah, Saudi Arabia

[*]all authors contributed equally

Corresponding Author:
Rowa Aljondi, PhD
Department of Applied Radiologic Technology
College of Applied Medical Sciences
University of Jeddah
Al Sharafeyah
Jeddah, 23218
Saudi Arabia
Phone: 966 530550559
Email: rowaaljondi@gmail.com

Abstract

Background: Coronavirus disease (COVID-19) is a serious infectious disease that causes severe respiratory illness. This pandemic represents a serious public health risk. Therefore, early and accurate diagnosis is essential to control disease progression. Radiological examination plays a crucial role in the early identification and management of infected patients.

Objective: The aim of this review was to identify the diagnostic value of different imaging modalities used for diagnosis of COVID-19.

Methods: A comprehensive literature search was conducted using the PubMed, Scopus, Web of Science, and Google Scholar databases. The keywords *diagnostic imaging*, *radiology*, *respiratory infection*, *pneumonia*, *coronavirus infection* and COVID-19 were used to identify radiology articles focusing on the diagnosis of COVID-19 and to determine the diagnostic value of various imaging modalities, including x-ray, computed tomography (CT), ultrasound, and nuclear medicine for identification and management of infected patients.

Results: We identified 50 articles in the literature search. Studies that investigated the diagnostic roles and imaging features of patients with COVID-19, using either chest CT, lung ultrasound, chest x-ray, or positron emission topography/computed tomography (PET/CT) scan, were discussed. Of these imaging modalities, chest x-ray and CT scan are the most commonly used for diagnosis and management of COVID-19 patients, with chest CT scan being more accurate and sensitive in identifying COVID-19 at early stages. Only a few studies have investigated the roles of ultrasound and PET/CT scan in diagnosing COVID-19.

Conclusions: Chest CT scan remains the most sensitive imaging modality in initial diagnosis and management of suspected and confirmed patients with COVID-19. Other diagnostic imaging modalities could add value in evaluating disease progression and monitoring critically ill patients with COVID-19.

(J Med Internet Res 2020;22(8):e19673) doi:10.2196/19673

KEYWORDS

diagnostic imaging; radiology; COVID-19; respiratory infection; pneumonia; imaging; CT; infectious disease; diagnosis; review

Introduction

Coronavirus disease (COVID-19) is a viral respiratory disease that first emerged in December 2019, when a cluster of patients with unknown pneumonia was reported in Wuhan City in Hubei Province in China. The causative agent of this unknown pneumonia was a novel coronavirus, later known as novel coronavirus pneumonia (NCP) [1,2]. This virus was then renamed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses based on phylogeny, taxonomy, and established practice [3,4]. Compared with previous coronaviruses, such as severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS-CoV), SARS-CoV-2 is highly contagious and

transmissible from person to person [5]. The disease caused by SARS-CoV-2 was officially named coronavirus disease 2019 (COVID‐19) by the World Health Organization (WHO) [6]. It quickly spread to other countries worldwide, causing an increasing number of deaths [7,8]. Accordingly, on January 30, 2020, the WHO declared COVID_19 an international public health emergency [9].

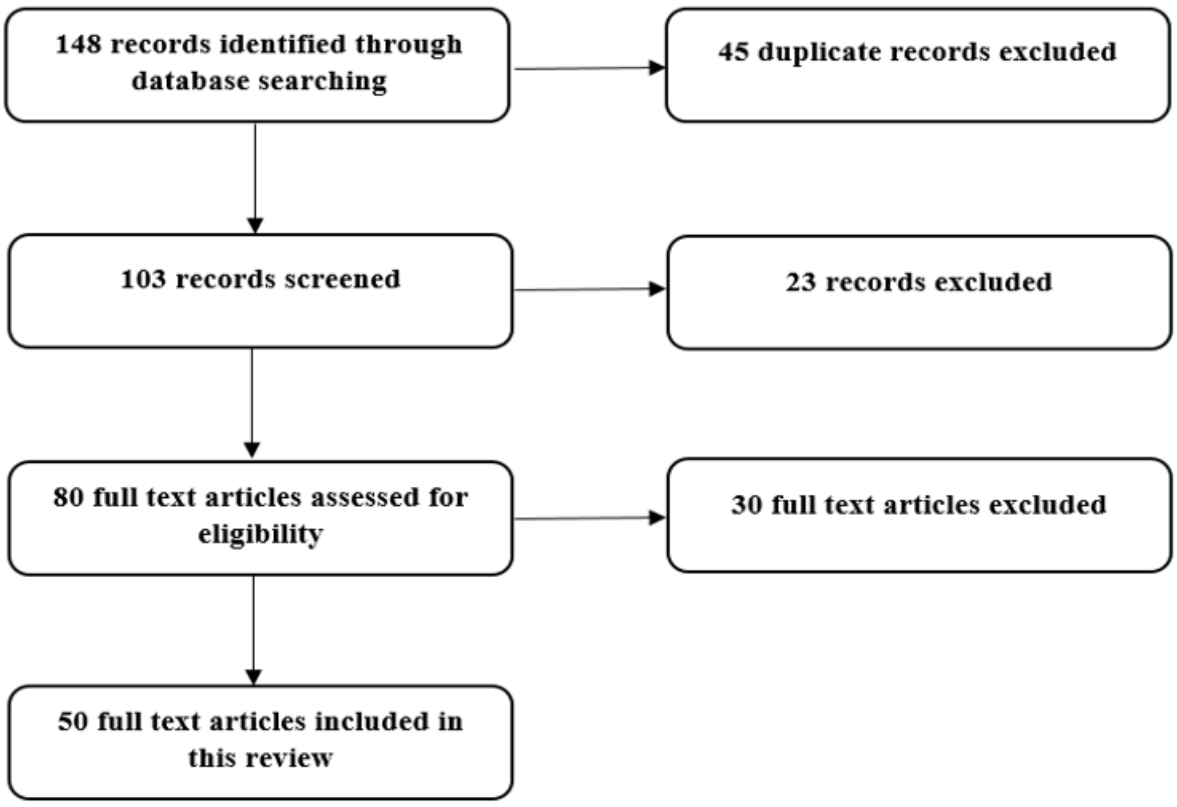
The most common clinical symptoms are fever, dry cough, fatigue, and gradual development of dyspnea [10-12]. The current gold standard clinical diagnostic tool for COVID-19 is the reverse transcription–polymerase chain reaction (RT-PCR) analysis of specimens from the respiratory tract. However, this test shows high false negative results due to inadequate cellular material or errors in detection and extraction techniques during nasopharyngeal swab sampling [13-15]. With an increasing number of infected patients and a shortage of RT‐PCR testing kits in affected areas, alternative diagnostic and screening strategies are needed [16]. As such, diagnostic imaging now plays a critical role in identifying and assessing the progression of COVID-19 [17].

Recently, radiological literature has focused on chest computed tomography (CT) findings in COVID-19 [18-23]. Excessive use of CT scans can place a substantial burden on radiology departments in practice and increase the risk of infection in CT units [19]. However, other imaging modalities, such as chest x-ray, ultrasound, and positron emission topography/computed tomography (PET/CT), have also been used in the diagnosis and management of patients with COVID-19 [19]. Thus, radiologists should be aware of the roles and diagnostic value of various other imaging modalities for COVID-19 that can help manage disease progression [7,24]. In this literature review, we discuss the diagnostic value of each imaging modality commonly used in the diagnosis and evaluation of patients with COVID-19.

Methods

A literature search was performed on April 28, 2020, using the PubMed, Scopus, and Web of Science databases. The keywords *diagnostic imaging*, *radiology*, *respiratory infection*, *pneumonia*, *coronavirus infection*, and *COVID-19* were used to identify articles focusing on the diagnostic value of different imaging modalities used for diagnosis and management of patients with COVID-19. To increase the sensitivity of the search, Google Scholar was employed with the same keywords, capturing the most recently published articles in the field of imaging for COVID-19. This Google Scholar search was limited to selected keywords in the article titles due to the large number of records identified from the literature. All searches were limited to articles published in 2020, with consideration of the earliest date of confirmed COVID-19 reports. For inclusion in this literature review, articles were required to be original research, peer-reviewed, and written in English. Nonscientific commentary and news articles were excluded. Figure 1 illustrates the literature search process and article identification.

Figure 1. Flowchart showing the article identification and selection process.



Results

In the literature search, 50 articles were identified, covering areas such as chest CT scan, chest radiography (x-ray), nuclear medicine, and ultrasound in diagnosis and management of suspected and confirmed cases of COVID-19 (Multimedia Appendix 1) [2,8,13-15,18-22,25-65]. The selected articles are discussed in the following sections.

Discussion

Computed Tomography

Chest computed tomography (CT) is considered to be the primary diagnostic modality for examining patients with COVID-19 [20,25,26]. A large number of existing studies have investigated the CT image manifestations in COVID-19 cases [8,19,25,26,28-30,32,64]. The most common chest CT imaging features of COVID-19 pneumonia include peripheral ground-glass opacities (GGOs) and consolidation in the lower and middle lung regions, usually bilaterally distributed and with multi-lobar involvement (Figure 2) [19,26,31,32,63]. On the initial CT images, some patients with COVID-19 have pulmonary nodules that increase in size and number in follow-up CT [22,25]. Regarding infection duration, Bernheim et al [20] identified the most common CT findings at a longer time after symptom onset, including consolidation, linear opacities, bilateral and peripheral disease, crazy paving pattern, and reserved halo sign, which indicate greater total lung involvement. However, pneumothorax, pleural effusion, lymphadenopathy, pericardial effusion, and lung cavitation are uncommon findings on chest CT imaging of patients with COVID-19 that can be seen with disease progression in follow-up CT images [22].

The chest CT patterns of COVID-19 pneumonia are likely related to the pathological changes in the lungs [27,29]. Although the pathological process of lung injury in patients with COVID-19 pneumonia has not yet been studied, recent research has reported that SARS-CoV-2 shares a similar pneumonia pathogenesis to the severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) coronaviruses [31,33,64]. In SARS patients, angiotensin-converting enzyme 2 is a molecule that is potentially involved in the development and progression of acute lung failure [33,64]. The SARS virus induces lung injury by affecting this enzyme, which contributes to injury of pulmonary epithelial cells, diffuse alveolar damage, and edema [33,64]. These results may explain the pathological basis of GGO and consolidation as well as the rapid changes in chest CT imaging in patients with COVID-19 [33,64].

The chest CT manifestations of COVID-19 vary at different stages of the disease, which helps differentiate the diagnosis of COVID-19 from those of other known pneumonia viruses, such as mycoplasma pneumonia and bacterial pneumonia [29,31,34]. In pneumonia caused by SARS and MERS, the chest CT shows a unifocal involvement of lung lesions more than multifocal involvement, which is found on the chest CT images of patients with COVID-19 [64]. In patients with MERS, the GGOs are mainly distributed in the subpleural and basilar lung regions in the chest CT images. The chest CT imaging of patients with SARS shows that multiple GGOs are distributed in the periphery of the lung, with interlobular septal thickening and intralobular interstitial hyperplasia. These findings are similar to the chest CT features of patients with COVID-19 [31]. With the progression of COVID-19 pneumonia, the number of GGOs increases and the consolidations become denser [64]. These findings indicate that there are differences as well as similarities between viruses of the same family [29,31,34].

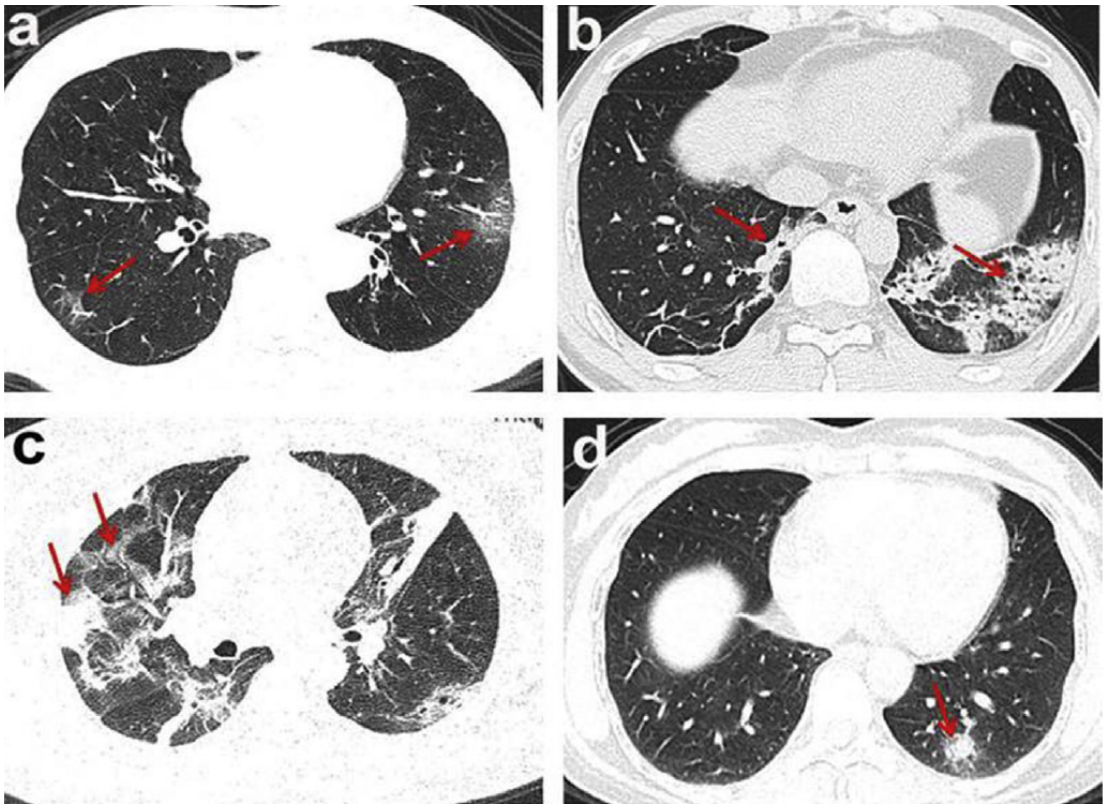
CT imaging has proven to be diagnostic in early stages of COVID-19 [25,29]. In a recent report involving 1014 patients in Wuhan, China, the sensitivity, specificity, diagnostic accuracy, positive predictive value, and negative predictive value for identifying COVID-19 infection using RT-PCR results as reference standards were 97%, 25%, 68%, 65%, and 83%, respectively [14]. Most patients with COVID-19 who initially showed negative RT-PCR results presented lung abnormalities on their chest CT images [13-15,35]. Moreover, abnormal chest CT findings have been observed in all patients with positive laboratory-confirmed COVID-19 infection by RT-PCR [2]. Similarly, Li and Xia [64] evaluated the diagnostic performance of chest CT for COVID-19 in cases confirmed by nucleic acid test and found that the initial chest CT scan showed a low misdiagnosis rate of COVID-19 viral pneumonia (3.9%) as a common infection. These studies suggest that chest CT imaging is an effective method for early diagnosis of COVID-19, particularly in regions with a shortage of RT - PCR testing kits.

The diagnostic value of chest CT examination mainly lies in its short examination time and high resolution in the detection and classification of lung lesions. Additionally, CT scanning is highly reproducible and easy to perform [29], and it provides a rapid and accurate estimation of disease progression. The main advantages of chest CT in detecting lung lesions in patients with COVID-19 are early characterization of lung lesions, assessment of disease severity, and improvement of lung lesions during treatment [29]. Thus, numerous studies emphasize follow-up CT examination in COVID-19 patients [21,25,26,29,36,42,64]. Among these studies, Pan et al [25] examined initial and follow-up imaging features in patients with confirmed COVID-19 using high resolution CT scans; they found diverse and rapidly changing imaging signs with disease progression in COVID-19 pneumonia. Li and Xia [64] also reported that 75% of their patients with COVID-19 showed disease progression in follow-up CT examination. One other study that examined changes in CT image findings from initial diagnosis of COVID-19 pneumonia until patient recovery observed that lung abnormalities increased to consolidations approximately 10 days after initial symptom onset [21]. In patients who recovered from COVID-19 pneumonia, chest CT abnormalities gradually decreased two weeks after initial onset of symptoms. Therefore, rapid and accurate diagnosis of COVID-19 based on CT imaging features in conjunction with clinical and laboratory findings may be useful in early control of potential transmission and optimizing management of patients with suspected disease so they can be treated and isolated promptly [36,64].

While chest CT imaging is the most sensitive modality for the early detection of lung disease and management of patients with COVID-19, it has low specificity for distinguishing lung lesions

of COVID-19 pneumonia from findings of other viral pneumonia caused by SARS and MERS [14,64]. In addition, it poses an increased risk of infection transmission to other patients or health care workers; as such, thorough cleaning is required, causing downtime of the scanning room and consumption of personal protection equipment [38]. Accordingly, the American College of Radiology (ACR) does not recommend using a CT scan for diagnosis of COVID-19 as a first-line test [41]. Moreover, before scanning subsequent patients, appropriate infection control techniques should be applied [41]. To control this pandemic, chest CT imaging should be reserved for screening patients with COVID-19 pneumonia complications [18,37,41].

Figure 2. Chest computed tomography findings in patients with coronavirus disease [63]. (a) Ground-glass opacities, (b) consolidations, (c) consolidations with ground-glass opacities, (d) solid nodules.



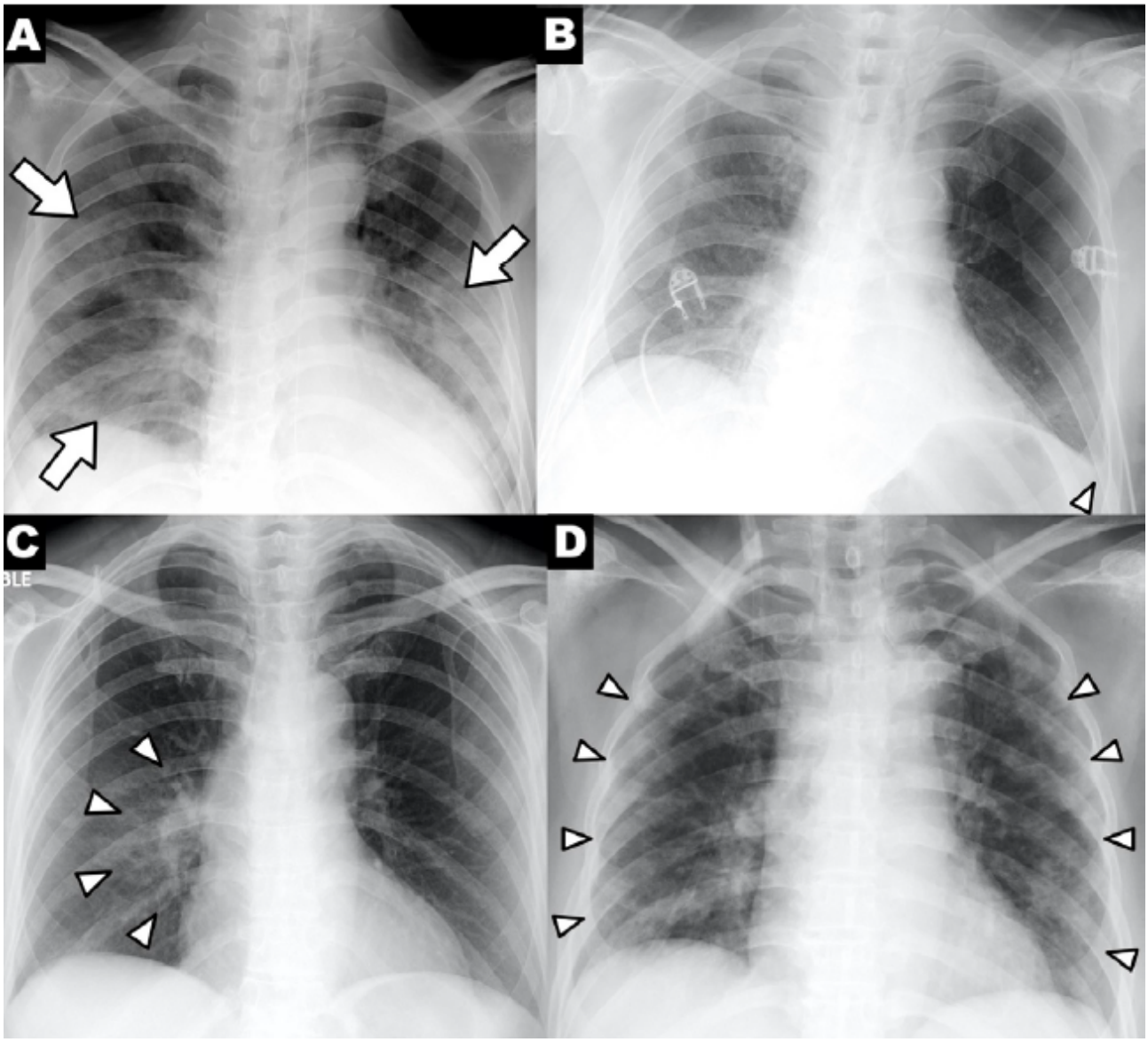
Chest Radiography

Chest x-ray is the most commonly used diagnostic imaging modality for patients with COVID-19 [44]. Still, chest radiography has limited sensitivity for COVID-19 in early stage of infection, as initial chest x-rays usually indicate normal lung appearance [18,19]. Therefore, it was not recommended as the first-line imaging modality for diagnosis of suspected COVID-19 patients [45]. Several studies report that chest x-ray often shows no image abnormalities in patients with COVID-19 at early stages (< 2-4 days) [18,19]. Wong et al [18] reported that the severity and abnormalities of chest x-ray in COVID-19 patients appeared 10 to 12 days after initial symptom onset. Using the RT-PCR results as the gold standard, the reported sensitivity of baseline chest x-ray for diagnosis of COVID-19 in mild to moderate cases was 69% [18]. When imaging is abnormal in severe cases, the most common features of chest x-rays are consolidation and GGOs with bilateral involvement and/or peripheral distribution (Figure 3) [18,28,43]. In COVID-19 patients, pleural effusions, lung cavitation, and pneumothorax were reported as rare findings on chest x-rays and could occur late in the disease course [22].

As the pandemic progresses, chest x-rays could play a vital role in disease identification for patients with high clinical suspicion of COVID-19 [44]. The ACR recommends using portable chest x-ray units to reduce the risk of cross-infection in radiology departments [41]. The advantages of using a portable chest x-ray unit is that it can detect the most common manifestations and patterns of COVID-19 lung abnormality in clinically confirmed cases of COVID-19, thereby limiting use of the CT scanner [44]. Jacobi et al [44] suggested using portable chest x-ray units for the diagnosis and follow-up of patients with high clinical suspicion of COVID-19 due to their cost-effectiveness and widespread availability; this modality can be used in emergency or intensive care units to minimize infection risks [44].

The limitation of chest x-ray imaging is its lack of sensitivity in detecting lung lesions in the early stage of COVID-19 pneumonia [18,19,43]. In addition, there is a paucity of reported specificity and diagnostic accuracy of chest x-ray in various stages of COVID-19 pneumonia. Thus, it is difficult to use chest x-rays to distinguish COVID-19 from pneumonia caused by other coronaviruses such as MERS and SARS. The positive predictive value and negative predictive value of chest x-rays for COVID-19 pneumonia have not yet been established.

Figure 3. Chest x-ray findings in a patient with coronavirus disease. (A) Patchy consolidations, (B) pleural effusion, (C) perihilar distribution, (D) peripheral distribution [18].



Nuclear Medicine

Combined positron emission topography/computed tomography (PET/CT) is an established nuclear medicine method for assessing and monitoring several lung diseases. ^{18}F-labelled fluorodeoxyglucose (^{18}F-FDG**)** is the most commonly used radiotracer in PET/CT imaging; it shows metabolic activity in inflammatory cells [46]. However, the role of nuclear medicine in the diagnosis and management of COVID-19 appears to be limited [47].

Qin et al [48] reported the findings of ^{18}F-FDG-PET/CT imaging in a case series of four patients with COVID-19. All patients presented with peripheral GGOs and/or consolidation in pulmonary lobes, showing high tracer uptake with a maximum standardized uptake value (SUV max) range of 1.8-12.2. In addition, increased nodal FDG uptake was observed in three out of four patients. Although lymphadenopathy was rare in CT imaging findings in patients with COVID-19 infection [32], these high tracer uptakes in pulmonary or lymph nodal lesions reflect a significant inflammatory burden, which is a hallmark of COVID-19 pulmonary infections [48]. As a noninvasive imaging modality, FDG-PET/CT can play a potential role in the evaluation of lung function in COVID-19 pneumonia.

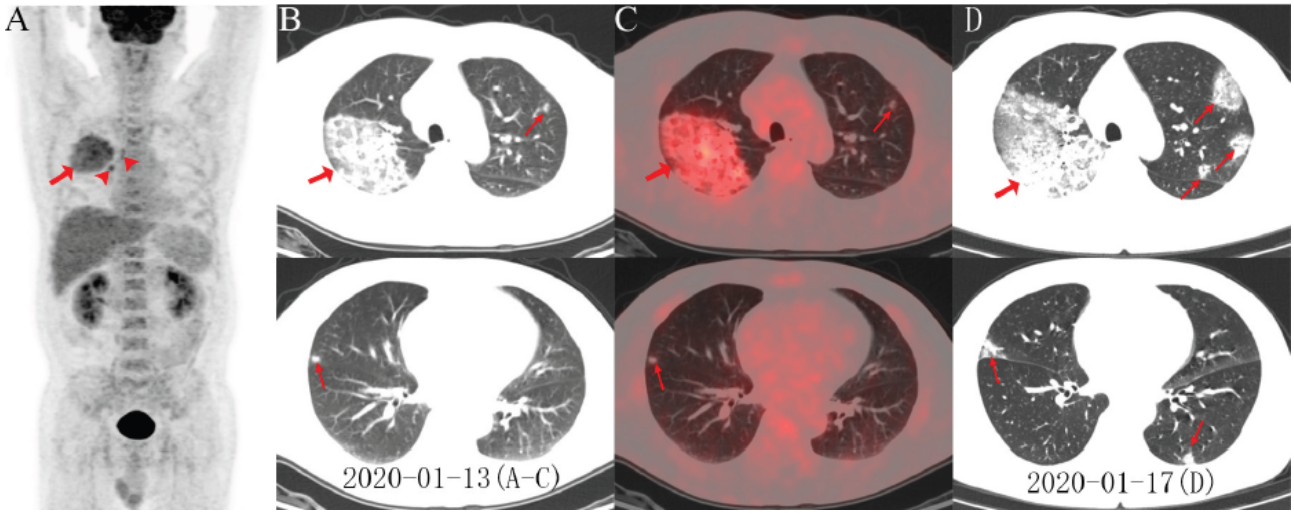
Zou and Zhu [49] also reported the case of a COVID-19 patient who was scanned using ^{18}F-FDG-PET/CT. They observed a positive FDG uptake (SUV max 4.9) in the right lung and increased accumulation of FDG in the right hilar and right paratracheal lymph nodes (Figure 4). In this case study, accumulation of FDG was noted in the bone marrow. In another case study reported by Czernin et al [50] of a 53-year-old patient with a neuroendocrine pancreatic tumor who was referred for staging and scanned with ^{18}F-FDG-PET/CT, they found positive uptake (SUV max 5.5) in a new hypermetabolic area in the right lower and upper lobes. This tracer uptake correlated topographically to predominantly peripheral and sub-pleural GGOs with beginning, partly round-shaped consolidations. The patient was asymptomatic when the PET scan was performed; however, COVID-19 infection was later confirmed [50]. These

observations suggest the diagnostic value of nuclear medicine imaging in the early stages of COVID-19, especially when clinical symptoms are unspecific in patients referred for other clinical concerns. As the progression and severity of the disease can harm other organs, such as the kidneys, bone marrow, heart and gastrointestinal tract, ^{18}F-FDG-PET/CT scans can provide a whole-body noninvasive assessment to detect damage in chronical and concomitant organs [47].

As an imaging modality, FDG PET/CT offers added value in diagnostic complications caused by COVID-19, observation of disease progression, and treatment responses [46,47]. While PET/CT cannot be routinely used in an emergency setting for COVID-19, this imaging modality could play a complementary diagnostic role in disease management [47]. ^{18}F-FDG PET/CT may hold special diagnostic value in estimating the extent of organ involvement during the course of COVID-19 and in monitoring treatment efficiency. It also may help predict the recovery time of patients with COVID-19 [47].

Based on available evidence, the advantage of using ^{18}F-FDG PET/CT for patients with COVID-19 is its sensitivity in detecting, diagnosing, and monitoring pathophysiological changes in inflamed and infected lung lesions [51]. However, recent published studies of FDG PET/CT for COVID-19 have been limited to small sample sizes and case reports [48,49]. Further investigations are required to identify the SUV cutoffs for different lung lesions in various stages of the disease to define the potential diagnostic accuracy and limitations of PET/CT scan in detecting lung lesions for COVID-19 pneumonia. Although radionuclide pulmonary ventilation and perfusion may play roles in the diagnosis of pulmonary embolism in patients with COVID-19, these modalities are not recommended for use in clinical practice during the pandemic due to the increased risk of infection transmission [52].

Figure 4. Fluorodeoxyglucose positron emission topography/computed tomography (FDG PET/CT) imaging findings in a patient with coronavirus disease. (A) The PET maximum intensity projection image shows an FDG-avid mass in the right lung with a maximum standardized uptake value of 4.9, as well as increased accumulation of FDG in the right hilar lymph nodes, in the right paratracheal stripe (arrowhead), and in the bone marrow. The axial images of the low-dose CT scan (B) and the PET/CT fusion (C) show ground-glass opacities in the right upper lobe with areas of focal consolidation (arrows) and focal opacities in the right middle and left upper lobes (arrows). Follow-up CT axial images obtained 4 days later (D) show lesion progression in the middle and bilateral upper lobes, with newly developed focal opacities in the left lower and upper lobes (arrows) [49].



Ultrasound

Commonly used for early diagnosis of pneumonia, lung ultrasound is an alternative to chest x-ray or CT scan [53,54]. The diagnostic accuracy of lung ultrasound has been found to be significantly better than that of chest x-ray in identifying the most common lung pathologies, such as consolidation, pneumothorax, pleural effusion, and interstitial syndrome [55]. The most common ultrasonography features of patients with confirmed COVID-19 include thickened pleural lines with irregularities, B-lines in various patterns, small peripheral consolidations, absence of pleural effusions, and appearance of A-lines during recovery [56]. Lung ultrasonography has specific features for alveolar-interstitial lung disease, including viral pneumonia and adult respiratory distress syndrome (ARDS), which distinguish it from bacterial pneumonia. These features are represented by the B-lines, small subpleural consolidation, and irregular pleural line [57]. In a case report of a young man infected with COVID-19 [59], his lung ultrasound results clearly showed signs suggestive of interstitial-alveolar damage, including areas of white lung, irregular pleural line with subpleural consolidations, and thick irregular vertical artifacts (B-lines) (Figure 5).

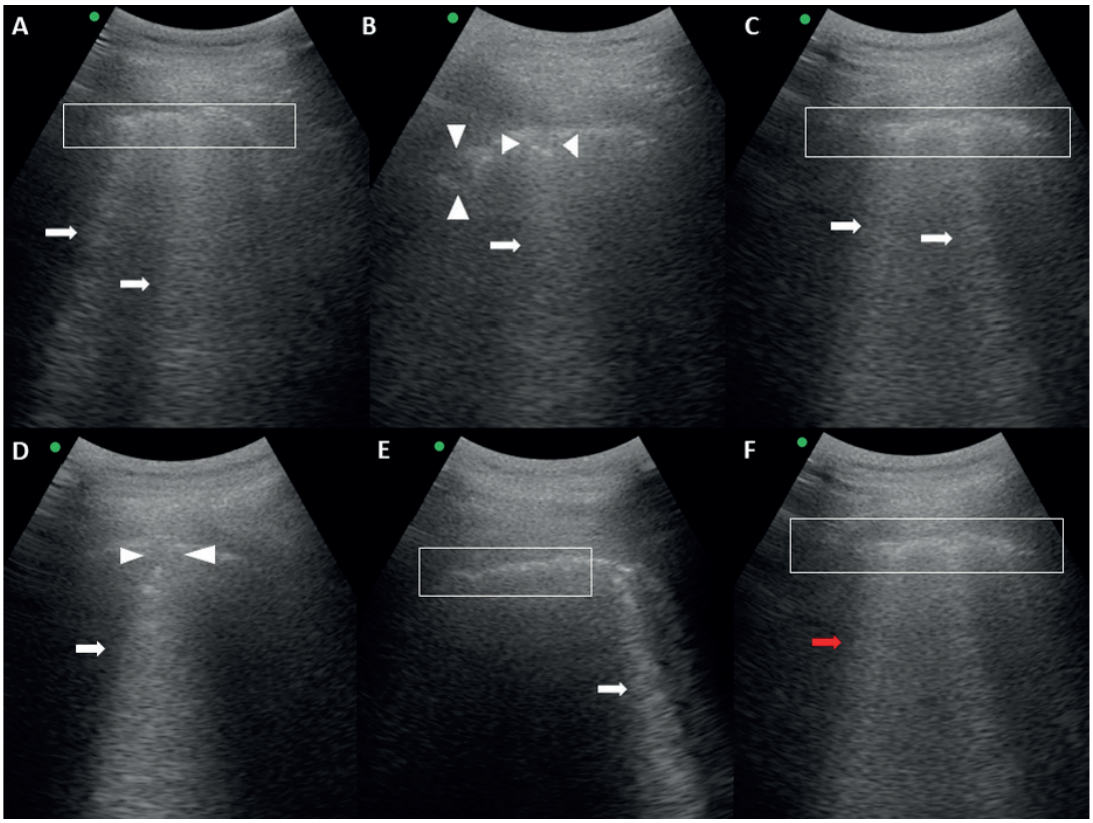
The advantages of lung ultrasound for detecting lung lesions in patients with COVID-19 are that it gives similar results to chest CT and it is superior to standard chest radiography [43,56]. Lomoro et al [43] performed lung ultrasound, chest radiography, and CT scans of confirmed COVID-19 patients and found an association between lung ultrasound features and CT findings for GGOs and consolidation. This could aid the rapid diagnosis and management of COVID-19 pneumonia and its progression toward ARDS [43,57]. A notable limitation of lung ultrasound is that aerated lungs may block transmission of ultrasonography, preventing the detection of deep lesions within the lung [57,58]. When pneumonia does not extend to the pleural surface, a CT scan is required to identify disease progression [57].

Taking chest CT as the gold standard, lung ultrasound showed relatively high diagnostic consistency value and diagnostic coincidence rate in asymptomatic patients with COVID-19 [60]. The positive predictive value and negative predictive value were 100% and 85.71%, respectively [60]. Thus, lung ultrasound can be used as a screening method for lung involvement at different stages of COVID-19 pneumonia [57,58]. In the early stage of the disease, the focal B-line was seen in the lung ultrasound, whereas in more severe and progressive stages, thickening of the plural line and irregular B-lines were found in patients with fibrosis and alveolar interstitial syndrome [57]. Lu et al [65] investigated the clinical value of ultrasound in the diagnosis of lung lesions at different stages of COVID-19. They found that the diagnostic accuracy (93.3 %), sensitivity (100.0 %), and specificity (85.7 %) of bedside ultrasound are high for severe lung lesions in COVID-19 patients; however, they are relatively low for mild (sensitivity 68.8%, specificity 85%, diagnostic accuracy 76.7%) and moderate (sensitivity 77.8%, specificity 76.2%, diagnostic accuracy 76.7%) lung lesions [65]. This study suggests that the use of nonionizing and dynamic ultrasound examination should be further considered in critically ill patients with suspected or documented COVID-19 infection to detect the severity of lung lesions [65]. The additional advantage of lung ultrasound is that it can be used in the emergency department or in the intensive care unit for scanning COVID-19 patients due to its portability, safety, absence of radiation, ease of use, repeatability, and low cost [59]. However, strict protection and operating procedures during the examination should be employed to minimize infection risk [65].

During the COVID-19 pandemic, lung ultrasound was rapidly established in Italy as a diagnostic tool in patients with suspected COVID-19 infection [61]. Their application of lung ultrasound scores offers added value for determining severity of lung involvement in infected patients. Vetrugno et al [61] also noted a reduction in the use of CT scans and chest x-ray along with improved management of patients. Thus, lung ultrasound may be helpful when planning a suitable diagnostic workup, depending on the patient's clinical condition and available technological resources.

Lung ultrasound could also play a crucial role in the diagnosis and monitoring of pregnant women with COVID-19 [62]. To our knowledge, there is limited data on pregnant women with COVID-19 infection. In a case study of 9 pregnant women with COVID-19, chest CT scans were used and showed multiple patchy ground-glass shadows in the lungs, similar to those of nonpregnant women with COVID-19 [39]. Liu et al [40] reported similar chest CT imaging features in pregnant women with COVID-19 but with more severe consolidations. In their study, a low-dose technique was used for CT scanning of pregnant women; this technique was sufficient to detect lesions but reduced the image quality [40]. While chest CT was the first-choice modality for early detection and assessment of disease severity, special attention was required in follow-up diagnosis and management of pregnant women with COVID-19. As lung ultrasound can reveal certain specific signs in patients with respiratory involvement of COVID-19 [65], it can be used in the diagnosis of pregnant women with COVID-19 [62].

Figure 5. Lung ultrasound findings in a patient with coronavirus disease. Irregular plural lines (A, C, E, and F, within the white boxes); thick irregular vertical artifacts (A, B, C, D, and E, white arrows); subpleural consolidations (B and D, white arrowheads); and areas of white lung (F, red arrow) [59].



Conclusion

In this literature review, the value and roles of different imaging modalities for the diagnosis and management of COVID-19 were discussed. Chest x-ray and CT scan are thoracic imaging techniques with key diagnostic value in suspected cases of COVID-19. While PET/CT and ultrasound may not be routinely used in diagnosing COVID-19, these modalities could play complementary roles and add value in managing disease progression. Ultimately, early and accurate diagnosis of patients infected with COVID-19 can effectively control disease progression.

Acknowledgments

The authors thank Dr Adnan Al-Homaidan, President of the University of Jeddah, and the Vice-President for Graduate Studies and Scientific Research, Dr Obaid Al-Modaf, for their support and encouragement of research related to COVID-19. Also, the authors would like to thank the College of Applied Medical Sciences and the department of Applied Radiologic Technology in the University of Jeddah for their full support. This work received no funding.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of the selected articles in the literature based on the different imaging modalities and their diagnostic value for COVID-19.
[PNG File , 41 KB - jmir_v22i8e19673_app1.png]

References

1. He F, Deng Y, Li W. Coronavirus disease 2019: What we know? J Med Virol 2020 Jul;92(7):719-725 [FREE Full text] [doi: 10.1002/jmv.25766] [Medline: 32170865]
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
3. Gorbalenya A, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. Severe acute respiratory syndrome-related coronavirus – The species and its viruses, a statement of the Coronavirus Study Group. bioRxiv 2020 Feb 11:preprint. [doi: 10.1101/2020.02.07.937862]
4. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020 Apr 2;5(4):536-544 [FREE Full text] [doi: 10.1038/s41564-020-0695-z] [Medline: 32123347]
5. Yan Y, Shin WI, Pang YX, Meng Y, Lai J, You C, et al. The First 75 Days of Novel Coronavirus (SARS-CoV-2) Outbreak: Recent Advances, Prevention, and Treatment. Int J Environ Res Public Health 2020 Mar 30;17(7):2323 [FREE Full text] [doi: 10.3390/ijerph17072323] [Medline: 32235575]
6. Lai C, Shih T, Ko W, Tang H, Hsueh P. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020 Mar;55(3):105924 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105924] [Medline: 32081636]
7. Kanne J. Chest CT Findings in 2019 Novel Coronavirus (2019-nCoV) Infections from Wuhan, China: Key Points for the Radiologist. Radiology 2020 Apr;295(1):16-17 [FREE Full text] [doi: 10.1148/radiol.2020200241] [Medline: 32017662]
8. Zhou S, Wang Y, Zhu T, Xia L. CT Features of Coronavirus Disease 2019 (COVID-19) Pneumonia in 62 Patients in Wuhan, China. American Journal of Roentgenology 2020 Jun;214(6):1287-1294. [doi: 10.2214/ajr.20.22975]
9. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
10. Zhang G, Zhang J, Wang B, Zhu X, Wang Q, Qiu S. Analysis of clinical characteristics and laboratory findings of 95 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a retrospective analysis. Respir Res 2020 Mar 26;21(1):74 [FREE Full text] [doi: 10.1186/s12931-020-01338-8] [Medline: 32216803]
11. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
12. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
13. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. Radiology 2020 Aug;296(2):E115-E117 [FREE Full text] [doi: 10.1148/radiol.2020200432] [Medline: 32073353]

14. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology 2020 Aug;296(2):E32-E40 [FREE Full text] [doi: 10.1148/radiol.2020200642] [Medline: 32101510]
15. Hao W, Li M. Clinical diagnostic value of CT imaging in COVID-19 with multiple negative RT-PCR testing. Travel Med Infect Dis 2020 Mar;34:101627 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101627] [Medline: 32179123]
16. Araujo-Filho J, Sawamura MVY, Costa AN, Cerri GG, Nomura CH. COVID-19 pneumonia: what is the role of imaging in diagnosis? J Bras Pneumol 2020 Mar 27;46(2):e20200114 [FREE Full text] [doi: 10.36416/1806-3756/e20200114] [Medline: 32236303]
17. Yang W, Sirajuddin A, Zhang X, Liu G, Teng Z, Zhao S, et al. The role of imaging in 2019 novel coronavirus pneumonia (COVID-19). Eur Radiol 2020 Apr 15:online [FREE Full text] [doi: 10.1007/s00330-020-06827-4] [Medline: 32296940]
18. Wong H, Lam H, Fong A, Leung S, Chin T, Lo C, et al. Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19. Radiology 2020 Aug;296(2):E72-E78 [FREE Full text] [doi: 10.1148/radiol.2020201160] [Medline: 32216717]
19. Ng M, Lee EY, Yang J, Yang F, Li X, Wang H, et al. Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review. Radiology: Cardiothoracic Imaging 2020 Feb 01;2(1):e200034. [doi: 10.1148/ryct.2020200034]
20. Bernheim A, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, et al. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. Radiology 2020 Jun;295(3):200463 [FREE Full text] [doi: 10.1148/radiol.2020200463] [Medline: 32077789]
21. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, et al. Time Course of Lung Changes at Chest CT during Recovery from Coronavirus Disease 2019 (COVID-19). Radiology 2020 Jun;295(3):715-721. [doi: 10.1148/radiol.2020200370]
22. Salehi S, Abedi A, Balakrishnan S, Gholamrezanezhad A. Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients. AJR Am J Roentgenol 2020 Jul;215(1):87-93. [doi: 10.2214/ajr.20.23034]
23. Li B, Li X, Wang Y, Han Y, Wang Y, Wang C, et al. Diagnostic value and key features of computed tomography in Coronavirus Disease 2019. Emerg Microbes Infect 2020 Dec 23;9(1):787-793 [FREE Full text] [doi: 10.1080/22221751.2020.1750307] [Medline: 32241244]
24. Kooraki S, Hosseiny M, Myers L, Gholamrezanezhad A. Coronavirus (COVID-19) Outbreak: What the Department of Radiology Should Know. J Am Coll Radiol 2020 Apr;17(4):447-451 [FREE Full text] [doi: 10.1016/j.jacr.2020.02.008] [Medline: 32092296]
25. Pan Y, Guan H, Zhou S, Wang Y, Li Q, Zhu T, et al. Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan, China. Eur Radiol 2020 Jun 13;30(6):3306-3309 [FREE Full text] [doi: 10.1007/s00330-020-06731-x] [Medline: 32055945]
26. Zhao W, Zhong Z, Xie X, Yu Q, Liu J. CT Scans of Patients with 2019 Novel Coronavirus (COVID-19) Pneumonia. Theranostics 2020;10(10):4606-4613 [FREE Full text] [doi: 10.7150/thno.45016] [Medline: 32292517]
27. Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J. Chest CT for Typical Coronavirus Disease 2019 (COVID-19) Pneumonia: Relationship to Negative RT-PCR Testing. Radiology 2020 Aug;296(2):E41-E45 [FREE Full text] [doi: 10.1148/radiol.2020200343] [Medline: 32049601]
28. Yoon S, Lee K, Kim J, Lee Y, Ko H, Kim K, et al. Chest Radiographic and CT Findings of the 2019 Novel Coronavirus Disease (COVID-19): Analysis of Nine Patients Treated in Korea. Korean J Radiol 2020 Apr;21(4):494-500 [FREE Full text] [doi: 10.3348/kjr.2020.0132] [Medline: 32100485]
29. Wang K, Kang S, Tian R, Zhang X, Zhang X, Wang Y. Imaging manifestations and diagnostic value of chest CT of coronavirus disease 2019 (COVID-19) in the Xiaogan area. Clin Radiol 2020 May;75(5):341-347 [FREE Full text] [doi: 10.1016/j.crad.2020.03.004] [Medline: 32216961]
30. Zhao W, Zhong Z, Xie X, Yu Q, Liu J. Relation Between Chest CT Findings and Clinical Conditions of Coronavirus Disease (COVID-19) Pneumonia: A Multicenter Study. AJR Am J Roentgenol 2020 May;214(5):1072-1077. [doi: 10.2214/ajr.20.22976]
31. Zhou Z, Guo D, Li C, Fang Z, Chen L, Yang R, et al. Coronavirus disease 2019: initial chest CT findings. Eur Radiol 2020 Aug;30(8):4398-4406 [FREE Full text] [doi: 10.1007/s00330-020-06816-7] [Medline: 32211963]
32. Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). Radiology 2020 Apr;295(1):202-207 [FREE Full text] [doi: 10.1148/radiol.2020200230] [Medline: 32017661]
33. Koo HJ, Lim S, Choe J, Choi S, Sung H, Do K. Radiographic and CT Features of Viral Pneumonia. Radiographics 2018 May;38(3):719-739. [doi: 10.1148/rg.2018170048] [Medline: 29757717]
34. Jin Y, Cai L, Cheng Z, Cheng H, Deng T, Fan YP, for the Zhongnan Hospital of Wuhan University Novel Coronavirus Management and Research Team, Evidence-Based Medicine Chapter of China International Exchange and Promotive Association for Medical and Health Care (CPAM). A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). Mil Med Res 2020 Feb 06;7(1):4 [FREE Full text] [doi: 10.1186/s40779-020-0233-6] [Medline: 32029004]
35. Long C, Xu H, Shen Q, Zhang X, Fan B, Wang C, et al. Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? Eur J Radiol 2020 May;126:108961 [FREE Full text] [doi: 10.1016/j.ejrad.2020.108961] [Medline: 32229322]

36. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. Lancet Infect Dis 2020 Apr;20(4):425-434. [doi: 10.1016/s1473-3099(20)30086-4]
37. Raptis C, Hammer M, Short R, Shah A, Bhalla S, Bierhals A, et al. Chest CT and Coronavirus Disease (COVID-19): A Critical Review of the Literature to Date. AJR Am J Roentgenol 2020 Apr 16:1-4. [doi: 10.2214/AJR.20.23202] [Medline: 32298149]
38. Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raoof S, et al. The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. Radiology 2020 Jul;296(1):172-180 [FREE Full text] [doi: 10.1148/radiol.2020201365] [Medline: 32255413]
39. Lei D, Li C, Fang C, Yang W, Cheng B, Wei M. Clinical characteristics of pregnancy with the 2019 novel coronavirus disease (COVID-19) infection. Article in Chinese. Chinese Journal of Perinatal Medicine 2020;23(3):online.
40. Liu H, Liu F, Li J, Zhang T, Wang D, Lan W. Clinical and CT imaging features of the COVID-19 pneumonia: Focus on pregnant women and children. J Infect 2020 May;80(5):e7-e13 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.007] [Medline: 32171865]
41. ACR Recommendations for the use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection. American College of Radiology. 2020 Mar 22. URL: https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection [accessed 2020-08-04]
42. Song F, Shi N, Shan F, Zhang Z, Shen J, Lu H, et al. Emerging 2019 Novel Coronavirus (2019-nCoV) Pneumonia. Radiology 2020 Apr;295(1):210-217 [FREE Full text] [doi: 10.1148/radiol.2020200274] [Medline: 32027573]
43. Lomoro P, Verde F, Zerboni F, Simonetti I, Borghi C, Fachinetti C, et al. COVID-19 pneumonia manifestations at the admission on chest ultrasound, radiographs, and CT: single-center study and comprehensive radiologic literature review. Eur J Radiol Open 2020;7:100231 [FREE Full text] [doi: 10.1016/j.ejro.2020.100231] [Medline: 32289051]
44. Jacobi A, Chung M, Bernheim A, Eber C. Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review. Clin Imaging 2020 Aug;64:35-42 [FREE Full text] [doi: 10.1016/j.clinimag.2020.04.001] [Medline: 32302927]
45. Zu Z, Jiang M, Xu P, Chen W, Ni Q, Lu G, et al. Coronavirus Disease 2019 (COVID-19): A Perspective from China. Radiology 2020 Aug;296(2):E15-E25 [FREE Full text] [doi: 10.1148/radiol.2020200490] [Medline: 32083985]
46. Capitanio S, Nordin AJ, Noraini AR, Rossetti C. PET/CT in nononcological lung diseases: current applications and future perspectives. Eur Respir Rev 2016 Sep 31;25(141):247-258 [FREE Full text] [doi: 10.1183/16000617.0051-2016] [Medline: 27581824]
47. Lütje S, Marinova M, Kütting D, Attenberger U, Essler M, Bundschuh RA. Nuclear medicine in SARS-CoV-2 pandemia: 18F-FDG-PET/CT to visualize COVID-19. Nuklearmedizin 2020 Jun;59(3):276-280 [FREE Full text] [doi: 10.1055/a-1152-2341] [Medline: 32259853]
48. Qin C, Liu F, Yen T, Lan X. F-FDG PET/CT findings of COVID-19: a series of four highly suspected cases. Eur J Nucl Med Mol Imaging 2020 May 22;47(5):1281-1286 [FREE Full text] [doi: 10.1007/s00259-020-04734-w] [Medline: 32088847]
49. Zou S, Zhu X. FDG PET/CT of COVID-19. Radiology 2020 Aug;296(2):E118-E118 [FREE Full text] [doi: 10.1148/radiol.2020200770] [Medline: 32142399]
50. Czernin J, Fanti S, Meyer P, Allen-Auerbach M, Hacker M, Sathekge M, et al. Nuclear Medicine Operations in the Times of COVID-19: Strategies, Precautions, and Experiences. J Nucl Med 2020 May;61(5):626-629. [doi: 10.2967/jnumed.120.245738] [Medline: 32238430]
51. Deng Y, Lei L, Chen Y, Zhang W. The potential added value of FDG PET/CT for COVID-19 pneumonia. Eur J Nucl Med Mol Imaging 2020 Jul 21;47(7):1634-1635 [FREE Full text] [doi: 10.1007/s00259-020-04767-1] [Medline: 32198615]
52. Ayan A, Kıraç FS. Guide for Nuclear Medicine Applications During the COVID-19 Outbreak. Mol Imaging Radionucl Ther 2020 Apr 29;29(2):49-58 [FREE Full text] [doi: 10.4274/mirt.galenos.2020.33600] [Medline: 32368875]
53. Laursen CB, Sloth E, Lassen AT, Christensen RD, Lambrechtsen J, Madsen PH, et al. Point-of-care ultrasonography in patients admitted with respiratory symptoms: a single-blind, randomised controlled trial. Lancet Respir Med 2014 Aug;2(8):638-646. [doi: 10.1016/s2213-2600(14)70135-3]
54. Zhou B, Bartholmai BJ, Kalra S, Osborn TG, Zhang X. Lung US Surface Wave Elastography in Interstitial Lung Disease Staging. Radiology 2019 May;291(2):479-484 [FREE Full text] [doi: 10.1148/radiol.2019181729] [Medline: 30835191]
55. Mojoli F, Bouhemad B, Mongodi S, Lichtenstein D. Lung Ultrasound for Critically Ill Patients. Am J Respir Crit Care Med 2019 Mar 15;199(6):701-714. [doi: 10.1164/rccm.201802-0236ci]
56. Convissar D, Gibson L, Berra L, Bittner E, Chang M. Application of Lung Ultrasound During the COVID-19 Pandemic: A Narrative Review. Anesth Analg 2020 Aug;131(2):345-350 [FREE Full text] [doi: 10.1213/ANE.0000000000004929] [Medline: 32366774]
57. Peng Q, Wang X, Zhang L, Chinese Critical Care Ultrasound Study Group (CCUSG). Findings of lung ultrasonography of novel corona virus pneumonia during the 2019-2020 epidemic. Intensive Care Med 2020 May 12;46(5):849-850 [FREE Full text] [doi: 10.1007/s00134-020-05996-6] [Medline: 32166346]
58. Sofia S, Boccatonda A, Montanari M, Spampinato M, D'ardes D, Cocco G, et al. Thoracic ultrasound and SARS-COVID-19: a pictorial essay. J Ultrasound 2020 Jun 16;23(2):217-221 [FREE Full text] [doi: 10.1007/s40477-020-00458-7] [Medline: 32297175]

59. Buonsenso D, Piano A, Raffaelli F, Bonadia N, de Gaetano Donati K, Franceschi F. Point-of-Care Lung Ultrasound findings in novel coronavirus disease-19 pnemoniae: a case report and potential applications during COVID-19 outbreak. Eur Rev Med Pharmacol Sci 2020 Mar;24(5):2776-2780 [FREE Full text] [doi: 10.26355/eurrev_202003_20549] [Medline: 32196627]
60. Lin H, Zhang B, Kou H, Zhao Y, Li K, Wu D. Application value of lung ultrasound in asymptomatic patients with confirmed COVID-19. AUDT 2020;4(2):67-72 [FREE Full text]
61. Vetrugno L, Bove T, Orso D, Barbariol F, Bassi F, Boero E, et al. Our Italian experience using lung ultrasound for identification, grading and serial follow-up of severity of lung involvement for management of patients with COVID-19. Echocardiography 2020 Apr 15;37(4):625-627 [FREE Full text] [doi: 10.1111/echo.14664] [Medline: 32239532]
62. Moro F, Buonsenso D, Moruzzi MC, Inchingolo R, Smargiassi A, Demi L, et al. How to perform lung ultrasound in pregnant women with suspected COVID-19. Ultrasound Obstet Gynecol 2020 May;55(5):593-598. [doi: 10.1002/uog.22028] [Medline: 32207208]
63. Li X, Zeng W, Li X, Chen H, Shi L, Li X, et al. CT imaging changes of corona virus disease 2019 (COVID-19): a multi-center study in Southwest China. J Transl Med 2020 Apr 06;18(1):154 [FREE Full text] [doi: 10.1186/s12967-020-02324-w] [Medline: 32252784]
64. Li Y, Xia L. Coronavirus Disease 2019 (COVID-19): Role of Chest CT in Diagnosis and Management. AJR Am J Roentgenol 2020 Jun;214(6):1280-1286. [doi: 10.2214/AJR.20.22954] [Medline: 32130038]
65. Lu W, Zhang S, Chen B, Chen J, Xian J, Lin Y, et al. A Clinical Study of Noninvasive Assessment of Lung Lesions in Patients with Coronavirus Disease-19 (COVID-19) by Bedside Ultrasound. Ultraschall Med 2020 Jun 15;41(3):300-307. [doi: 10.1055/a-1154-8795] [Medline: 32294796]

Abbreviations

ACR: American College of Radiology
ARDS: adult respiratory distress syndrome
COVID-19: coronavirus disease
CT: computed tomography
FDG: fluorodeoxyglucose
GGO: ground-glass opacity
MERS: Middle East respiratory syndrome
MERS-CoV: Middle East respiratory syndrome coronavirus
NCP: novel coronavirus pneumonia
PET/CT: positron emission topography/computed tomography
RT-PCR: reverse transcription–polymerase chain reaction
SARS: severe acute respiratory syndrome
SARS-CoV: severe acute respiratory syndrome coronavirus
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
SUV max: maximum standardized uptake value
WHO: World Health Organization

Edited by G Eysenbach; submitted 27.04.20; peer-reviewed by A Tajaldeen, I Abdelaziz, P Lei; comments to author 15.06.20; revised version received 28.06.20; accepted 21.07.20; published 19.08.20.

Please cite as:
Aljondi R, Alghamdi S
Diagnostic Value of Imaging Modalities for COVID-19: Scoping Review
J Med Internet Res 2020;22(8):e19673
URL: http://www.jmir.org/2020/8/e19673/
doi:10.2196/19673
PMID:32716893

Original Paper

Impact of Wearing Masks, Hand Hygiene, and Social Distancing on Influenza, Enterovirus, and All-Cause Pneumonia During the Coronavirus Pandemic: Retrospective National Epidemiological Surveillance Study

Nan-Chang Chiu[1,2*], MD; Hsin Chi[1,2*], MD; Yu-Lin Tai[3], MD; Chun-Chih Peng[1,2], MD; Cheng-Yin Tseng[3], MD; Chung-Chu Chen[3,4], MD; Boon Fatt Tan[5], MD; Chien-Yu Lin[2,3], MD

[1]MacKay Children's Hospital, Taipei, Taiwan

[2]Department of Medicine, MacKay Medical College, New Taipei, Taiwan

[3]Hsinchu MacKay Memorial Hospital, Hsinchu, Taiwan

[4]Teaching Center of Natural Science, Minghsin University of Science and Technology, Hsinchu, Taiwan

[5]National Taiwan University, Hsinchu Branch, Hsinchu, Taiwan

[*]these authors contributed equally

Corresponding Author:
Chien-Yu Lin, MD
Hsinchu MacKay Memorial Hospital
690 Guanfu Road
East District
Hsinchu, 300
Taiwan
Phone: 886 36119595 ext 2582
Email: mmhped.lin@gmail.com

Abstract

Background: The coronavirus disease (COVID-19) pandemic is an important health crisis worldwide. Several strategies were implemented to combat COVID-19, including wearing masks, hand hygiene, and social distancing. The impact of these strategies on COVID-19 and other viral infections remains largely unclear.

Objective: We aim to investigate the impact of implemented infectious control strategies on the incidences of influenza, enterovirus infection, and all-cause pneumonia during the COVID-19 pandemic.

Methods: We utilized the electronic database of the Taiwan National Infectious Disease Statistics System and extracted incidences of COVID-19, influenza virus, enterovirus, and all-cause pneumonia. We compared the incidences of these diseases from week 45 of 2016 to week 21 of 2020 and performed linear regression analyses.

Results: The first case of COVID-19 in Taiwan was reported in late January 2020 (week 4). Infectious control strategies have been promoted since late January. The influenza virus usually peaks in winter and decreases around week 14. However, a significant decrease in influenza was observed after week 6 of 2020. Regression analyses produced the following results: 2017, R^2=0.037; 2018, R^2=0.021; 2019, R^2=0.046; and 2020, R^2=0.599. A dramatic decrease in all-cause pneumonia was also reported (R^2 values for 2017-2020 were 0.435, 0.098, 0.352, and 0.82, respectively). Enterovirus had increased by week 18 in 2017-2019, but this was not observed in 2020.

Conclusions: Using this national epidemiological database, we found a significant decrease in cases of influenza, enterovirus, and all-cause pneumonia during the COVID-19 pandemic. Wearing masks, hand hygiene, and social distancing may contribute not only to the prevention of COVID-19 but also to the decline of other respiratory infectious diseases. Further studies are warranted to elucidate the causal relationship.

(J Med Internet Res 2020;22(8):e21257) doi:10.2196/21257

KEYWORDS

novel coronavirus; COVID-19; SARS-CoV-2; pandemic; influenza; pneumonia; hygiene; social distancing; prevention; incidence; surveillance

Introduction

Coronavirus disease (COVID-19), which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has engendered a substantial health burden worldwide, although the full impact of this virus remains largely unknown [1,2]. As of May 2020, Taiwan had succeeded in containing COVID-19 without a lockdown, recording just 441 confirmed cases (19.2 cases per 1 million residents) [3]. Many strategies have been implemented since mid-January, such as boundary control, use of masks, hand hygiene, and social distancing [4,5]. Advances in technology also contributed to the control of this novel pandemic, including big data analysis, proactive tests, and a real-time, web-based dashboard to track COVID-19 [4,6]. Big data analytics with smart contact tracing and automated alert messaging for self-restriction were used to effectively contain infected patients [7]. Novel technologies contributed to Taiwan's COVID-19 response.

Although the full pathophysiology of COVID-19 was not known, droplet and contract transmission were believed to be the major transmission route [2]. Wearing a mask was a simple way to prevent viral transmission and decrease disease spread, but the public attitude toward masks varied across countries [8]. The World Health Organization's recommendation of mask use also varied from time to time [9]. A recent systematic review and meta-analysis showed a significantly lower risk of viral transmission by maintaining a physical distance of 1 meter or more (pooled adjusted odds ratio [aOR] 0.18) [10]. Mask and eye protection use also resulted in a large reduction in the risk of infection (mask use: aOR 0.15; eye protection: aOR 0.22). Additionally, the use of masks by all residents was a key component to successfully combat COVID-19 and may have reduced fear and anxiety [5,11,12]. Briefly, although these customary strategies may have marginal benefits based on current evidence, experts recommend their use during the COVID-19 pandemic [13]. However, the effectiveness of these strategies on other respiratory infections apart from SARS-CoV-2 remains largely unclear. Jefferson et al [14] investigated the effectiveness of physical interventions to reduce the spread of respiratory viruses in their 2011 study and found that wearing masks and hand hygiene were effective against viral transmission; social distancing was not. Barasheed et al [15] explored uptake and effectiveness of masks during mass gatherings in 2016; they found a pooled protective effectiveness with a relative risk of 0.89, but an extremely wide range in the uptake of masks was reported (0.02%-92.8%) [15]. The effectiveness of these traditional strategies and public compliance was not fully disclosed.

The Taiwan government executed a name-based mask rationing plan since late January, and mask factories were recruited. Masks were re-allocated to the general public to ensure availability for all citizens. Mask use and medical care were believed to be key strategies for successful control in Taiwan [11]. Hand hygiene was also promoted in late January, and a significant increase of Google searches for "washing hands" was observed since January 19, 2020 [16]. Citizens of Taiwan were highly motivated to curb the pandemic, and this led to successful outcomes. High uptake and compliance of these traditional practices of infection control, including wearing masks, hand hygiene, and social distancing, were observed in Taiwan. Furthermore, we also observed a noticeable decrease in the number of cases of influenza infection during the COVID-19 pandemic. Taiwan is located in the northern hemisphere where influenza infection is usually prevalent starting in October and peaks in February. We hypothesized that it may be affected by the strategies in place for COVID-19 control and prevention. Therefore, we conducted this retrospective study to investigate the prevalence of other respiratory viral infections using the national surveillance database.

Methods

Study Design and Database

Our study was approved by the ethical committee of MacKay Memorial Hospital (No. 20MMHIS140e). Taiwan Centers for Disease Control (CDC) had a comprehensive surveillance system and epidemiological data regarding communicable diseases, influenza virus, enterovirus, and pneumonia that were available on its website [17]. The Taiwan National Infectious Disease Statistics System is a public and nationwide database that provides real-time epidemiologic information to health care personnel.

All suspected cases of COVID-19 received nucleic acid testing following the standard procedure by World Health Organization. All confirmed cases had to be quarantined in hospital and might be discharged after three consecutive negative tests. A diagnosis of influenza virus, enterovirus, and pneumonia were made by physicians based on clinical manifestations, physical examinations, laboratory tests, and imaging studies. The diagnoses were uploaded to the national health insurance system and surveillance system using International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes [18].

We extracted epidemiological data related to the influenza virus, enterovirus, all-cause pneumonia, and COVID-19. We then compared the weekly cases from October to May (weeks 45 to 21 of next year) for the 2017-2020 period. We compared the incidences of these diseases during the same period and plotted the trendlines. Furthermore, policies and strategies were obtained from the CDC website to demonstrate the time sequences.

Statistics

The weekly incidences of reported cases were plotted using Microsoft Office, version 2019 (Microsoft Corp), and SPSS, version 23.0 (IBM Corp). Linear regression analyses were performed and R^2 values were calculated for each year. The equation of linear trend estimation was presented as $y=\alpha x+ß$.

A positive α coefficient denoted an increase, and a negative α value indicated a decrease. The value of α reflected the slope of the trendline and the magnitude of effects. R^2, also known as the coefficient of determination, represented the degree of dispersion between individual data and the regression line. The R^2 value is always between 0% and 100%, and the higher the R^2 value, the lower the discrepancies between data. An R^2 value close to 1 represents a reliable fitted regression line.

Results

In Taiwan, a unique name-based mask rationing plan was executed, and hand hygiene has been promoted since January 2020 [3]. Social distancing policies recommended a distance of at least 1 meter and 1.5 meters from others in outdoor and indoor settings, respectively, in week 14 of 2020 [3]. We extracted epidemiological data of target diseases from the Taiwan National Infectious Disease Statistics System. As of week 21 of 2020, there were 441 confirmed cases of COVID-19; Taiwan had a relatively controllable situation [2,3]. The incidences of influenza, enterovirus, and pneumonia between week 45 of 2016 and week 21 of 2020 are shown in Figure 1. Seasonality of each disease was observed and a significant decrease in all diseases since week 6 of 2020 were found. The age distribution from week 45 to week 21 of the next year is summarized in Table 1. Fewer patients with influenza, enterovirus, and pneumonia were reported in 2020. Among those with influenza and pneumonia, patients aged 15-24 years had the lowest rates, and approximately one third of patients were younger than 15 years. For patients with enterovirus infection, the majority were younger than 10 years. Compared with 2019, less than half of patients had enterovirus infection in 2020 (106,985 vs 220,865).

Figure 1. Weekly incidences of influenza, enterovirus, and pneumonia in 2017-2020. Dotted line indicates a significant decrease in all three diseases after week 6 of 2020.

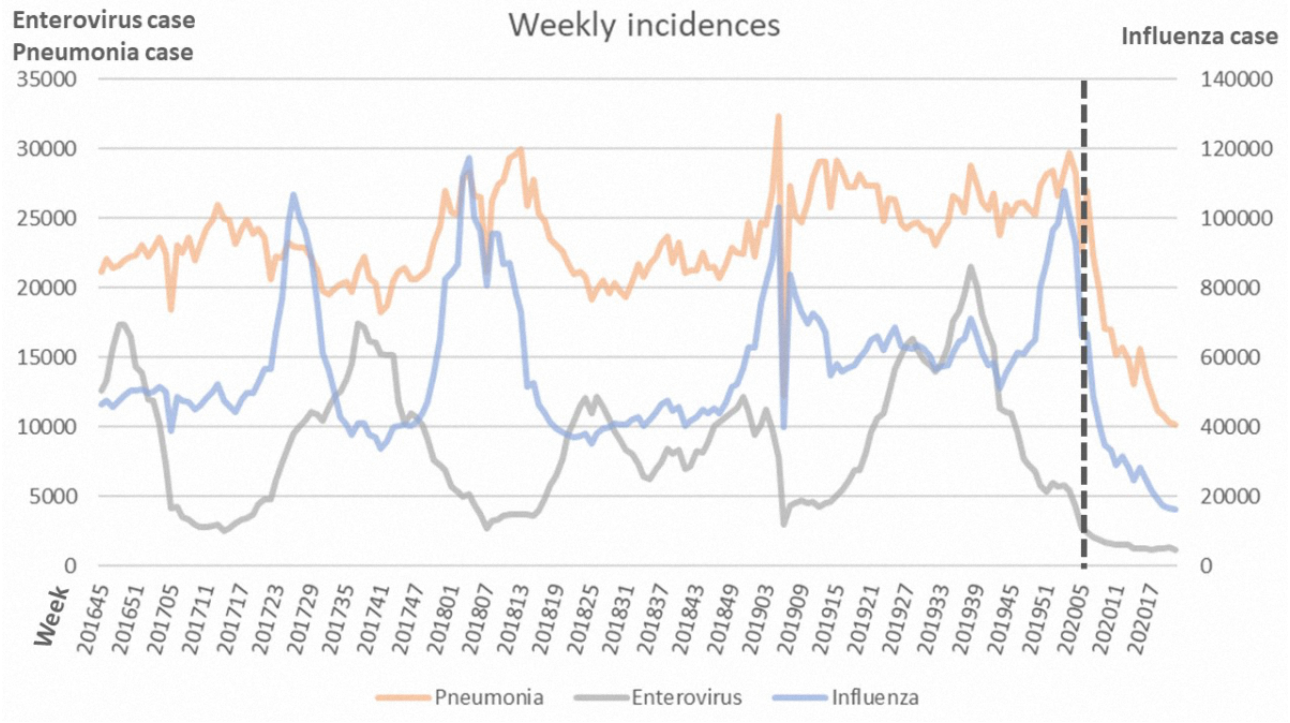


Table 1. Age distribution of people with respiratory viral infection during the study period (week 45 to week 21 of the next year).

Disease	Year			
	2017	2018	2019	2020
Influenza, n	1,405,539	1,960,252	1,838,406	1,518,787
0-4 years, n (%)	222,203 (15.81)	267,395 (13.64)	269,477 (14.66)	205,911 (13.56)
5-14 years, n (%)	263,076 (18.72)	418,752 (21.36)	377,639 (20.54)	311,008 (20.48)
15-24 years, n (%)	114,929 (8.18)	182,663 (9.32)	161,034 (8.76)	129,450 (8.52)
25-64 years, n (%)	567,302 (40.36)	806,152 (41.12)	759,040 (41.29)	643,141 (42.35)
≥65 years, n (%)	238,029 (16.94)	285,290 (14.55)	271,216 (14.75)	229,277 (15.1)
Enterovirus, n	218,969	171,401	220,865	106,985
0-2 years, n (%)	69,812 (31.88)	47,332 (27.61)	58,277 (26.39)	30,712 (28.71)
3-4 years, n (%)	61,824 (28.23)	45,190 (26.37)	64,160 (29.05)	28,397 (26.54)
5-9 years, n (%)	60,562 (27.66)	51,266 (29.91)	71,069 (32.18)	30,079 (28.12)
10-14 years, n (%)	11,312 (5.17)	11,913 (6.95)	12,450 (5.64)	5908 (5.52)
≥15 years, n (%)	15,459 (7.06)	15,700 (9.16)	14,909 (6.75)	11,889 (11.11)
Pneumonia, n	668,070	725,014	730,414	593,292
0-4 years, n (%)	163,835 (24.52)	160,827 (22.18)	161,670 (22.13)	117,323 (19.77)
5-14 years, n (%)	142,120 (21.27)	148,408 (20.47)	151,033 (20.68)	115,580 (19.48)
15-24 years, n (%)	28,368 (4.25)	31,242 (4.31)	29,487 (4.04)	25,201 (4.25)
25-64 years, n (%)	198,359 (29.69)	229,681 (31.68)	233,855 (32.02)	203,163 (34.24)
≥65 years, n (%)	135,388 (20.27)	154,857 (21.36)	154,369 (21.13)	132,025 (22.25)

We further plotted the reported cases of influenza from week 45 to week 21 of the next year in Figure 2. There was no significant variation between weeks in 2017, but influenza increased rapidly around week 50 of 2018 and decreased around week 14 of 2019. In 2020, the COVID-19 pandemic began in week 4, during which a dramatic decrease in influenza was observed. The R^2 values for 2017-2020 were 0.037, 0.021, 0.046, and 0.599, respectively (Table 2). This dramatic decrease was more significant in all-cause pneumonia from week 6 onward (Figure 3). Enterovirus infection was common in Taiwan and usually increased by week 16 (Figure 4), although no such increase occurred in 2020.

Figure 2. Incidences of influenza in 2017-2020. COVID-19: coronavirus disease.

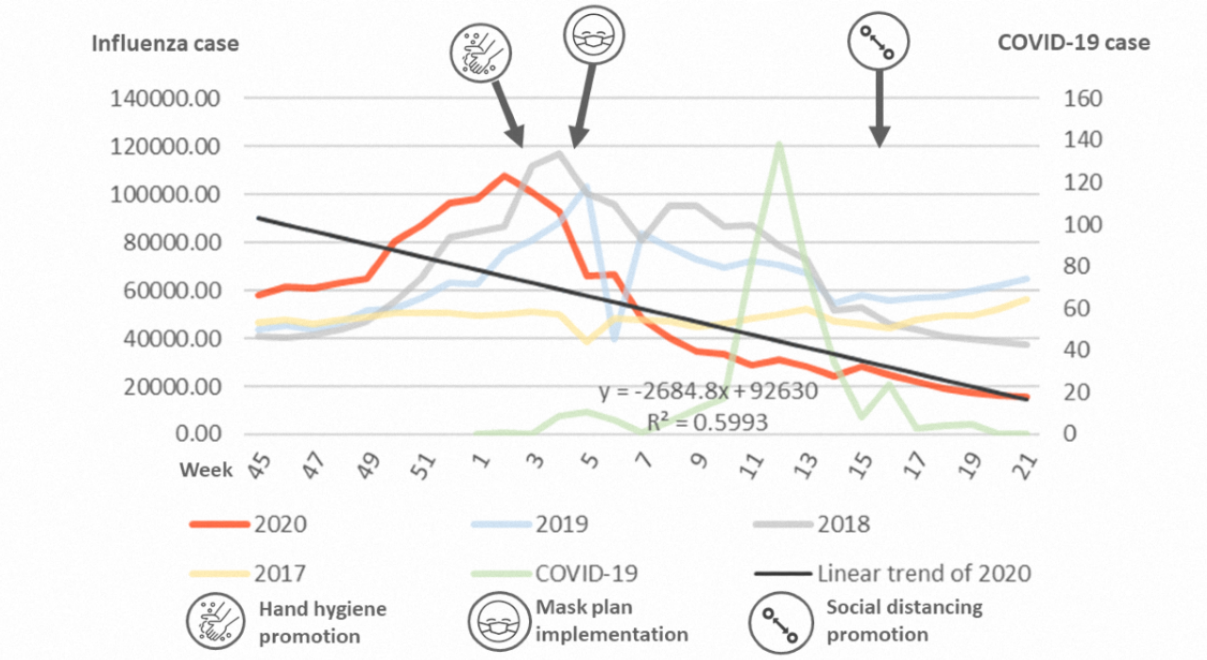


Table 2. Regression analyses results (R^2) for 2017-2020.

Disease	Year			
	2017	2018	2019	2020
Influenza	0.037	0.021	0.046	0.599
Enterovirus	0.72	0.256	0.359	0.834
All-cause pneumonia	0.435	0.098	0.352	0.82

Figure 3. Incidences of all-cause pneumonia in 2017-2020. A dramatic decrease in pneumonia from week 6 of 2020 is shown by the dotted line. COVID-19: coronavirus disease.

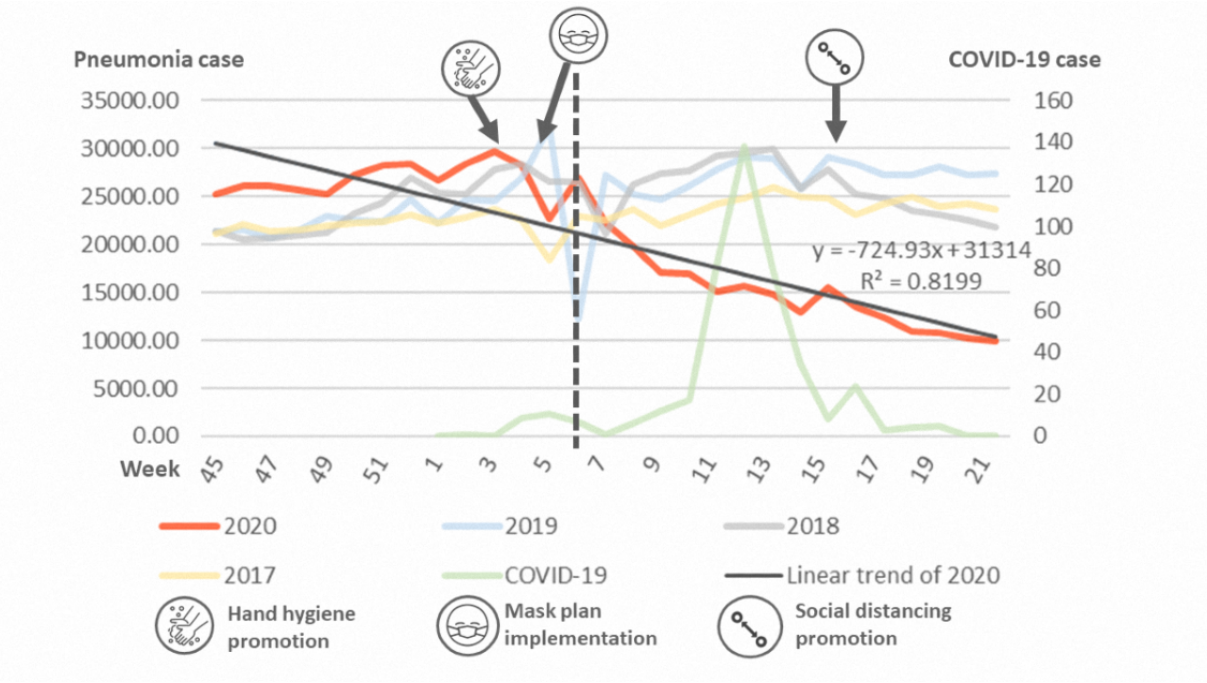
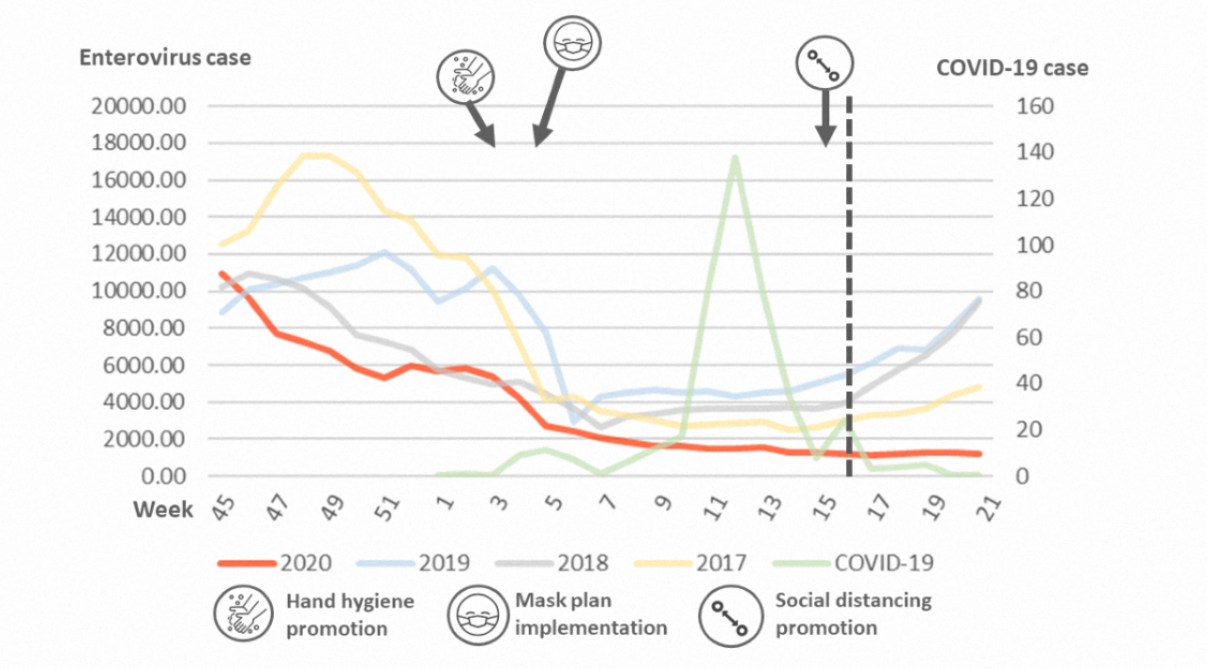


Figure 4. Incidences of enterovirus in 2017-2020. Dotted line marks the increase that is usually seen in enterovirus infection cases in week 16, although this was not observed in 2020. COVID-19: coronavirus disease.



Discussion

Utilizing the national electronic epidemiologic database, we found a concomitant decrease in influenza, enterovirus, and all-cause pneumonia during the COVID-19 pandemic. Infectious control measures, including wearing masks, hand hygiene, and social distancing, may contribute not only to the prevention of COVID-19 but also to decreases in the incidence of other viral infections and pneumonia.

Advances in technology also contributed to COVID-19 response. Several novel techniques had been applied to control COVID-19, including an interactive web-based dashboard, big data analysis, mobile technology, and social media platforms. Internet-based digital citizen science is the crucial component for tackling pandemics in the 21th century [19-23]. Our study utilized electronic epidemiologic statistics and provided timely preliminary findings. A combination of novel technology and traditional infectious control measurements played crucial roles to fight the pandemic. Controversies regarding the effectiveness of traditional strategies exist and the uptake of these strategies varied across countries. Although these strategies are straightforward for reducing respiratory viral transmission, convincing evidence supporting their effectiveness is lacking. Most studies were observational studies or simulation models, and a strong recommendation was not achieved [24]. For COVID-19, asymptomatic patients may spread disease so universal masking in communities was recommended in some areas [8,12,13,25]. However, the effectiveness of masks is doubted; different fitted filtration efficiencies were observed in different masks [26]. Furthermore, wearing masks, especially N95 respirators, is uncomfortable, which can contribute to noncompliance. Generally speaking, uptake of masks was widely accepted in Asian countries, and high compliance of mask use was observed during the pandemic [8,27]. Additionally, a shortage of masks was an important issue, and Taiwan implemented a name-based mask distribution system and rationing plan to ensure availability of masks for purchase [5,28]. Although a direct comparison with randomized controlled trials was not available, our study found a relatively controllable COVID-19 situation in Taiwan. As of the 21st week of 2020, there were 441 confirmed cases of COVID-19 in Taiwan. These traditional strategies were effective for prevention and control.

We found these strategies were effective not only in reducing COVID-19 but also other respiratory viruses. However, it is difficult to investigate the independent and combined effects of each strategy. Although wearing masks, hand hygiene, and social distancing are straightforward, it is difficult, and may be unethical, to conduct a randomized controlled trial to compare their protective effects during the pandemic. In Taiwan, policies for wearing masks and hand hygiene were implemented in late January and social distancing was promoted in week 14. The observed decrease began in week 6, and success may be mainly attributed to wearing a mask and hand hygiene. Moreover, previous studies found social distancing had not significantly reduced transmission of influenza and other viruses [14,29,30]. SARS-CoV-2 is highly contagious; the effects of these practices may vary in areas with different viruses, societies, cultures, health care resources, population densities, disease prevalence, and proportion of subclinical carriers. It is a challenge to determine the effectiveness of these practices on outcomes since the results are drawn from descriptive analyses. There were many factors such as comorbidity, health care access during COVID-19, age, sex, time of the year, mask wearing behaviors (eg, all the time, sometimes), hand washing frequencies, etc, that may affect incidence rates. Nevertheless, the entire causal relationship between infectious control measurement and viral transmission was not easily clarified. Further studies are warranted to investigate the independent and combined effects of these practices.

Our study demonstrated a significant decrease in respiratory viral diseases after the implementation of these practices using linear trend estimation. Time-series analysis was also a useful tool to predict the trendline based on previous epidemiological data [31]. However, COVID-19 is a novel situation; thus, a precise prospective prediction may be not feasible at present. Additionally, seasonality is an important factor for infectious diseases, and the onset of the pandemic occurred approximately 6 months ago. The impact of seasonality on COVID-19 remains unclear. Therefore, we decided to compare the incidences of the same period (week 45 to week 21 of the next year) in different years and plot the trendlines using linear regression analyses [32,33], which is more meaningful and intuitive. However, the COVID-19 situation changes rapidly and the optimal strategies to combat COVID-19 have also changed rapidly. Timely and continuous surveillance and international cooperation are crucial for successful epidemiological studies.

Since the onset of the COVID-19 pandemic, aggressive infection-control measures were implemented, and a decrease in the occurrence of other infectious diseases was observed. The reason for the observed decrease may be multifactorial, with other factors such as virus competition contributing as well [34-36]. Although coinfection with 2 or more viruses were not uncommon, competition for resources between viruses was observed [35,37]. Competition of different subtypes of the influenza A virus was reported in previous study [36]. The observed decreases of other respiratory infections may be the result of competition by SARS-CoV-2. However, there were relatively few cases of COVID-19 in Taiwan. We believe that mask use, hand hygiene, and social distancing controlled the spread of COVID-19 as well as influenza, enterovirus, and all-cause pneumonia. Further studies are warranted to clarify the causal relationship and elucidate the complex interactions between people, viruses, and the environment.

The strength of our study was the use of a national, real-time database on a large population. Our study has some limitations. First, our study was a retrospective epidemiological study. Therefore, the underpinning mechanisms and causal relationships cannot be established; further studies are required for this. Second, the people of Taiwan were strongly motivated to control and prevent infection spread; hence, the independent effects of every single strategy are not easy to confirm. Although hand hygiene and mask use were implemented in the early phase of the pandemic response and seemed to be responsible for the successful decrease in viral transmission, more data on the efficacy of individual strategies is required. Finally, health care resources, accessibility of network, availability of masks, and

attitudes toward mask use varied across different societies and countries. The impact of COVID-19 was also different in different areas. Further studies investigating the prevalence of other respiratory viruses in different countries contribute to further understanding the entire impact of COVID-19 and infectious control measurements.

In conclusion, our nationwide epidemiologic study found a significant decrease in influenza, enterovirus, and all-cause pneumonia during the COVID-19 pandemic. Wearing a mask, hand hygiene, and social distancing not only reduced the impact of COVID-19; these strategies also led to a decline in other respiratory infections. Further studies are warranted to clarify this causal relationship.

Authors' Contributions

N-CC, HC, and C-YL contributed to the study concept and design. N-CC, Y-LT, C-CP, C-YT, C-CC, and BFT contributed to data acquisition. N-CC, HC, and C-YL contributed to the statistical analysis. N-CC and HC contributed to this work equally. C-YL drafted the manuscript. All authors gave final approval for the manuscript.

Conflicts of Interest

None declared.

References

1. Easom N, Moss P, Barlow G, Samson A, Taynton T, Adams K, et al. Sixty-eight consecutive patients assessed for COVID-19 infection: Experience from a UK Regional infectious diseases Unit. Influenza Other Respir Viruses 2020 Jul 08;14(4):374-379 [FREE Full text] [doi: 10.1111/irv.12739] [Medline: 32223012]
2. Coronavirus disease (COVID-19) outbreak situation. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-04-01]
3. Coronavirus disease 2019 (COVID-19). Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov.tw/En [accessed 2020-08-12]
4. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA 2020 Mar 03. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
5. Chang C, Tan T, Ho T, Chen C, Su T, Lin C. COVID-19: Taiwan's epidemiological characteristics and public and hospital responses. PeerJ 2020;8:e9360 [FREE Full text] [doi: 10.7717/peerj.9360] [Medline: 32551205]
6. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May 19;20(5):533-534. [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
7. Chen C, Jyan H, Chien S, Jen H, Hsu C, Lee P, et al. Containing COVID-19 Among 627,386 Persons in Contact With the Diamond Princess Cruise Ship Passengers Who Disembarked in Taiwan: Big Data Analytics. J Med Internet Res 2020 May 05;22(5):e19540 [FREE Full text] [doi: 10.2196/19540] [Medline: 32353827]
8. Feng S, Shen C, Xia N, Song W, Fan M, Cowling BJ. Rational use of face masks in the COVID-19 pandemic. The Lancet Respiratory Medicine 2020 May;8(5):434-436. [doi: 10.1016/s2213-2600(20)30134-x]
9. Advice on the use of masks in the community, during home care and in health care settings in the context of the novel coronavirus (2019-nCoV) outbreak: interim guidance. World Health Organization. 2020. URL: https://apps.who.int/iris/handle/10665/330987 [accessed 2020-05-20]
10. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 2020 Jun 27;395(10242):1973-1987 [FREE Full text] [doi: 10.1016/S0140-6736(20)31142-9] [Medline: 32497510]
11. Su VYF, Yen Y, Yang K, Su W, Chou K, Chen Y, et al. Masks and medical care: Two keys to Taiwan's success in preventing COVID-19 spread. Travel Med Infect Dis 2020 Jun 04:101780 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101780] [Medline: 32505623]
12. Klompas M, Morris CA, Sinclair J, Pearson M, Shenoy ES. Universal Masking in Hospitals in the Covid-19 Era. N Engl J Med 2020 May 21;382(21):e63. [doi: 10.1056/NEJMp2006372] [Medline: 32237672]
13. Greenhalgh T, Schmid MB, Czypionka T, Bassler D, Gruer L. Face masks for the public during the covid-19 crisis. BMJ 2020 Apr 09;369:m1435. [doi: 10.1136/bmj.m1435] [Medline: 32273267]
14. Jefferson T, Del Mar CB, Dooley L, Ferroni E, Al-Ansary LA, Bawazeer GA, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses. Cochrane Database Syst Rev 2011 Jul 06(7):CD006207 [FREE Full text] [doi: 10.1002/14651858.CD006207.pub4] [Medline: 21735402]
15. Barasheed O, Alfelali M, Mushta S, Bokhary H, Alshehri J, Attar A, et al. Uptake and effectiveness of facemask against respiratory infections at mass gatherings: a systematic review. Int J Infect Dis 2016 Jun;47:105-111 [FREE Full text] [doi: 10.1016/j.ijid.2016.03.023] [Medline: 27044522]

16. Lin Y, Liu C, Chiu Y. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun 2020 Jul;87:30-32 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.020] [Medline: 32283286]
17. Taiwan National Infectious Disease Statistics System. Centers for Disease Control. 2020. URL: https://nidss.cdc.gov.tw/en/ [accessed 2020-05-28]
18. Lin C, Liu J, Wu C, Hsu R, Hsu W. Decreased Risk of Renal Calculi in Patients Receiving Androgen Deprivation Therapy for Prostate Cancer. Int J Environ Res Public Health 2020 Mar 09;17(5):e [FREE Full text] [doi: 10.3390/ijerph17051762] [Medline: 32182733]
19. Ahmad AR, Murad HR. The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
20. Geldsetzer P. Use of Rapid Online Surveys to Assess People's Perceptions During Infectious Disease Outbreaks: A Cross-sectional Survey on COVID-19. J Med Internet Res 2020 Apr 02;22(4):e18790 [FREE Full text] [doi: 10.2196/18790] [Medline: 32240094]
21. Huang Y, Wu Q, Wang P, Xu Y, Wang L, Zhao Y, et al. Measures Undertaken in China to Avoid COVID-19 Infection: Internet-Based, Cross-Sectional Survey Study. J Med Internet Res 2020 May 12;22(5):e18718 [FREE Full text] [doi: 10.2196/18718] [Medline: 32396516]
22. Katapally TR. A Global Digital Citizen Science Policy to Tackle Pandemics Like COVID-19. J Med Internet Res 2020 May 26;22(5):e19357 [FREE Full text] [doi: 10.2196/19357] [Medline: 32408267]
23. Wang P, Lu W, Ko N, Chen Y, Li D, Chang Y, et al. COVID-19-Related Information Sources and the Relationship With Confidence in People Coping with COVID-19: Facebook Survey Study in Taiwan. J Med Internet Res 2020 Jun 05;22(6):e20021 [FREE Full text] [doi: 10.2196/20021] [Medline: 32490839]
24. Nussbaumer-Streit B, Mayr V, Dobrescu AI, Chapman A, Persad E, Klerings I, et al. Quarantine alone or in combination with other public health measures to control COVID-19: a rapid review. Cochrane Database Syst Rev 2020 Apr 08;4:CD013574. [doi: 10.1002/14651858.CD013574] [Medline: 32267544]
25. Cheng VC, Wong S, Chuang VW, So SY, Chen JH, Sridhar S, et al. The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. J Infect 2020 Jul;81(1):107-114 [FREE Full text] [doi: 10.1016/j.jinf.2020.04.024] [Medline: 32335167]
26. Sickbert-Bennett EE, Samet JM, Clapp PW, Chen H, Berntsen J, Zeman KL, et al. Filtration Efficiency of Hospital Face Mask Alternatives Available for Use During the COVID-19 Pandemic. JAMA Intern Med 2020 Aug 11;173(1):W22-W23 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4221] [Medline: 32780113]
27. Universal mask-wearing is the most overlooked COVID-19 lifesaver. maskssavelives.org. 2020. URL: https://reurl.cc/8GdRLj [accessed 2020-07-10]
28. Wu HL, Huang J, Zhang CJP, He Z, Ming WK. Facemask shortage and the novel coronavirus disease (COVID-19) outbreak: Reflections on public health measures. EClinicalMedicine 2020 Apr 03;8(4):100329-100328 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100329] [Medline: 32292898]
29. Ahmed F, Zviedrite N, Uzicanin A. Effectiveness of workplace social distancing measures in reducing influenza transmission: a systematic review. BMC Public Health 2018 Apr 18;18(1):518 [FREE Full text] [doi: 10.1186/s12889-018-5446-1] [Medline: 29669545]
30. Yu D, Lin Q, Chiu AP, He D. Effects of reactive social distancing on the 1918 influenza pandemic. PLoS One 2017;12(7):e0180545 [FREE Full text] [doi: 10.1371/journal.pone.0180545] [Medline: 28704460]
31. Dayer MJ, Jones S, Prendergast B, Baddour LM, Lockhart PB, Thornhill MH. Incidence of infective endocarditis in England, 2000–13: a secular trend, interrupted time-series analysis. The Lancet 2015 Mar 28;385(9974):1219-1228 [FREE Full text] [doi: 10.1016/S0140-6736(14)62007-9] [Medline: 25467569]
32. Marill KA. Advanced statistics: linear regression, part I: simple linear regression. Acad Emerg Med 2004 Jan;11(1):87-93 [FREE Full text] [Medline: 14709436]
33. Kraemer MUG, Yang C, Gutierrez B, Wu C, Klein B, Pigott DM, Open COVID-19 Data Working Group, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 2020 May 01;368(6490):493-497 [FREE Full text] [doi: 10.1126/science.abb4218] [Medline: 32213647]
34. Pinky L, Dobrovolny HM. Coinfections of the Respiratory Tract: Viral Competition for Resources. PLoS One 2016;11(5):e0155589 [FREE Full text] [doi: 10.1371/journal.pone.0155589] [Medline: 27196110]
35. Trinh JT, Zeng L. Virus interactions: cooperation or competition? Future Microbiol 2017 Jun;12:561-564. [doi: 10.2217/fmb-2017-0048] [Medline: 28604103]
36. Latorre-Margalef N, Brown JD, Fojtik A, Poulson RL, Carter D, Franca M, et al. Competition between influenza A virus subtypes through heterosubtypic immunity modulates re-infection and antibody dynamics in the mallard duck. PLoS Pathog 2017 Jun 22;13(6):e1006419 [FREE Full text] [doi: 10.1371/journal.ppat.1006419] [Medline: 28640898]
37. Lin C, Hwang D, Chiu N, Weng L, Liu H, Mu J, et al. Increased Detection of Viruses in Children with Respiratory Tract Infection Using PCR. Int J Environ Res Public Health 2020 Jan 15;17(2):e [FREE Full text] [doi: 10.3390/ijerph17020564] [Medline: 31952364]

Abbreviations

aOR: adjusted odds ratio
CDC: Taiwan Centers for Disease Control
COVID-19: coronavirus disease
ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 13.06.20; peer-reviewed by TR Katapally, CF Yen; comments to author 13.07.20; revised version received 29.07.20; accepted 01.08.20; published 20.08.20.

Please cite as:
Chiu NC, Chi H, Tai YL, Peng CC, Tseng CY, Chen CC, Tan BF, Lin CY
Impact of Wearing Masks, Hand Hygiene, and Social Distancing on Influenza, Enterovirus, and All-Cause Pneumonia During the Coronavirus Pandemic: Retrospective National Epidemiological Surveillance Study
J Med Internet Res 2020;22(8):e21257
URL: http://www.jmir.org/2020/8/e21257/
doi:10.2196/21257
PMID:32750008

Original Paper

Digital Inequality During a Pandemic: Quantitative Study of Differences in COVID-19–Related Internet Uses and Outcomes Among the General Population

Alexander JAM van Deursen[1], Prof Dr

University of Twente, Enschede, Netherlands

Corresponding Author:
Alexander JAM van Deursen, Prof Dr
University of Twente
Drienerlolaan 5
Enschede, 7500AE
Netherlands
Phone: 31 622942142
Fax: 31 534893200
Email: a.j.a.m.vandeursen@utwente.nl

Abstract

Background: The World Health Organization considers coronavirus disease (COVID-19) to be a public emergency threatening global health. During the crisis, the public's need for web-based information and communication is a subject of focus. Digital inequality research has shown that internet access is not evenly distributed among the general population.

Objective: The aim of this study was to provide a timely understanding of how different people use the internet to meet their information and communication needs and the outcomes they gain from their internet use in relation to the COVID-19 pandemic. We also sought to reveal the extent to which gender, age, personality, health, literacy, education, economic and social resources, internet attitude, material access, internet access, and internet skills remain important factors in obtaining internet outcomes after people engage in the corresponding uses.

Methods: We used a web-based survey to draw upon a sample collected in the Netherlands. We obtained a dataset with 1733 respondents older than 18 years.

Results: Men are more likely to engage in COVID-19–related communication uses. Age is positively related to COVID-19–related information uses and negatively related to information and communication outcomes. Agreeableness is negatively related to both outcomes and to information uses. Neuroticism is positively related to both uses and to communication outcomes. Conscientiousness is not related to any of the uses or outcomes. Introversion is negatively related to communication outcomes. Finally, openness relates positively to all information uses and to both outcomes. Physical health has negative relationships with both outcomes. Health perception contributes positively to information uses and both outcomes. Traditional literacy has a positive relationship with information uses and both outcomes. Education has a positive relationship with information and communication uses. Economic and social resources played no roles. Internet attitude is positively related to information uses and outcomes but negatively related to communication uses and outcomes. Material access and internet access contributed to all uses and outcomes. Finally, several of the indicators and outcomes became insignificant after accounting for engagement in internet uses.

Conclusions: Digital inequality is a major concern among national and international scholars and policy makers. This contribution aimed to provide a broader understanding in the case of a major health pandemic by using the ongoing COVID-19 crisis as a context for empirical work. Several groups of people were identified as vulnerable, such as older people, less educated people, and people with physical health problems, low literacy levels, or low levels of internet skills. Generally, people who are already relatively advantaged are more likely to use the information and communication opportunities provided by the internet to their benefit in a health pandemic, while less advantaged individuals are less likely to benefit. Therefore, the COVID-19 crisis is also enforcing existing inequalities.

(J Med Internet Res 2020;22(8):e20073) doi:10.2196/20073

KEYWORDS

COVID-19; digital inequality; internet use; survey; personality; literacy; internet skills; information; communication

Introduction

Background

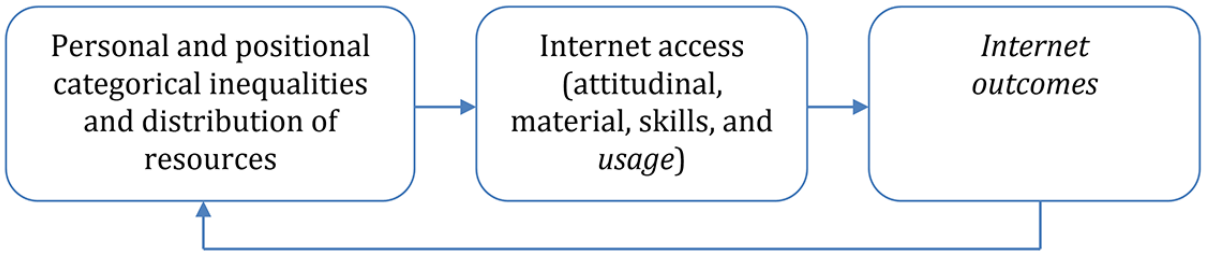
The World Health Organization considers coronavirus disease (COVID-19) to be a public emergency threatening global health [1]. Governments worldwide have taken stringent action, including requiring social distancing, closing public services, schools and universities, and canceling cultural events [2,3]. People are being advised or ordered to stay at home and socially isolate themselves to avoid being infected [4]. The ongoing pandemic represents an outbreak of an unparalleled scale, and it has induced widespread fear and uncertainty.

In this paper, we focus on the role of the internet during the crisis. The internet has become a crucial source for the general public, as it provides access to general information, the latest national and international developments, and guidelines on behavioral norms during the crisis. In this respect, the internet plays an important role in the great challenges facing governments regarding the transfer of knowledge and guidelines to the population at large. When individuals understand the need and rationale behind government-enforced measures, they are more motivated to comply and even adopt measures voluntarily [5,6]. In addition to informational purposes, the internet enables individuals to share news and experiences with people they cannot meet face-to-face, remain in contact with friends and family, seek support, and ask questions of official agencies, including health agencies. Further, the internet enables people to take initiatives such as raising money or preparing packaged meals for people in need, such as health workers or people who have lost their jobs. In sum, the internet plays a vital role for people of all social strata and backgrounds during a time of worldwide crisis. All people should thus be able to use the internet as a source of information and communication.

However, digital inequality research has shown that internet access is not evenly distributed among the general population [7,8]. The basic idea of digital inequality stems from a comparative perspective of social and information inequality, as there are benefits associated with internet access and negative consequences of lack of access [9]. Calamities are often a story of inequality [10]; therefore, in this paper, we aimed to gain a deeper and broader understanding of the differences in how people use the internet to cope during the COVID-19 crisis. Van Dijk's resources and appropriation theory [8] explains differences or inequalities of internet access by considering personal and positional categories of individuals and the individuals' resources. Internet access itself is considered to be a process of appropriation involving attitudinal access, material access, skills access, and in the final stage, usage access. The latter entails differences in the type of activities that people perform on the internet. The consequences of the process are the outcomes of internet use. These outcomes in turn reinforce personal and positional inequalities and an unequal distribution of resources [8] (Figure 1). The first goal of this paper is to provide a timely understanding of how different people use the internet and the outcomes they gain from it in relation to the COVID-19 pandemic.

Internet use and outcome differences between groups of people are likely to have profound consequences on how people manage a crisis. For example, older people are most in danger of being infected with the virus and most likely to die from the infection [11], and they also use the internet less and have the fewest internet outcomes [12]. The latter may further endanger their peculiar situation, as limited internet use and outcomes may result in a lack of critical information or necessary support.

Figure 1. Simplified model of the resources and appropriation theory [8].



COVID-19–Related Internet Uses and Outcomes

To study differences in internet uses and outcomes during the COVID-19 pandemic, it is necessary to understand the types of uses and outcomes that are at play. Typically, uses and outcomes are studied by following conceptual classifications that distinguish different domains, such as economic, social, cultural, or personal domains [13]. Here, we take the COVID-19 pandemic as the domain of interest. Within this domain, we consider two main and conceptually different types of uses and outcomes: information and communication [14,15]. Information internet uses involve searching for information on all aspects of COVID-19. Potential information outcomes include becoming better informed about the disease, understanding why certain measures are necessary, and limiting the risk of becoming infected by developing greater awareness of one's own behavior. Communication internet uses include talking to friends about the crisis, asking questions on social media or online fora, giving advice, or offering support to others. Communication outcomes include finding people on the internet who can offer support or share concern, being less lonely, and protecting others from potential COVID-19 risks. Studying both types of uses and outcomes is important, as prior research has shown that communication uses can compensate for information uses to attain beneficial internet outcomes [16].

Determinants of COVID-19–Related Internet Uses and Outcomes

Digital inequality research suggests that the vast amount of web-based information and communication possibilities around the COVID-19 pandemic are likely to be difficult to grasp and conceptualize for sections of the general population [7]. Some frequently observed personal categorical inequalities are gender, age, personality, and health [7]. Earlier research revealed that men and women differ in their internet activities; women are more likely to use email and social media, whereas men are more likely to use the internet to obtain information [17,18]. Age in general has a negative impact on all types of internet uses and outcomes [7]. In the COVID-19 crisis, older people are especially vulnerable; therefore, it is very important for them to know how to behave and be safe. We hypothesize that (H1) men are more likely to be involved in information-related uses and outcomes while women are more likely to be involved in communication-related internet uses and outcomes regarding COVID-19-related internet uses and outcomes. We also hypothesize that (H2) age contributes negatively to COVID-19–related internet uses and outcomes.

An individual's personality may hinder or stimulate their engagement in certain COVID-19–related activities. Cognitive appraisal theory suggests that individuals complete two types of cognitive appraisal processes in a crisis [19]. The process starts with an evaluation of the crisis as a potential source of danger or life disruption. If the crisis is not determined to be dangerous, it is not considered a stressor and does not require intervention. If the crisis is determined to be relevant, it is considered a stressor and must be evaluated further by balancing the demands of the crisis and the person's resources [20]. At this point, personality enters the equation [20]. There is a general consensus regarding the Big Five model when personality traits are studied. This model proposes five personality traits of agreeableness, neuroticism, conscientiousness, introversion, and openness [21]. However, there is no agreement as to whether these traits contribute to or detract from resisting disturbance [20]. There is also no consensus on how the Big Five personality traits relate to internet use [7,22]. For example, conscientiousness relates to people who abide by rules. On one hand, one might argue that this would result in a greater need for information on how to behave. On the other hand, the internet is unstructured, and rules and policies are absent to a large extent. When linking personality traits to internet use for psychological adjustments to the COVID-19 crisis, it is not evident whether these traits will support or hinder COVID-19–related internet uses and outcomes. We hypothesize that (H3a) agreeableness, (H3b) neuroticism, (H3c) contentiousness, (H3d) introversion, and (H3e) openness are related to COVID-19–related internet uses and outcomes.

An individual's health may play an important role in how they approach COVID-19. To gain an elaborate understanding of how health relates to COVID-19–related internet uses and outcomes, we followed earlier research that distinguishes between different health aspects [23]: A person's physical functioning or the degree to which their health currently interferes with activities such as sports, carrying groceries, climbing stairs, and walking, their mental health or psychological distress and well-being, and their health perception concerning their own health rating in general. During a crisis, we expect that people with health issues are more likely to turn to the internet for comfort and reassurance. We hypothesize that (H4a) physical functioning, (H4b) mental health, and (H4c) health perception contribute negatively to COVID-19–related internet uses and outcomes.

The final type of personal inequality considered in this study is traditional literacy, which is known to have a substantial impact on how the internet is used [24,25]. We consider literacy to be the ability to read, write, and understand text, which is also framed under the umbrella terms functional literacy or fundamental literacy [24]. Functional or traditional literacy can be considered as the basic dimension of all literacy concepts [26]. Considering the crucial role the internet is playing in the COVID-19 crisis, a low level of literacy is a potentially large inhibitor of understanding information and being involved in web-based communication. We hypothesize that (H5) traditional literacy contributes positively to COVID-19–related internet uses and outcomes.

Education is the most observed positional categorial inequality in digital divide research, and it is likely to play a role in the current context. People with higher levels of education are better equipped to comprehend web-based information and benefit from internet use [7]. We hypothesize that (H6) education contributes positively to COVID-19–related internet uses and outcomes.

When studying differences in internet uses and outcomes, the resources people can access are often derived from Pierre Bourdieu's capital theory [27], which stresses the importance of including not only economic but also social and cultural resources to determine one's status and position in society. In the COVID-19 pandemic, economic and social resources are likely to be important, as earlier research has shown that people with greater economic resources—mostly operationalized as income in digital inequality research—are known to use the internet more efficaciously and productively [7,28]. People with more social resources are more likely to have access to family, friends, or other contacts on the internet [29]. We hypothesize that (H7a) economic and (H7b) social resources contribute positively to COVID-19–related internet uses and outcomes.

The Internet Appropriation Process

The core of the resources and appropriation theory is access to technology, which is considered as a process of appropriation involving attitudinal, material, skills, and usage access. Attitudinal access concerns a person's attitude towards the internet; according to theories of technology adoption, this type of access is crucial for using the internet [30]. Material access can be defined in terms of the different devices that people use to access the internet and all other web-based resources, including desktop computers, laptop computers, tablets, smartphones, game consoles, and interactive televisions [31]. Skills access concerns the skills necessary to use the internet, ranging from operational and information skills to social and content creation skills [32]. Prior research has revealed that all three types of internet access directly affect internet uses and outcomes [16]. We hypothesize that (H8a) attitudinal internet

access, (H8b) material internet access, and (H8c) skills internet access contribute positively to COVID-19–related internet uses and outcomes.

The Effects of COVID-19–Related Internet Uses on Their Corresponding Outcomes

A recent multifaceted consideration of digital inequality revealed a strong effect of internet uses on outcomes [12]. Further, people's internet activities appeared to be more important than their personal characteristics with regard to inequalities in outcomes of internet use. This suggests that the variables discussed in the prior sections will become less important for obtaining information outcomes when people are involved in COVID-19–related internet information uses. This is also true for COVID-19–related communication uses and outcomes. The second goal of this paper is to reveal the extent to which the indicators discussed remain important for obtaining internet outcomes after people are involved in the corresponding uses.

Methods

Recruitment

This study used a web-based survey and drew upon a sample collected in the Netherlands. To obtain a representative sample of the population, we used PanelClix, a professional organization for market research, to provide a panel of approximately 110,000 people. Members of the panel received a small incentive for every survey they completed. In the Netherlands, 98% of the population uses the internet; therefore, the internet user population is very closely representative of the general population in terms of its sociodemographic makeup. The panel included novice and advanced internet users. In total, we aimed to obtain a dataset with approximately 1700 respondents over the age of 18. Eventually, this resulted in the collection of 1733 responses over a 1-week period in April 2020. During the data collection period, three amendments to the sampling frame were made to ensure the representativity of the Dutch population. Accordingly, the analyses revealed that the gender, age, and formal education of our respondents largely matched official census data. As a result, only very small post hoc corrections were needed.

The web-based survey used software that checked for missing responses and prompted users to respond. The survey was pilot-tested with 10 internet users over two rounds. Amendments were made based on the feedback provided. No major comments were provided in the second round. The average time required to complete the survey was 20 minutes.

Measures

We initially developed 11 survey items pertaining to COVID-19–related internet use. Respondents were asked to indicate the extent to which they used the internet for various activities in the past month using a 5-point scale ("not" to "multiple times a day") as an ordinal-level measure. Principal component analysis with varimax rotation was used to determine two underlying usage clusters, one related to information and one to communication. Factor loadings were employed at 0.4 and above for each item [33]. In total, 8 items (3 for information and 5 for communication) were retained in a two-factor structure with eigenvalues over 1.0, together accounting for 76% of the total variance.

For COVID-19-related information and communication internet outcomes, we developed 14 items mapped onto the use items. A 5-point agreement scale as an ordinal level measure was used. Principal component analysis with varimax rotation resulted in a structure that matched the conceptual definition of information outcomes (4 items) and communication outcomes (4 items). The two factors showed eigenvalues over 1.0 and explained 65% of the variance.

Gender was included as a dichotomous variable, and age was directly asked (mean 50.2, SD 17.0).

Personality was measured with the Quick Big Five personality questionnaire [34], which consists of 30 adjectives reflecting a valid and reliable measure of the Big Five traits. Participants were asked to rate the extent to which a particular adjective applied to them on a 7-point scale, ranging from completely untrue to completely true. The Cronbach α values for the five traits were .89 for agreeableness, .88 for neuroticism, .88 for conscientiousness, .87 for introversion, and .81 for openness.

Physical health, mental health, and health perception were measured with the Dutch version of the Medical Outcomes Study (MOS) Short-Form General Health Survey (SF-20) [35]. This instrument enables respondents to assess their general health and generates composite summary scores representing different types of health. We normalized the scales, with higher scores representing better functioning. Physical health was measured with 5 items (2-point scale; α=.89; mean 1.75, SD 0.34), mental health with 5 items (5-point scale; α=.85; mean 3.65, SD 0.77), and health perception with 5 items (5-point scale; α=.86; mean 3.39, SD 0.85).

To measure traditional literacy, we used the validated 11-item Diagnostic Illiteracy Scale [36]. Sample items included "I have difficulties with reading and understanding information from my municipality" and "I find it difficult to read and understand my telephone bill." A 5-point agreement scale was used. Scores on the scale exhibited high internal consistency. Items were recoded so that higher scores corresponded with higher levels of literacy (α=.94; mean 4.33, SD 0.71).

To assess education, data regarding degrees earned were collected and used to create three groups: low (primary), middle (secondary), and high (tertiary) educational achievement.

Economic resources were objectively measured by seeking the annual family income in the last 12 months. Twelve categories were recoded into three categories of low for <€30,000 (US $35,503.50), middle for €30,000 to €70,000 (US $35,503.50 to $82841.50), and high for >€70,000 (>US $82841.50). For social resources, we used the MOS Social Support Survey [37]. Respondents completed 18 items covering emotional support (eg, "Someone you can count on to listen when you need to talk"), informational support (eg, "Someone to give you good advice about a crisis"), and tangible support (eg, "Someone to help you if you were confined to bed"). All items were rated on a 5-point Likert scale with anchors of none of the time (1) and most of the time (5). We computed an aggregate measure of support availability (α=.96; mean 3.83, SD 0.85).

Attitudinal internet access was measured by three items adapted from the Digital Motivation Scale [38]. A 5-point agreement scale was used, and all items were balanced for the direction of response (α=.74; mean 4.10, SD 0.70). An example statement is "Technologies such as the internet and mobile phones make life easier." To measure material internet access, we considered 7 devices used to connect to the internet (mean 3.43, SD 1.53). Included were desktop computer, laptop computer, tablet, smartphone, smart TV, game console, and smart device (eg, activity tracker). Finally, skills internet access was adapted from the conceptual idea behind the Internet Skills Scale [32]. We proposed 30 items reflecting operational, information navigation, social, and creative internet skills. A 20-item single skills construct resulted from the principal component analysis. All items were scored on a 5-point scale that ranged from "not at all true of me" to "very true of me" and exhibited high internal consistency (α=.96; mean 3.67, SD 0.97). Example items are "I know how to open downloaded files," "I find it hard to decide what the best keywords are to use for online searches," and "I know which information I should and shouldn't share online."

Statistical Analysis

To test the hypotheses and account for the sequentiality between COVID-19–related internet uses and outcomes, hierarchical regression analyses were used. In the first model, we tested our hypotheses by analyzing the significant determinants for the two types of COVID-19–related internet uses and the two corresponding outcomes. In the second model, we sought to determine the changes in the significance of the determinants after the internet uses were added to the models.

Results

Table 1 provides an overview of the sample of people surveyed in the study.

Table 2 shows the mean scores of the survey questions related to internet uses and internet outcomes.

The first goal of this paper was addressed in the first model, as presented in Table 3, where several significant determinants for COVID-19 uses and outcomes are revealed.

Table 1. Demographic profile of the Dutch internet user sample (N=1733), n (%).

Characteristic	Value
Gender	
Male	874 (50.4)
Female	859 (49.6)
Age (years)	
18-30	280 (16.2)
31-40	271 (15.6)
41-50	293 (16.9)
51-60	338 (19.5)
61-70	324 (18.7)
>70	227 (13.1)
Education level[a]	
Low	519 (29.9)
Middle	602 (34.7)
High	612 (35.3)

[a]Low: primary; middle: secondary; high: tertiary.

Table 2. Survey questions and responses on the 5-point Likert scale.

Category and questions	α	Mean (SD)
COVID-19[a]–related informational internet uses	.80	3.13 1.53
Search the internet for information about COVID-19		3.76 1.91
Consult websites of public agencies (eg, RIVM[b], municipality, hospital, or government)		3.21 1.83
Search the internet for measures to prevent the further spread of COVID-19		2.44 1.71
COVID-19–related communication internet uses	.92	1.56 1.13
Provide advice on COVID-19 to others via social media		1.56 1.31
Starting an action against COVID-19 via the internet (eg collecting money, offering help)		1.41 1.17
Ask questions about COVID-19 on forums or social media		1.54 1.30
Comment on the internet on COVID-19 discussions (eg, on social media)		1.58 1.34
Offering help online to people who need it now		1.70 1.41
COVID-19–related information internet outcomes	.80	3.17 0.95
The internet makes me better informed about COVID-19		3.58 1.13
The internet makes me understand the measures against COVID-19 better		3.25 1.15
The internet helps me to reduce the risk of getting COVID-19		3.15 1.16
Information about COVID-19 on the internet has made me more aware of my own behavior		2.70 1.26
COVID-19–related communication internet outcomes	.80	1.91 0.89
Through the internet I found someone who can help me in this time of COVID-19		1.67 1.04
Through the internet I have found people with whom I can share my concerns about COVID-19		1.83 1.10
Via the internet I contributed to the COVID-19 crisis (eg, collecting money, helping people)		1.83 1.13
The internet makes me less lonely now		2.29 1.25

[a]COVID-19: coronavirus disease.

[b]RIVM: Rijksinstituut voor Volksgezondheid en Milieu.

Table 3. Hierarchical regression analysis summary for coronavirus disease–related internet uses and outcomes (Model 1).

Characteristic	Information				Communication			
	Use		Outcome		Use		Outcome	
	β	*P* value	β	*P* value	β	*P* value	β	*P* value
Gender and age								
Gender (male or female)	.01	.61	.00	.98	–.08	<.001	.01	.83
Age	.08	.01	–.03	.35	–.08	.003	–.11	<.001
Big Five personality traits								
Agreeableness	-.07	.03	–.01	.75	–.13	<.001	-.08	.003
Neuroticism	.15	<.001	.15	<.001	.05	.20	.08	.02
Conscientiousness	.01	.60	–.02	.52	.01	.54	–.04	.14
Introversion	–.04	.11	.02	.56	–.09	<.001	-.06	.02
Openness	.08	.004	.03	.30	.14	<.001	.15	<.001
Health status								
Physical health	–.04	.15	–.03	.31	–.15	<.001	–.10	<.001
Mental health	–.06	.15	.03	.41	–.06	.11	–.01	.82
Health perception	.07	.05	.04	.31	.16	<.001	.10	<.001
Literacy and education								
Traditional literacy	.09	<.001	.10	.18	.31	<.001	.33	<.001
Education	.08	.002	.02	.36	.07	.003	.02	.34
Resources								
Economic resources	.03	.23	.04	.13	–.01	.57	–.03	.23
Social resources	.02	.40	.00	.91	–.02	.36	–.01	.69
Access								
Attitudinal access	.14	<.001	.29	<.001	–.06	.01	–.04	.08
Material access	.10	<.001	.06	.02	.08	<.001	.07	.008
Skills access	.08	.006	.08	.008	.09	<.001	.09	<.001

Table 3 shows that men are more likely to be involved in COVID-19–related communication uses. Age is positively related to COVID-19–related information uses and negatively related to COVID-19 communication uses and outcomes. Concerning personality traits, agreeableness is negatively related to COVID-19–related information and communication uses and to communication outcomes. Neuroticism is positively related to both uses and to communication outcomes.

Conscientiousness is not related to any of the uses or outcomes. Introversion is negatively related to COVID-19–related communication uses and outcomes, suggesting that this is performed more by extraverted people. Finally, openness relates positively to information uses and to both outcomes.

The results further show that concerning the three health indicators, physical health is negatively related to communication uses and outcomes. Mental health did not contribute to any uses or outcomes. Health perception contributes positively to information uses and to both outcomes.

Traditional literacy has a positive relationship with information-type uses and with both outcomes, and education has a positive relationship with COVID-19–related information and communication uses. Economic and social resources were not related to any COVID-19 uses or outcomes.

Attitudinal internet access is positively related to information uses and outcomes but is negatively related to communication uses and outcomes. Material internet access contributes positively to all uses and outcomes, and skills access has a positive relationship with all uses and outcomes. Table 4 provides an overview of the hypotheses.

Table 4. Overview of the hypotheses.

Number	Hypothesis	Information uses	Information outcomes	Communication uses	Communication outcomes	Validation
H1	Gender (male or female)	ns[a]	ns	–[b]	ns	R[c]
H2	Age	+[d]	ns	–	–	PS[e]
H3a	Agreeableness	–	ns	–	–	PS
H3b	Neuroticism	+	+	ns	+	PS
H3c	Conscientiousness	ns	ns	ns	ns	R
H3d	Introversion	ns	ns	–	–	PS
H3e	Openness	+	ns	+	+	PS
H4a	Physical health	ns	ns	–	–	PS
H4b	Mental health	ns	ns	ns	ns	R
H4c	Health perception	+	ns	+	+	R
H5	Traditional literacy	+	ns	+	+	PS
H6	Education	+	ns	+	ns	PS
H7a	Economic resources	ns	ns	ns	ns	R
H7b	Social resources	ns	ns	ns	ns	R
H8a	Attitudinal access	+	+	–	–	PS
H8b	Material access	+	+	+	+	S[f]
H8c	Skills access	+	+	+	+	S

[a]ns: no significant contribution.

[b]–: significant negative contribution.

[c]R: reject.

[d]+: significant positive contribution.

[e]PS: partial support.

[f]S: support.

Finally, to address the second goal of the study, we tested what would happen to the contribution of the outcome determinants when the corresponding uses were added to the analyses (Model 2: see Tables 5 and 6). Adding the uses significantly increased the explained variance; also, several of the relationships between personal and positional categories and between resources and outcomes became insignificant. The relationships that remained significant for information outcomes were age, health perception, and traditional literacy. Furthermore, attitudinal internet access remained significant. For communication outcomes, the relationships that remained significant were age, openness, and traditional literacy.

Table 5. Hierarchical regression analysis summary for coronavirus disease–related internet outcomes (Model 2).

Characteristic	Information outcomes		Communication outcomes	
	β	*P* value	β	*P* value
Gender and age				
Gender (male or female)	–.01	.71	.04	.05
Age	–.07	.003	–.08	<.001
Big Five personality traits				
Agreeableness	.03	.25	–.02	.38
Neuroticism	.06	.04	.06	.06
Conscientiousness	–.02	.26	–.04	.05
Introversion	.04	.08	–.02	.40
Openness	-.02	.49	.08	<.001
Health status				
Physical health	.07	.05	–.03	.26
Mental health	.01	.79	.02	.59
Health perception	–.05	.02	.03	.34
Literacy and education				
Traditional literacy	.05	.02	.19	<.001
Education	–.02	.33	–.01	.67
Resources				
Economic resources	.02	.30	–.02	.28
Social resources	–.01	.67	.00	.99
Access				
Attitudinal access	.21	<.001	–.02	.49
Material access	.01	.74	.03	.21
Skills access	.03	.17	.05	.06
Information uses	.55	<.001	N/A[a]	N/A
Communication uses	N/A	N/A	.45	<.001

[a]N/A: not applicable.

Table 6. Changes in the significance of the determinants after internet uses were added to the models (*P*<.001).

Model and measures	Information outcomes		Communication outcomes	
	Use	Outcome	Use	Outcome
Model 1				
r^2	.09	.13	.23	.21
F	22.15	15.05	30.13	26.54
Model 2				
r^2	N/A[a]	.41	N/A	.37
r^2 change	N/A	.28	N/A	.16
F	N/A	63.71	N/A	54.72

[a]N/A: not applicable.

Discussion

Principal Results

This paper aims to provide a comprehensive examination of digital inequality in the case of an unprecedented health pandemic. The first goal of the study was to reveal how inequality manifests itself in COVID-19–related internet information and communication uses and outcomes. The findings revealed several relationships between the background variables and the two types of internet uses and outcomes.

Older people were found to be less equipped to use the internet for information and communication during a time of crisis. However, they were more likely to engage in information-type COVID-19–related internet uses, possibly because they are at greatest risk from the disease [11]. This did not result in more beneficial information outcomes. Internet skills play an important role in translating internet uses into beneficial internet outcomes [39], and prior research has shown that older people have lower internet skill levels in general [32]. The finding that older people are less likely to perform communication activities or obtain communication-related outcomes is in line with prior studies [15]; however, these outcomes are important, as older people are more at risk of having severe complications when diagnosed with COVID-19. Regarding gender, contrary to general internet use, men were found to be more likely to engage in communication-type COVID-19–related internet uses during the crisis than women. A possible explanation is that men and women may respond to crisis news in different ways [40].

The positive effect of neuroticism suggests that a tendency to experience negative emotions such as anger, anxiety, or depression fuels the need to turn to the internet for COVID-19–related information and communication. People who score higher on the neuroticism scale may be more in need of guidelines on how to mitigate risks or may need more support from others to be comforted. Also, the openness trait supports both information and communication internet use and outcomes. A possible explanation is that a major crisis triggers adventure, unconventional ideas, imagination, awareness of feelings, curiosity, or a variety of experiences, all of which are aspects linked to high openness [21]. The negative contribution of agreeableness raises questions. A possible explanation is that agreeable people are less frequently sought out for communication activities. However, the internet may also be a very inviting environment for less agreeable people. Conscientiousness did not appear to be a significant determinant. People who are more stubborn and focused or more flexible and spontaneous both appear to be involved in information- and communication-type COVID-19–related internet uses and outcomes. Extroversion emerged as a trait that supports using the internet for communication uses and outcomes; this can be expected, as extroversion is marked by pronounced engagement with the external world [21].

Although we expected that psychological distress would play a role in the current context, as there would be a relatively high need for information and support from others, mental health did not surface as a significant contributor. Furthermore, we did find that physical health problems appear to encourage web-based COVID-19–related communication uses and outcomes. The most likely explanation is that people with underlying health problems are more at risk (and thus more bound to their homes) and thus have higher needs for communication with friends and family. A possible reason for the positive effect of health perception is that people who believe their personal health to be good may feel better equipped to support others during the COVID-19 pandemic.

As expected, traditional literacy played an important role. A lack of general ability to read, write, and understand text further disadvantages individuals in the case of the COVID-19 pandemic, as they have less access to information and communication sources. COVID-19 is a new, unknown, and complicated disease with characteristics that are often described in difficult medical language that is not easy to read. Similar findings were found for educational attainment. Research has long shown that education is one of the most prominent positional variables in digital divide research [7]. However, our results suggest that when less educated individuals are involved in information and communication internet uses, they are as likely to achieve the corresponding outcomes as people with higher levels of education. This is an important finding for designing interventions for those of lower levels of education.

An effect of economic resources did not emerge in relation to COVID-19–related internet uses and outcomes. The participants' income did not make a difference in obtaining information and communication COVID-19–related internet outcomes. Earlier research often showed that income is especially important to consumptive and work-related internet uses [17], topics that are not considered here. Unexpectedly, social resources were not found to be influential. Apparently, a person who has an offline support network will not necessarily turn more to web-based information and communication support during a crisis.

Concerning internet access, we can first conclude that a person's internet attitude is important for engaging in information uses and gaining information outcomes. Unexpectedly, there was a negative contribution of internet attitude to communication uses and outcomes, suggesting that individuals who have a negative evaluation of the internet in general are more likely to engage in communication uses in the event of a major crisis. Both material and skills internet access played important roles in achieving all uses and outcomes. Using a higher diversity of devices was related to higher COVID-19–related internet use and to more outcomes. The opportunities devices offer are known to be related to inequalities in internet uses and outcomes. As each device offers its own specific characteristics and advantages, a higher diversity of devices supports a larger range of use activities and outcomes [31]. Furthermore, internet skills play a fundamental role in COVID-19–related uses and in obtaining beneficial outcomes [12].

In this paper, several indicators surfaced for people's web-based COVID-19–related uses and outcomes. The variety of important indicators raises the question of whether general policies to address digital inequalities in a time of crisis will be effective. The complex relationships between the different indicators on one hand and internet uses and outcomes on the other hand demand more focused policies, such as those related to health

indicators and the need for information to enhance health outcomes. This study reveals that the greater an individual's existing advantages, the more they benefit from the internet at a time of crisis; the converse is true as well. Marginalized people are likely to have fewer types of access available to take actions, behave as requested, or be comforted by help, creating a vicious cycle where already marginalized groups are further marginalized in a time of crisis.

To end on a positive note, the situation may become slightly less complex when we address the second goal of this paper. When people engage in information and communication internet uses in a crisis situation, their personal characteristics become less important to achieving the corresponding outcomes. This suggests that to achieve information and communication outcomes, policy or research should especially focus on encouraging people to engage in the corresponding internet uses, as we can assume to some extent that engagement with information and communication-related COVID-19 uses is the best way to achieve beneficial outcomes at a time when they are most needed.

Limitations

The current study was conducted in the Netherlands, a country whose citizens have very high household internet penetration and high levels of educational attainment. Although differences in educational background and income are present and were taken into consideration, the observed inequalities may be even stronger in countries with a less homogeneous population. Given that the greatest burden of deaths has been in countries with very diverse populations, race and associated factors are likely to play a major role.

The aim of this study was to provide a broader picture of inequality in relation to how the internet is used in the case of a major global health crisis. A broad range of determinants was considered, and the relative importance of these indicators was revealed. However, a deeper understanding and further investigation to reveal the exact underlying mechanisms that cause these indicators to play a role would provide additional explanations. This suggests that further qualitative research is needed not only to obtain in-depth understanding of the mechanisms but also to understand the consequences of the observed inequalities to complement the findings of the current quantitative approach.

Conclusions

Digital inequality is a major concern among national and international scholars and policy makers. In this paper, we aimed to provide a broader understanding in the case of a major health pandemic by using the ongoing COVID-19 crisis as a context for empirical work. Several groups of people were identified as vulnerable, such as older people and people with lower levels of education, physical health problems, higher levels of neuroticism, low literacy levels, and low levels of trust. The general conclusion is that people who are already relatively advantaged are more likely to use the information and communication opportunities provided by the internet to their benefit in a health pandemic, while more disadvantaged individuals are less likely to benefit. Therefore, the COVID-19 crisis is also an enforcer of existing inequalities.

Conflicts of Interest

None declared.

References

1. Mahase E. China coronavirus: WHO declares international emergency as death toll exceeds 200. BMJ 2020 Jan 31;368:m408. [doi: 10.1136/bmj.m408] [Medline: 32005727]
2. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet 2020 Mar;395(10228):931-934 [FREE Full text] [doi: 10.1016/s0140-6736(20)30567-5]
3. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
4. Horton R. Offline: 2019-nCoV—"A desperate plea". Lancet 2020 Feb;395(10222):400. [doi: 10.1016/s0140-6736(20)30299-3]
5. Cowper A. Covid-19: are we getting the communications right? BMJ 2020 Mar 06;368:m919. [doi: 10.1136/bmj.m919] [Medline: 32144115]
6. Deci E, Ryan R. Intrinsic motivation. In: Weiner I, Craighead WE, editors. The Corsini Encyclopedia Of Psychology. Hoboken, NJ: John Wiley & Sons; 2010:2.
7. Scheerder A, van Deursen AJAM, van Dijk JAGM. Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telemat Inform 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
8. Van Dijk JAGM. The Deepening Divide: Inequality in the Information Society. London, UK: Sage Publications; 2005.
9. van Deursen AJ, van Dijk JA. The digital divide shifts to differences in usage. New Media Soc 2013 Jun 07;16(3):507-526. [doi: 10.1177/1461444813487959]
10. Madianou M. Digital Inequality and Second-Order Disasters: Social Media in the Typhoon Haiyan Recovery. SM+S 2015 Sep 30;1(2):205630511560338. [doi: 10.1177/2056305115603386]

11. Wang L, He W, Yu X, Hu D, Bao M, Liu H, et al. Coronavirus disease 2019 in elderly patients: Characteristics and prognostic factors based on 4-week follow-up. J Infect 2020 Jun;80(6):639-645 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.019] [Medline: 32240670]
12. Van Deursen AJ, Helsper EJ. Collateral benefits of Internet use: Explaining the diverse outcomes of engaging with the Internet. New Media Soc 2018 Jul 30;20(7):2333-2351 [FREE Full text] [doi: 10.1177/1461444817715282] [Medline: 30581362]
13. Helsper EJ. A Corresponding Fields Model for the Links Between Social and Digital Exclusion. Commun Theor 2012 Oct 15;22(4):403-426. [doi: 10.1111/j.1468-2885.2012.01416.x]
14. Kraut R, Mukhopadhyay T, Szczypula J, Kiesler S, Scherlis B. Information and Communication: Alternative Uses of the Internet in Households. Inf Sys Res 1999 Dec;10(4):287-303. [doi: 10.1287/isre.10.4.287]
15. Blank G, Groselj D. Dimensions of Internet use: amount, variety, and types. Information, Communication & Society 2014 Feb 28;17(4):417-435. [doi: 10.1080/1369118X.2014.889189]
16. van Deursen AJAM, Courtois C, van Dijk JAGM. Internet Skills, Sources of Support, and Benefiting From Internet Use. International Journal of Human-Computer Interaction 2014 Mar 07;30(4):278-290. [doi: 10.1080/10447318.2013.858458]
17. van Deursen AJ, van Dijk JA, ten Klooster PM. Increasing inequalities in what we do online: A longitudinal cross sectional analysis of Internet activities among the Dutch population (2010 to 2013) over gender, age, education, and income. Telemat Inform 2015 May;32(2):259-272. [doi: 10.1016/j.tele.2014.09.003]
18. Zillien N, Hargittai E. Digital Distinction: Status‐Specific Types of Internet Usage. Soc Sci Q 2009;90(2):274-291. [doi: 10.1111/j.1540-6237.2009.00617.x]
19. Lazarus R, Folkman S. Cognitive theories of stress and the issue of circularity. In: Appley MH, Trumbull R, editors. Dynamics of Stress. Boston, MA: Spinger; 1986:63-80.
20. Lazarus R. Stress and Emotion: A New Synthesis. New York, NY: Springer; 2006:A.
21. John O, Srivastava S. The Big 5 trait taxonomy: History, measurement, and theoretical perspectives. In: Pervin LA, John OP, editors. Handbook of Personality: Theory and Research (2nd ed.). New York, NY: Guilford Press; 1999:102-138.
22. Landers RN, Lounsbury JW. An investigation of Big Five and narrow personality traits in relation to Internet usage. Comput Hum Behav 2006 Mar;22(2):283-293. [doi: 10.1016/j.chb.2004.06.001]
23. Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey. Reliability and validity in a patient population. Med Care 1988 Jul;26(7):724-735. [doi: 10.1097/00005650-198807000-00007] [Medline: 3393032]
24. Coiro J. Exploring Literacy on the Internet: Reading Comprehension on the Internet: Expanding Our Understanding of Reading Comprehension to Encompass New Literacies. Read Teach 2003;56(5):458-464.
25. van Deursen A, van Dijk J. Modeling Traditional Literacy, Internet Skills and Internet Usage: An Empirical Study. Interact Comput 2014 Jul 16;28(1):13-26. [doi: 10.1093/iwc/iwu027]
26. Frisch A, Camerini L, Diviani N, Schulz PJ. Defining and measuring health literacy: how can we profit from other literacy domains? Health Promot Int 2012 Mar;27(1):117-126. [doi: 10.1093/heapro/dar043] [Medline: 21724626]
27. Bourdieu P. The forms of capital. In: Richardson J, editor. Handbook of Theory and Research for the Sociology of Education. New York, NY: Greenwood; 1986:241-258.
28. DiMaggio P, Hargittai E, Celeste C, Shafer S. From unequal access to differentiated use: A literature review and agenda for research on digital inequality. In: Neckerman K, editor. Social Inequality. New York: Russell Sage Foundation; 2004:A-400.
29. Woolcock M, Narayan D. Social Capital: Implications for Development Theory, Research, and Policy. World Bank Res Obs 2000 Aug 01;15(2):225-249. [doi: 10.1093/wbro/15.2.225]
30. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
31. van Deursen AJ, van Dijk JA. The first-level digital divide shifts from inequalities in physical access to inequalities in material access. New Media Soc 2019 Feb 07;21(2):354-375 [FREE Full text] [doi: 10.1177/1461444818797082] [Medline: 30886536]
32. van Deursen AJ, Helsper EJ, Eynon R. Development and validation of the Internet Skills Scale (ISS). Inf Commun Soc 2015 Aug 25;19(6):804-823. [doi: 10.1080/1369118x.2015.1078834]
33. Field A. Discovering Statistics Using IBM SPSS Statistics. London, UK: Sage Publications; 2013.
34. Vermulst AA, Gerris JRM. Quick Big Five Personality Questionnaire. Guideline. Leeuwarden NL: LDC Publications; 2006.
35. Kempen GIJM, Brilman EI, Heyink J, Ormel J. MOS Short-Form General Health Survey. Groningen, Netherlands: Rijksuniversiteit Groningen; 1995.
36. De Greef M, van Deursen AJAM, Tubbing M. Development of the DIS-scale (Diagnostic Illiteracy Scale) in order to reveal illiteracy among adults. J Study Adult Educ Learn 2013;1:37-48.
37. Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med 1991 Jan;32(6):705-714. [doi: 10.1016/0277-9536(91)90150-B]
38. Helsper E, Smirnova S, Robinson D. DiSTO Youth. London School of Economics and Political Science. 2017. URL: http://www.lse.ac.uk/media-and-communications/research/research-projects/disto/disto-youth [accessed 2020-08-14]

39. Van Deursen AJAM, Helsper E, Eynon R, van Dijk JAGM. The compoundness and sequentiality of digital inequality. Int J Commun 2017;11:452-473.
40. Lachlan KA, Spence PR, Nelson LD. Gender Differences in Negative Psychological Responses to Crisis News: The Case of the I-35W Collapse. Communication Research Reports 2010 Feb;27(1):38-48. [doi: 10.1080/08824090903293601]

Abbreviations

COVID-19: coronavirus disease
MOS: Medical Outcomes Study

Edited by G Eysenbach; submitted 11.05.20; peer-reviewed by S LaValley, T Hale; comments to author 13.07.20; revised version received 16.07.20; accepted 03.08.20; published 20.08.20.

Please cite as:
van Deursen AJAM
Digital Inequality During a Pandemic: Quantitative Study of Differences in COVID-19–Related Internet Uses and Outcomes Among the General Population
J Med Internet Res 2020;22(8):e20073
URL: http://www.jmir.org/2020/8/e20073/
doi:10.2196/20073
PMID:32750005

Review

Digital Tools to Ameliorate Psychological Symptoms Associated With COVID-19: Scoping Review

Melvyn Zhang[1], MBBS, MRCPsych; Helen Elizabeth Smith[1], DM, FFPHM

Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore

Corresponding Author:
Melvyn Zhang, MBBS, MRCPsych
Family Medicine and Primary Care
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
11 Mandalay Road
Level 11, Family Medicine and Primary Care
308322
Singapore
Phone: 65 98556631
Email: melvynzhangweibin@gmail.com

Abstract

Background: In the four months after the discovery of the index case of coronavirus disease (COVID-19), several studies highlighted the psychological impact of COVID-19 on frontline health care workers and on members of the general public. It is evident from these studies that individuals experienced elevated levels of anxiety and depression in the acute phase, when they first became aware of the pandemic, and that the psychological distress persisted into subsequent weeks. It is becoming apparent that technological tools such as SMS text messages, web-based interventions, mobile interventions, and conversational agents can help ameliorate psychological distress in the workplace and in society. To our knowledge, there are few publications describing how digital tools have been used to ameliorate psychological symptoms among individuals.

Objective: The aim of this review was to identify existing SMS text message, web-based, mobile, and conversational agents that the general public can access to ameliorate the psychological symptoms they are experiencing during the COVID-19 pandemic.

Methods: To identify digital tools that were published specifically for COVID-19, a search was performed in the PubMed and MEDLINE databases from the inception of the databases through June 17, 2020. The following search strings were used: “NCOV OR 2019-nCoV OR SARS-CoV-2 OR Coronavirus OR COVID19 OR COVID” and “mHealth OR eHealth OR text”. Another search was conducted in PubMed and MEDLINE to identify existing digital tools for depression and anxiety disorders. A web-based search engine (Google) was used to identify if the cited web-based interventions could be accessed. A mobile app search engine, App Annie, was used to determine if the identified mobile apps were commercially available. Results: A total of 6 studies were identified. Of the 6 identified web-based interventions, 5 websites (83%) could be accessed. Of the 32 identified mobile interventions, 7 apps (22%) could be accessed. Of the 7 identified conversational agents, only 2 (29%) could be accessed.

Results: A total of 6 studies were identified. Of the 6 identified web-based interventions, 5 websites (83%) could be accessed. Of the 32 identified mobile interventions, 7 apps (22%) could be accessed. Of the 7 identified conversational agents, only 2 (29%) could be accessed.

Conclusions: The COVID-19 pandemic has caused significant psychological distress. Digital tools that are commercially available may be useful for at-risk individuals or individuals with pre-existing psychiatric symptoms.

(J Med Internet Res 2020;22(8):e19706) doi:10.2196/19706

KEYWORDS

COVID-19; digital tool; psychiatry; mental health; digital health; psychology; distress; stress; anxiety; depression

Introduction

On December 31, 2019, the World Health Organization (WHO) was alerted to a case of pneumonia of unknown cause that originated from China [1]. The causative pathogen was a novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and the disease it causes is now referred to as coronavirus disease (COVID-19) [1]. As numerous other individuals became afflicted, the WHO declared the outbreak of COVID-19 to be an international public health emergency [1]. The problem escalated, with many individuals across different countries being afflicted; on March 11, 2020, the WHO increased their alert to a pandemic [1]. The rapid escalation of the number of individuals infected, the increasing number of deaths, and the measures undertaken by governments have resulted in significant psychological impact not only on health care workers, but also on the public [2].

In the four months after the discovery of the index case, several studies highlighted the psychological impact of COVID-19 on frontline health care workers and on members of the public. Li et al [3] used a mobile phone app to administer the Chinese version of a vicarious traumatization questionnaire to 214 individuals from the public and 526 nurses (234 frontline nurses and 292 nurses not on the front line). They found that the vicarious traumatization scores were higher among the public and nonfrontline health care workers. Wang et al [4] investigated the mental health status of 1120 members of the public living in China using both the Impact of Event Scale–Revised (IES-R) and the Depression, Anxiety, and Stress Scale (DASS-21). Notably, 53.8% of respondents reported the psychological impact to be moderate to severe, with 16.5% having moderate to severe depressive symptoms, 28.8% having moderate to severe anxiety symptoms, and 8.1% having moderate to severe stress. These studies focused on the immediate psychological impact of the COVID-19 pandemic; meanwhile, in their most recent study, Wang et al [5] examined the longitudinal changes in the mental health of the general population in China. A total of 1738 participants were included, and they were administered the IES-R questionnaire and the DASS-21 at baseline and after 4 weeks. While there was a mean reduction in the overall scores across a period of 4 weeks, the mean scores for the IES-R scale were still high, suggesting the presence of posttraumatic stress disorder (PTSD) symptoms [5]. It is evident from these studies that individuals experienced elevated levels of anxiety and depression in the acute phase, when they first became aware of the pandemic, and that this psychological distress persisted into subsequent weeks.

Since the onset of the pandemic, the priorities for government have understandably been the treatment of people infected with COVID-19 and steps to limit spread. It is evident in many countries that the need for clinical services has exceeded the supply, requiring the construction of temporary medical facilities and redeployment of staff. Beyond hospitals, the impact of COVID-19 is multifactorial; economies are being affected, individuals are suffering bereavement after loss of loved ones, and others are physically isolated and quarantined. Social distancing and lockdowns have created difficulties in accessing mental health services. Individuals with psychiatric disorders are likely to have increasing difficulty accessing conventional mental health services. Hao et al [6] reported that mean PTSD, anxiety, depression, and insomnia scores were elevated in psychiatric patients compared to those in the general population. A need indeed exists for web-based mental health interventions that use digital tools. An editorial has also been published describing how web-based tools and social media have been used in China to support the mental health needs of frontline workers as well as of people who are infected or living in quarantine facilities [7]. It is becoming apparent that technological tools such as SMS text messages, web-based interventions, mobile interventions, and conversational agents can help ameliorate psychological distress in the workplace and society. In an opinion paper, Zhou et al [8] described the efforts of the Australian government to provide telemental health solutions to address the psychological impact of COVID-19. Unfortunately, the scope of this paper was limited to the identification of relevant services in Australia. In another recent paper, Cosic et al [9] highlighted the potential of digital tools in dealing with the psychological distress associated with COVID-19 and how their prior experiences can be applied in developing relevant apps; however, the authors failed to identify any existing digital tools that individuals can use. To our knowledge, except for the paper by Zhou et al [8], few publications have described how digital tools are being used to ameliorate psychological symptoms among individuals. Thus, our aim in this paper was to identify existing SMS text message, web-based, mobile, and conversational agents that the public can access to ameliorate the psychological symptoms they are facing during the COVID-19 pandemic.

Methods

To identify digital tools that have been published specifically for COVID-19, a search was performed through PubMed and MEDLINE from the inception of the databases through June 17, 2020. The following search strings were used: "*NCOV* OR *2019-nCoV* OR *SARS-CoV-2* OR *Coronavirus* OR *COVID19* OR *COVID*" and "*mHealth* OR *eHealth* OR *text*". Given that the aim was to identify potential digital tools, the terms electronic health (*eHealth*), mobile health (*mHealth*), and *text* were used because these terminologies would identify all potential web-based, mobile, and SMS text message interventions.

Another search was conducted on PubMed and MEDLINE to identify existing digital tools for depression and anxiety disorders. To identify these tools, reviews of digital tools (SMS text messaging, web-based interventions, mobile apps, and conversational agents) were identified. Only articles in the English language were considered. A narrative synthesis of the identified tools was conducted.

A web-based search engine (Google) was used to identify if the cited web-based interventions could be accessed. A mobile app search engine, App Annie [10], was used to identify if the mobile apps identified in the literature search were commercially available. All available data have been included in the manuscript.

Results

Based on the search strategy, a total of 9829 articles were identified from the databases. Of these, 80/9829 (0.8%) were duplicated references. Upon further screening, a total of 24/9829 articles (0.2%) were identified as potentially relevant to COVID-19. Upon further examination of the full texts of these 24 articles, only one article described how an SMS text messaging intervention was applied to address mental health issues resulting from COVID-19 [11]. The remaining articles did not describe or mention how the discussed SMS text message, web-based, or mobile interventions helped ameliorate the psychological symptoms associated with COVID-19.

In our search for existing digital tools to manage depression and anxiety disorders, a total of 5 articles were identified. Two web-based reviews were identified for depressive disorders, and another was identified for anxiety disorders. Of the identified web-based interventions, we managed to access 5 of the 6 listed websites (83%), namely Beating the Blues [12], Living Life to the Full [13], Deprexis [14], moodgym [15], and Interapy [16]. One review of mobile interventions for depressive and anxiety disorders was identified. Of the identified mobile interventions, only 7/32 apps (22%) were available in commercial stores, namely Angesthjalpen, AnxietyCoach, SmartCAT, Headgear, MoodHacker, SuperBetter, and Thought Challenger. Another review highlighted conversational agents for psychiatric disorders. Of the 7 identified conversational agents, only 2 (29%) were commercially available. Table 1 provides a summarized overview of the identified studies.

Table 1. Overview of the studies identified in the literature search.

Study	Year	Mechanism of delivery of digital tools	Identified digital tools and prior evaluations	Availability
Agyapong et al [11]	2020	SMS text messaging (specific to COVID-19[a])	Text4Hope (specific to COVID-19) enables subscribers to receive 3 months of daily supportive SMS text messages with or without web links to web-based mental health resources.	Only available to individuals living in Alberta, Canada.
Rodriguez-Pulido et al [17]	2020	Web-based Interventions	Beating the Blues is a web-based intervention for depressive disorder.	Available to users in the United Kingdom.
Burger et al [18]	2020	Web-based interventions	Living Life to the Full (2 comparative trials were performed involving a total of 659 participants), Deprexis (6 comparative trials were performed involving a total of 1863 participants), and SHADE (3 comparative trials performed involving a total of 475 participants) were evaluated. moodgym was extensively evaluated, with a total of 11 comparative trials involving a total of 7294 participants. All the above interventions have been evaluated as websites that provide psychological therapy for depressive disorders.	The websites for Living Life to the Full, Deprexis, and moodgym can be accessed.
Anderson et al [19]	2019	Web-based interventions	The Interapy program from the Netherlands was highlighted as a program that assisted individuals with symptoms of depression, panic disorder, posttraumatic stress disorder, and burnout. moodgym was also highlighted as a commercially available option for anxiety and depression.	The websites for moodgym and Interapy can be accessed.
Miralles et al [20]	2020	Mobile interventions	7 Cups, Be Good to Yourself, Bluewatch, Dcombat, Get Happy Program, Headgear, iCare Prevent, MedLink, Mobile Sensing and Support, Moodhacker, Moodivate, MyGamePlan, PRIME-D, Push-D, SocioEmpathy, SPSRS, SuperBetter, The Sound Advice, Thought Challenger, TODAC, Kokoro-App, Agoraphobia free, Stress Free, Angesthjalpen, AnxietyCoach, CBT Assistant, Challenger, Lantern, Psych Assist, Public Speech Trainer, SmartCAT, and GET.ON.PAPP have been previously evaluated and reported in published research.	Headgear, MoodHacker, SuperBetter, Thought Challenger, Angesthjalpen, AnxietyCoach, and SmartCAT are commercially available.
Gaffney et al [21]	2019	Conversational agents or chatbots	Woebot, Tess, and eSMART-TH have been evaluated previously for depressive disorder, SABORI has been evaluated for psychological distress, and Tess has been evaluated for anxiety disorder.	Woebot and Tess are commercially available.

[a]COVID-19: coronavirus disease.

Discussion

Principal Findings

This review is one of the first to examine the literature for digital interventions that can be used by the general public as well as specific groups, such as workers and health care professionals, to ameliorate the psychological distress they are experiencing during the COVID-19 pandemic as well as specific symptoms such as panic buying [22-25]. The findings from our paper complement those of Zhou et al [8], who highlighted tools in their article that can be accessed by individuals in Australia. Our review helped address some of the inherent limitations of the work by Zhou et al [8], given that the authors only listed available resources without providing any evidence-based justification of the suggested interventions. We identified an SMS text message–based intervention that was designed to address the mental health needs of individuals in Canada. We managed to identify web-based, mobile, and conversational agents that are commercially available and have been previously validated by research.

Our review only identified one publication that describes how SMS text messaging technologies are used as a form of psychological support. As mentioned in the Introduction, numerous studies have been published that characterize the immediate and delayed psychological impact of COVID-19 on medical workers and members of the public [2,4,5]. There is still a lack of evaluation of psychological tools to address the identified psychological concerns, namely heightened levels of stress, anxiety, and depression. Psychological distress occurs frequently in everyday life; however, during epidemics and pandemics such as COVID-19, the prevalence of distress is extremely high and existing mental health services are unable to function normally. This provides strong justification for the rapid identification of tools that have an evidence base and can be promoted rapidly to address the unmet psychological needs of individuals. Our research highlighted commercially available digital tools (web-based, mobile, and conversational agents) that individuals can access during the COVID-19 pandemic. It is challenging for members of the public to identify tools that have been proven to be clinically effective and are available commercially; hence, our review is important. We focused primarily on reviews of digital tools to identify such tools. Our methods helped address the limitations of the prior work by Zhou et al [8], as they suggested tools but made no attempt to review the evidence base of those tools. The evidence-based websites, smartphone apps, and conversational agents we identified can help ameliorate symptoms of depression and anxiety. These tools have wide application; they can help individuals who are at risk of developing an illness or help individuals with pre-existing illness to cope with these symptoms. This is important because it is anticipated that it will be challenging for people to obtain appropriate psychiatric care as governments impose lockdowns and curb movement.

We identified a variety of commercially available tools; however, there may be limitations to some of these tools. Validations may have been conducted in certain localities or regions, and we cannot be sure that these tools will be as effective in other localities. However, we propose that a tool that has undergone validation, if only in a precise locality, is likely to be superior to a tool without any validation. We also recognize that accessing smartphone apps may be difficult in a different region or country.

It is evident from this paper that psychological tools to help individuals cope with heightened stress, anxiety, and depression due to COVID-19 are lacking. While some commercial tools are available, they are not without limitations. It is important for academic researchers, clinicians, and developers to work jointly to conceptualize tools that can be used by the general population to ameliorate their symptoms of psychological distress. It may also be valuable to consider participatory action research design when conceptualizing new tools to ensure that the created tools better meet the needs of individuals. It may be wise to consider modification of existing tools so that versatile tools can be rapidly deployed to meet the increasing need. In the interim period, while comprehensive treatment tools may not yet be available, we learned from the editorial by Liu et al [7] that in China, helplines and social media platforms are being used to extend support to individuals who are experiencing psychological distress. Similarly, in other countries such as Singapore, the government has set up a mental health hotline to address the needs of the public and to refer at-risk individuals to appropriate mental health services. These clinical services provide some form of supportive therapy; however, there is still a need for tools that provide individuals with more comprehensive treatment, such as cognitive behavioral therapy for depression or anxiety.

Strengths and Limitations

The strength of this paper is that we examined the literature for digital tools that have been validated, which can help with depressive and anxiety symptoms. In addition, we conducted a search to determine if these tools were available commercially. It is possible, despite these strengths, that we have missed some tools; to mitigate this, we included recently published reviews, but we acknowledge that their search for digital interventions may not be as recent.

Conclusions

The COVID-19 pandemic has caused significant psychological distress. Commercially available digital tools may be useful for at-risk individuals or individuals with pre-existing psychiatric symptoms. The tools we identified may help address the psychological distress individuals are experiencing during the COVID-19 pandemic.

Acknowledgments

This work received no funding.

Authors' Contributions

MZ and HS conceptualized the study. MZ worked on and wrote the initial draft of the manuscript. HS revised the initial draft. All authors approved the manuscript prior to submission.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-03-26]
2. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. Asian J Psychiatr 2020 Apr 10;52:102066 [FREE Full text] [doi: 10.1016/j.ajp.2020.102066] [Medline: 32302935]
3. Li Z, Ge J, Yang M, Feng J, Qiao M, Jiang R, et al. Vicarious traumatization in the general public, members, and non-members of medical teams aiding in COVID-19 control. Brain Behav Immun 2020 Aug;88:916-919 [FREE Full text] [doi: 10.1016/j.bbi.2020.03.007] [Medline: 32169498]
4. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
5. Wang C, Pan R, Wan X, Tan Y, Xu L, McIntyre RS, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. Brain Behav Immun 2020 Jul;87:40-48 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.028] [Medline: 32298802]
6. Hao F, Tan W, Jiang L, Zhang L, Zhao X, Zou Y, et al. Do psychiatric patients experience more psychiatric symptoms during COVID-19 pandemic and lockdown? A case-control study with service and research implications for immunopsychiatry. Brain Behav Immun 2020 Jul;87:100-106 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.069] [Medline: 32353518]
7. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. Lancet Psychiat 2020 Apr;7(4):e17-e18 [FREE Full text] [doi: 10.1016/S2215-0366(20)30077-8] [Medline: 32085841]
8. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The Role of Telehealth in Reducing the Mental Health Burden from COVID-19. Telemed J E Health 2020 Apr 23;26(4):377-379 [FREE Full text] [doi: 10.1089/tmj.2020.0068] [Medline: 32202977]
9. Ćosić K, Popović S, Šarlija M, Kesedžić I. Impact of Human Disasters and COVID-19 Pandemic on Mental Health: Potential of Digital Psychiatry. Psychiatr Danub 2020;32(1):25-31 [FREE Full text] [doi: 10.24869/psyd.2020.25] [Medline: 32303026]
10. App Annie. URL: https://www.appannie.com/en/ [accessed 2020-04-25]
11. Agyapong VI. Coronavirus Disease 2019 Pandemic: Health System and Community Response to a Text Message (Text4Hope) Program Supporting Mental Health in Alberta. Disaster Med Public Health Prep 2020 Apr 22:1-2 [FREE Full text] [doi: 10.1017/dmp.2020.114] [Medline: 32317038]
12. Beating the Blues. URL: https://www.beatingtheblues.co.uk/ [accessed 2020-08-05]
13. Living Life to the Full. URL: https://lltf.com/ [accessed 2020-08-05]
14. Deprexis. URL: https://info.deprexis.com/ [accessed 2020-08-05]
15. moodgym. URL: https://moodgym.com.au/ [accessed 2020-08-05]
16. Interapy. Webpage in Dutch. URL: https://www.interapy.nl/?gclid=EAIaIQobChMI6_eczO6E6wIVEIzICh2vjgMSEAAYASAAEgIpJvD_BwE [accessed 2020-08-05]
17. Rodriguez-Pulido F, Castillo G, Hamrioui S, Martin LD, Vazquez-Beltrán P, de la Torre-Díez I, et al. Treatment of Depression in Primary Care with Computerized Psychological Therapies: Systematic Reviews. J Med Syst 2020 Feb 13;44(3):67. [doi: 10.1007/s10916-020-1543-7] [Medline: 32060635]
18. Burger F, Neerincx MA, Brinkman W. Technological State of the Art of Electronic Mental Health Interventions for Major Depressive Disorder: Systematic Literature Review. J Med Internet Res 2020 Jan 20;22(1):e12599 [FREE Full text] [doi: 10.2196/12599] [Medline: 31958063]
19. Andersson G, Carlbring P, Titov N, Lindefors N. Internet Interventions for Adults with Anxiety and Mood Disorders: A Narrative Umbrella Review of Recent Meta-Analyses. Can J Psychiatry 2019 Jul 16;64(7):465-470 [FREE Full text] [doi: 10.1177/0706743719839381] [Medline: 31096757]
20. Miralles I, Granell C, Díaz-Sanahuja L, Van Woensel W, Bretón-López J, Mira A, et al. Smartphone Apps for the Treatment of Mental Disorders: Systematic Review. JMIR mHealth uHealth 2020 Apr 02;8(4):e14897 [FREE Full text] [doi: 10.2196/14897] [Medline: 32238332]
21. Gaffney H, Mansell W, Tai S. Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]

22. Tan W, Hao F, McIntyre RS, Jiang L, Jiang X, Zhang L, et al. Is returning to work during the COVID-19 pandemic stressful? A study on immediate mental health status and psychoneuroimmunity prevention measures of Chinese workforce. Brain Behav Immun 2020 Jul;87:84-92 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.055] [Medline: 32335200]
23. Chew NW, Lee GK, Tan BY, Jing M, Goh Y, Ngiam NJ, et al. A multinational, multicentre study on the psychological outcomes and associated physical symptoms amongst healthcare workers during COVID-19 outbreak. Brain Behav Immun 2020 Aug;88:559-565 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.049] [Medline: 32330593]
24. Tan BY, Chew NW, Lee GK, Jing M, Goh Y, Yeo LL, et al. Psychological Impact of the COVID-19 Pandemic on Health Care Workers in Singapore. Ann Intern Med 2020 Apr 06:online. [doi: 10.7326/m20-1083]
25. Ho C, Chee C, Ho R. Mental Health Strategies to Combat the Psychological Impact of COVID-19 Beyond Paranoia and Panic. Ann Acad Med Singapore 2020 Mar 16;49(3):155-160 [FREE Full text] [Medline: 32200399]

Abbreviations

COVID-19: coronavirus disease
DASS-21: Depression, Anxiety, and Stress Scale
eHealth: electronic health
IES-R: Impact of Event Scale–Revised
mHealth: mobile health
PTSD: posttraumatic stress disorder
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach; submitted 28.04.20; peer-reviewed by R Ho, A Powell; comments to author 12.06.20; revised version received 18.06.20; accepted 24.07.20; published 21.08.20.

Please cite as:
Zhang M, Smith HE
Digital Tools to Ameliorate Psychological Symptoms Associated With COVID-19: Scoping Review
J Med Internet Res 2020;22(8):e19706
URL: http://www.jmir.org/2020/8/e19706/
doi:10.2196/19706
PMID:32721922

Original Paper

Chinese Public's Engagement in Preventive and Intervening Health Behaviors During the Early Breakout of COVID-19: Cross-Sectional Study

Zhaomeng Niu[1*], PhD; Tingting Wang[2*], PhD; Pengwei Hu[3*], PhD; Jing Mei[3], PhD; Zhihan Tang[4], MD

[1]Rutgers Cancer Insititute of New Jersey, New Brunswick, NJ, United States

[2]Sun Yat-Sen University, Guangzhou, China

[3]IBM Research, Beijing, China

[4]Hengyang Medical College, University of South China, Hengyang, China

[*]these authors contributed equally

Corresponding Author:
Zhihan Tang, MD
Hengyang Medical College
University of South China
28 West Changsheng Road
Hengyang, 421001
China
Phone: 86 135 7478 0245
Email: tangzhihan98@163.com

Abstract

Background: Since January 2020, the coronavirus disease (COVID-19) swept over China and then the world, causing a global public health crisis. People's adoption of preventive and intervening behaviors is critical in curbing the spread of the virus.

Objective: The aim of this study is to evaluate Chinese people's adoption of health behaviors in responding to COVID-19 and to identify key determinants for their engagement.

Methods: An anonymous online questionnaire was distributed in early February 2020 among Mainland Chinese (18 years or older) to examine their engagement in preventive behaviors (eg, frequent handwashing, wearing masks, staying at home) and intervening behaviors (eg, advising family to wash hands frequently), and to explore potential determinants for their adoption of these health behaviors.

Results: Out of 2949 participants, 55.3% (n=1629) reported frequent engagement in preventive health behaviors, and over 84% (n=2493) performed at least one intervening health behavior. Greater engagement in preventive behaviors was found among participants who received higher education, were married, reported fewer barriers and greater benefits of engagement, reported greater self-efficacy and emotional support, had greater patient-centered communication before, had a greater media literacy level, and had greater new media and traditional media use for COVID-19 news. Greater engagement in intervening behaviors was observed among participants who were married, had lower income, reported greater benefits of health behaviors, had greater patient-centered communication before, had a lower media literacy level, and had a greater new media and traditional media use for COVID-19 news.

Conclusions: Participants' engagement in coronavirus-related preventive and intervening behaviors was overall high, and the associations varied across demographic and psychosocial variables. Hence, customized health interventions that address the determinants for health behaviors are needed to improve people's adherence to coronavirus-related behavior guidelines.

(J Med Internet Res 2020;22(8):e19995) doi:10.2196/19995

KEYWORDS

COVID-19; China; preventive health behaviors; intervening health behaviors; psychosocial; health literacy; behavior; prevention; cross-sectional

Introduction

In late December 2019, a novel coronavirus disease (COVID-19) emerged in Wuhan, Hubei, China, causing acute pneumonia by the severe acute respiratory syndrome coronavirus 2. COVID-19, being highly infectious and capable of human-to-human transmission, rapidly swept over China (85,921 total confirmed cases in China by July 23, 2020) and the world (15.4 million confirmed cases worldwide by July 23, 2020), developing into a global pandemic [1,2].

The Chinese government promptly implemented nationwide public health emergency measures to control the spread of COVID-19 [3]. Governmental public health strategies have been proven effective in containing infectious diseases [4]. Apart from governmental efforts, the general public plays a crucial role in conquering diseases [5]. There has been compelling evidence that the public's compliance with precautionary behaviors helps effectively curb the spread of many diseases [5,6]. Thus, it is of great value to evaluate Chinese people's adoption of health behaviors in responding to COVID-19.

This paper focuses on two distinct types of individual-level health behaviors—preventive health behaviors (PHBs) and intervening health behaviors (IHBs)—in responding to COVID-19. PHB refers to the activities undertaken by a healthy person for the purpose of preventing diseases [7]. In other words, people adopt PHBs to achieve the goal of lessening their own chance of contracting a disease. With the rising number of deaths caused by COVID-19, the Chinese people have significant concerns over COVID-19 for themselves, which would prompt them to adopt PHB. Critically, when facing a public health crisis, people also persuade other people to adopt precautionary behaviors that serve to reduce other people's risk of contracting the diseases [8,9]. We define the behaviors with a coherent objective of reducing other individuals' risk to a disease as IHB. As suggested by the definitions, PHB and IHB differ in the locus of intention, with the former serving to protect oneself, while the latter aims to protect others from potential risks. Chinese people not only worry for themselves but also have concern for their significant others over COIVD-19, which motivates IHB. Besides, home quarantine provided Chinese people with ample time and opportunity to communicate with and to influence their significant others (ie, engaging in IHB) both online and offline.

Both PHB and IHB contribute to curbing the spread of infectious diseases. On the one hand, PHB is self-serving, lowering one's own vulnerability to a disease [10-12]. However, despite the health benefits of PHB, there exist variations in people's adoption of preventive behaviors [5]. Hence, a close examination of the prevalence and potential correlates of people's engagement in preventive behaviors toward COVID-19 is called for. On the other hand, IHB is other-serving, reducing others' risks. A person's active intervening health behaviors targeted at others may successfully persuade other people to adopt precautionary measures against the disease because, as social beings, people's behaviors are subject to the influence of social relationships [13]. In essence, intervening behaviors can be treated as performing PHBs on behalf of others. Henceforth, we examine the influence of the same set of potential determinants on people's engagement in PHB and IHB.

In particular, we employed the key components of the Preventive Health Model (PHM) [14]. PHM posits that people's adoption of preventive behaviors are subject to the impacts of social influence (ie, social support and doctor-patient communication), psychological variables (ie, barriers, benefits, and self-efficacy of conducting precautionary behaviors), and program factors (eg, promotional communication or health information in media) [14]. All these factors were examined in our study. Additionally, we examined people's media use behaviors. A considerable amount of research has found that media use of different platforms has an influence on people's health behaviors [15-17]. Media literacy, defined as "the ability to access, analyze, evaluate, and create media in a variety of forms" [18], also has impact on people's different health behaviors [19-21].

Apart from the previously mentioned psychographic variables, demographic variables have also been revealed to partially explain the variation in people's propensity to adopt disease-related health behaviors [22]. As the Chinese population varies significantly along with demographics, we also included demographic variables in our investigation.

Ever since its outbreak in China, COVID-19 has seized national attention. China's unprecedented and relentless efforts started to pay off in late March [1]. Unfortunately, the number of confirmed cases in other countries is rising [1]. The world's fight against the coronavirus has just begun. Under these circumstances, investigating the Chinese public's engagement in coronavirus-related health behaviors and identifying the psychological and demographic variables that are significantly associated with these behaviors are urgently called for. This paper aims to examine the demographic and psychological correlates of preventive and intervening behaviors during the outbreak of COVID-19, which could generate insights for implementing health interventions among the general public and helping with effective containment of COVID-19.

Methods

Recruitment

Using the service of a Chinese survey company, an online survey was distributed on different local social media platforms in China, such as WeChat and Baidu Post Bar, to access Mainland Chinese residents from February 2, 2020, to February 12, 2020, when COVID-19 began breaking out in China. The foci of the survey were to evaluate Chinese people's propensity to engage in preventing health behaviors and IHBs, and to disentangle key determinants for people's adoption of such protective measures. The study was approved by the Institutional Review Board at Lingnan (University) College, Sun Yat-sen University.

An electronic consent form was presented at the beginning of the survey. Only participants who were 18 years or older, currently located in China, and agreed to participate after reading the consent form were allowed to proceed in the survey. Respondents who completed the survey entered a lucky draw for a monetary incentive of around ¥6.00 (US $0.86).

Measures

Our operationalizations of PHB and IHB followed the precaution behaviors that are recommended by the World Health Organization (WHO) for healthy people in responding to COVID-19 [23]. Specifically, *PHBs* were measured with five 5-point Likert scale (1=not at all, 5=very frequent) items that asked participants to report their frequency of engaging in the following behaviors: "wearing masks," "washing hands," "sanitizing clothes or other items," "sneezing into their elbows," and "staying at home" (α=.72). To assess participants' engagement in coronavirus-related *IHBs*, we instructed participants to indicate whether or not they had persuaded their social others such as family and friends to "wear masks," "wash hands," "sanitize clothes or other items," "staying at home," and "sneeze into elbows" (dichotomous variables; 0=no, 1=yes).

Next, we assessed the potential psychosocial determinants for PHB and IHB: perceived barriers and benefits of taking preventive measures, self-efficacy, emotional support, and patient-centered communications. All variables were measured along 5-point scales, anchoring from 1 (strongly disagree or not at all) to 5 (strongly agree or very much).

We assessed participants' perceived *barriers* (two items: "It is hard to buy masks" and "It is difficult to get sanitizers"; α=.84) and perceived *benefits* of preventive behaviors (two items: "Wearing face masks can help prevent the spread of the coronavirus" and "Using sanitizers can help prevent the spread of the coronavirus"; α=.87).

Following this, we evaluated participants' *self-efficacy* by measuring their confidence at addressing the risk of COVID-19 (two questions: "How confident are you at your preventing behaviors toward the coronavirus?" and "How confident are you that you will not be infected with the coronavirus?" [24]; α=.86).

Emotional support was measured by one question on a 5-point Likert scale (1=strongly disagree, 5=strongly agree): "During the outbreak of the coronavirus, my friends or family have provided me with emotional support when I need it - such as talking over problems" (mean 4.01, SD 1.06) [25].

Afterwards, *patient-centered communication* was assessed by instructing participants to evaluate their previous experience with health care providers along with four items: "In general, my feelings were taken seriously," "I was given a chance to ask all the health-related questions," "My healthcare providers made sure I understand the things I needed to do to take care of my health," and "My healthcare providers explained things in a way that I could understand" [26] (α=.94).

Additionally, *media literacy* was measured by four 5-point Likert scale items including "I look for more information before I believe something I see in messages," "It is important to think twice about what messages say," "I think about the purpose behind messages I see," and "I think about the truthfulness of messages before I accept them as believable" (α=.84) [27]. One question was used to ask participants' media use for COVID-19 news: "How frequently do you receive coronavirus-related news and/or information from the following media channels?" (1=never, 5=very frequent). *Social media use* was measured by four items: "Weibo," "Wechat messages," "Wechat public news accounts," and "QQ messages or Qzone" (α=.62). *Traditional media use* was measured by three items: "TV," "broadcast," and "newspapers" (α=.76). Digital news media was measured by one item: "news app, news on websites or other format of news on the internet other than social media" (mean 3.30, SD 1.28).

Finally, participants provided their basic demographic information (age, gender, marital status, education background, and income level) and ended the survey. Details of scales used for variables under study are shown in Table 1.

Table 1. Measurement of study variables.

Variable and items	Cronbach α	Range
Preventive behaviors: how frequently are you engaging in the following behaviors?	.72	1-5
Wearing masks		
Washing hands		
Sanitizing clothes or other items		
Sneezing into your elbows		
Staying at home (avoid going out)		
Intervening behaviors: please indicate whether or not you have persuaded your social others such as families and friends to:[a]	N/A[b]	0-5
Wear masks		
Wash hands		
Sanitize clothes or other items		
Stay at home (avoid going out)		
Sneeze into elbows		
Barriers	.84	1-5
It is hard to buy masks.		
It is difficult to get sanitizers.		
Benefits	.87	1-5
Wearing face masks can help prevent the spread of the coronavirus.		
Using sanitizers can help prevent the spread of the coronavirus.		
Self-efficacy	.85	1-5
How confident are you at your preventing behaviors toward the coronavirus?		
How confident are you that you will not be infected with the coronavirus?		
Emotional support	N/A	1-5
During the outbreak of coronavirus, my friends or family have provided me with emotional support when I need it, such as talking over problems.		
Patient-centered communication	.94	1-5
In general, my feelings were taken seriously.		
I was given a chance to ask all the health-related questions.		
My health care providers made sure I understand the things I needed to do to take care of my health.		
My health care providers explained things in a way that I could understand.		
Media literacy	.84	1-5
I look for more information before I believe something I see in messages.		
It is important to think twice about what messages say.		
I think about the purpose behind messages I see.		
I think about the truthfulness of messages before I accept them as believable.		
Social media: how frequently do you receive coronavirus-related news or information from the following media channels?	.62	1-5
Weibo		
Wechat messages		
Wechat public news accounts		
QQ messages or Qzone		
Traditional media	.76	1-5
TV		

Variable and items	Cronbach α	Range
Broadcast		
Newspapers		
Internet news channels other than social media	N/A	1-5
News apps, news on websites, or other format of news on the internet other than social media		

[a]The response was dichotomous.

[b]N/A: not applicable.

Statistical Analysis

Since our hypotheses were developed based on the PHB model, we conducted two multiple regression analyses to examine the associations between the independent variables and the two behavioral outcomes. All analyses were performed in SPSS 25 (IBM Corp).

Results

User Statistics

A response rate of 55.30% (2980/5388) was obtained. A pretest of our survey with 7 volunteers revealed that the time spent on the questionnaire ranged from 5 to 31 minutes. Following the advice of the survey company and glancing over the answers, completed surveys that took less than 5 minutes were most likely invalid. Therefore, questionnaires that took less than 5 minutes were excluded from the analyses. Additionally, we excluded respondents who had missing data on key variables (independent and dependent variables) in this study. Since outliers of the data set may affect regression results [28], we dropped extreme data points using the explore function and checking the box plot; 2949 participants (18-85 years, mean age 31, SD 0.65 years) were included in the final analyses. Among all participants, 51.2% (n=1509) were female, 54.5% (n=1607) were married, 22.2% (n=656) had an annual household income of ¥100,000-¥150,000 (US $14,389-$21,584), and 49.8% (n=1467) had a college degree or above. Demographic information of the sample is shown in Table 2.

Table 2. Demographic and socioeconomic characteristics of the sample.

Characteristic	Participants (N=2949), n (%)
Sex	
Male	1440 (48.8)
Female	1509 (51.2)
Age (years)	
18-24	786 (26.7)
25-29	549 (18.6)
30-34	653 (22.1)
35-39	477 (16.2)
40-44	189 (6.4)
45-49	138 (4.7)
50-54	79 (2.7)
55-59	53 (1.8)
≥60	25 (0.8)
Marital status	
Single	1342 (45.5)
Married	1607 (54.5)
Education	
High school graduate or less	727 (24.7)
Professional school	755 (25.6)
Bachelor's degree	1131 (38.4)
Postgraduate degree	336 (11.4)
Income, ¥ (US $)	
<70,000 (10,072)	1030 (35.2)
70,001-100,000 (10,073-14,389)	519 (17.6)
100,001-150,000 (14,390-21,583)	656 (22.2)
150,001-300,000 (21,584-43,167)	479 (16.2)
>300,001 (43,168)	255 (8.6)

Evaluation Outcomes

Preventive Health Behaviors

On average, participants' total frequency score of engaging in preventive behaviors had a mean of 4.00 (SD 0.65), with a possible range of 1 to 5. Of the 2949 respondents, approximately 55.3% of the participants reported frequent (ie, 4; n=935) or very frequent (ie, 5; n=694) engagement in preventive behaviors. Among the five preventive behaviors measured, the mean score for wearing a face mask was 3.98 (SD 1.14), washing hands had a mean of 4.42 (SD 0.76), using sanitizer had a mean of 3.62 (SD 1.12), sneezing into an elbow had a mean of 3.68 (SD 1.19), and not going out had a mean of 4.12 (SD 1.09). The distribution of each preventive behavior is shown in Figure 1.

A significant regression equation was found ($F_{14,2934}$=31.07, P<.001; R^2=0.13). Among the demographic predictors, only education level and marital status were significantly associated with preventive behaviors regarding COVID-19 (see Table 3). Individuals with higher education level (B=0.038, SE=0.012, P=.002) and those who were married (B=0.117, SE=0.030, P<.001) reported greater engagement in preventive behaviors. Among the psychosocial and behavioral predictors, perceived barriers and benefits of preventive behaviors; self-efficacy; emotional support; previous patient-centered communication with health providers; media literacy; and frequency of social media use, traditional media use, and internet news use other than social media for COVID-19 news were significantly associated with preventive behaviors. Individuals who reported fewer barriers of engaging in preventive behaviors (B=–0.055, SE=0.009, P<.001), higher benefits of the behaviors (B=0.098, SE=0.017, P<.001), greater self-efficacy (B=0.042, SE=0.014, P=.002), greater emotional support (B=0.031, SE=0.012, P=.01), greater previous patient-centered communication with health providers (B=0.029, SE=0.014, P=.04), higher media literacy (B=0.033, SE=0.016, P=.04), more frequent social media use (B=0.046, SE=0.016, P=.005), greater traditional media use (B=0.079, SE=0.013, P<.001), and greater use of internet news

channels other than social media (B=0.038, SE=0.010, *P*<.001) for COVID-19 news engaged in greater preventive behaviors at the outbreak of COVID-19.

Figure 1. Frequencies of each preventive behavior.

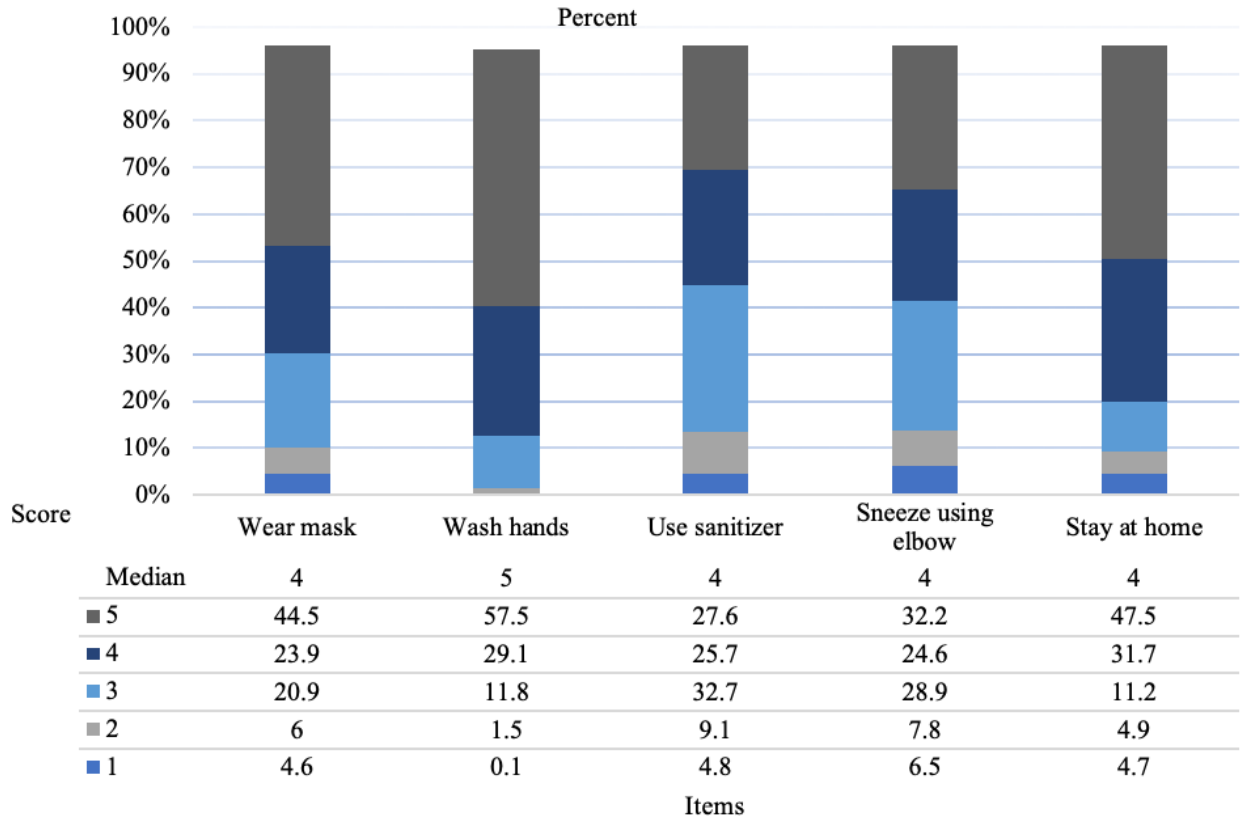


Table 3. Correlates of preventive behaviors among Chinese during the outbreak of the coronavirus disease.

Variable	Unstandardized coefficients		Standardized coefficients β	*P* value
	B	SE		
Sex	.024	.023	.018	.30
Age	.001	.002	.018	.44
Education	.038	.012	.063	.002
Marital status	.117	.030	.089	<.001
Income	.015	.008	.038	.06
Barriers	–.055	.009	–.104	<.001
Benefits	.098	.017	.109	<.001
Self-efficacy	.042	.014	.060	.002
Emotional support	.031	.012	.050	.01
Patient-centered communication	.029	.014	.042	.04
Media literacy	.033	.016	.043	.04
Social media	.046	.016	.057	.005
Traditional media	.079	.013	.129	<.001
Internet news channels other than social media	.038	.010	.074	<.001

Intervening Health Behaviors

The averaged index of IHB (mean 4.67, SD 0.77) revealed that overall participants engaged in more than four intervening behaviors out of the measured five behaviors. More than 97% (2864/2949) of the participants reported that they had ever advised others to wear masks, wash hands, and stay at home. Approximately 89.9% (2652/2949) of the participants had

advised others to use sanitizer and 84.5% (2493/2949) had suggested others to sneeze into an elbow.

A significant regression equation was found ($F_{14,2934}$=15.11, P<.001; R^2=0.07). Income and marital status were significantly associated with people's engagement in IHB (see Table 4). Particularly, individuals who were married (B=0.146, SE=0.037, P<.001) and had lower income (B=–0.040, SE=0.010, P<.001) reported greater engagement in IHB. That is, these people were more active in persuading other people to adopt protective measures against the disease.

Among the psychosocial and behavioral predictors, benefits of engaging in preventive behaviors; self-efficacy; previous patient-centered communication with health providers; media literacy; and frequency of social media use, traditional media use, and internet news channels other than social media for COVID-19 news had significant correlations with IHB. Particularly, individuals who reported greater benefits of engaging in preventive behaviors (B=0.052, SE=0.021, P=.01), greater self-efficacy (B=0.038, SE=0.017, P=.02), greater previous patient-centered communication with health providers (B=0.036, SE=0.017, P=.04), lower media literacy (B=–0.039, SE=0.019, P=.046), greater social media use (B=0.062, SE=0.020, P=.002), greater traditional media use (B=0.072, SE=0.016, P<.001), and greater use of internet news channels other than social media (B=0.039, SE=0.012, P=.002) for COVID-19 news were more engaged in performing IHBs.

Table 4. Correlates of intervening behaviors among Chinese during the outbreak of the coronavirus disease.

Variable	Unstandardized coefficients		Standardized coefficients β	*P* value
	B	SE		
Sex	0.026	0.028	.017	.35
Age	–0.001	0.002	–.018	.46
Education	0.014	0.015	.019	.36
Marital status	0.146	0.037	.095	<.001
Income	–0.040	0.010	–.086	<.001
Barriers	0.007	0.012	.011	.54
Benefits	0.052	0.021	.049	.01
Self-efficacy	0.038	0.017	.047	.02
Emotional support	0.011	0.015	.015	.47
Patient-centered communication	0.036	0.017	.045	.04
Media literacy	–0.039	0.019	–.043	.046
Social media	0.062	0.020	.066	.002
Traditional media	0.072	0.016	.099	<.001
Internet news channels other than social media	0.039	0.012	.065	.002

Discussion

Principal Results

Using a national online survey, we examined the potential predictors of two different types of epidemic-related health behaviors—the self-focused PHBs and the other-focused intervening behaviors—among Chinese people in the face of COVID-19 and pinned down key psychological determinants for Chinese public's behavioral engagement. Our findings offer valuable implications that might be applicable to other regions with similar policies or cultures in attempts to encourage the general public's adoption of precautionary measures.

Our surveyed participants' reported locations covered most of the provinces and areas in China, and most of them reported highly active adoption of PHBs to protect themselves (1629/2949, 55.3% reported frequent or very frequent engagement) and of IHBs with the goal of protecting social others (2493/2949, 84.5% reported engagement in at least one of five behaviors). The mainland Chinese people's active engagement in protective behaviors during the early outbreak of COVID-19 surely contributed to effective control of the epidemic. Mainland China's daily new confirmed cases decreased from 2022 on February 11, 2020, (toward the end of survey distribution) to 21 on July 23, 2020.

Our attempts to find key factors that facilitate or debilitate participants' propensity to take PHBs and IHBs revealed interesting findings. First, demographic variables have been found to exert differential effects on individuals' adoption of protective measures. In 2015, around 75.34% of the Chinese population received an educational level of high school and above [29]. Therefore, we used high school or less as the reference group. Education level was found to be positively associated with PHBs, suggesting that individuals with higher education were more likely to engage in preventive behaviors to protect themselves, which is consistent with previous studies [30]. However, education level was not associated with people's intervening behaviors. Regarding the null effect of education level on one's engagement in IHB, we believe the reason lies in the other-oriented nature of IHBs, that is, persuading others to follow health measures to protect themselves. Hence, we

believe this behavior is more likely to be affected by social factors or whether the individual has a significant other, such as one's marital status and their interactions with important people in their life.

Interestingly, income was negatively associated with intervening behaviors, such that people with lower income have weaker other-oriented motivation than those with higher income. Said otherwise, faced with a health crisis, poorer participants are less likely to behave in other-serving manners, suggesting their greater self-focus. This finding might be attributed to the fact that low-income groups face more and tougher challenges in the face of a health crisis due to lack of critical resources such as health insurance [31]. Marital status was found to be significantly associated with both preventive behaviors and intervening behaviors. Specifically, married individuals engaged in more preventive behaviors and more behaviors that promoted other individual's self-protection against COVID-19. This finding, in line with previous research [32,33], showcases health-related benefits of marriage. Marriage encourages adoption of healthy behaviors and motivates people to monitor, influence, and even control partners' health conditions [32].

However, among married individuals, engagement in intervening behaviors could be potentially associated with more interactions with spouses, children, parents, relatives, and even friends. Future studies should further investigate and tease apart the differential influences of those different types of interaction on people's engagement in intervening behaviors among married individuals.

Moreover, we observed interesting relationships between the examined psychological factors and participants' engagement in PHB and IHB. On the one hand, we found negative association between barriers and participants' engagement in PHB. Specifically, more perceived barriers deterred people's adoption of PHB against COVID-19. We also found that greater benefits, self-efficacy in preventing COVID-19, and emotional support had positive relationships with adoption of PHBs, which are consistent with previous research [14,33]. On the other hand, we found greater engagement in IHB among participants who perceived higher benefits of preventing coronavirus and with a higher self-efficacy. Taken together, these findings suggest that communication with the general public on COVID-19 should highlight the benefits of health behaviors, reduce perceived barriers of taking actions, and enhance self-efficacy. Additionally, it can be beneficial to advise people to seek emotional support from close others in the face of COVID-19.

Further, it is found that participants who experienced high-quality patient-doctor communication prior to COVID-19 were more active in adopting precautionary behaviors and intervening behaviors. People who had high-quality interaction with doctors tend to build trust in precautionary measures that are recommended and hence have greater motivation to comply with these recommendations. These findings shed light on the benefits of building and maintaining good patient-doctor relationships in the face of public health emergencies.

This study also investigates the effects of media literacy and media use during an outbreak of an epidemic on health behaviors. Interestingly, media literacy was positively related to preventive behaviors and negatively associated with intervening behaviors. These findings suggested that individuals with a higher ability to distinguish media messages were more likely to engage in preventive behaviors for themselves. On the contrary, individuals with a lower ability to judge a media message or news related to COVID-19 tended to intervene more toward other people's health behaviors [20,21]. Their trust of sentential or misinformation might potentially boost their intervening behaviors.

Media are usually the critical platforms to deliver news and health information, and could potentially contribute to the engagement of preventive or intervening behaviors. We found that more frequent use of both new media and traditional media for coronavirus news and information were associated with greater engagement in both preventive and intervening behaviors, indicating social media, traditional media, and internet news channels other than social media were effective platforms to disseminate COVID-19–related information to promote health behaviors. However, based on the values of coefficients, people used social media and traditional media more frequently to get information related to COVID-19 than internet news channels other than social media. Our study extends similar findings from a previous study conducted in the United States, which found general health information online was positively related to preventive behavior [34]. Our findings indicate the potential roles of social media and traditional media to deliver effective preventive campaigns related to COVID-19 [35,36].

Limitations

We acknowledge several limitations of this study and point in directions for future research. The nature of cross-sectional survey data limited the causal relationships between variables being inferred. However, our findings are largely in alignment with previous research findings regarding health behaviors. Besides, a convenient sampling approach was used. Participants were largely those who owned a social media account or who had internet access, which may undermine the generalizability of the conclusions to the whole population of China. Yet, we believe this is an issue of less significance given the urgency of the issue and the commonality of risks imposed by COVID-19 on the general public. Future studies should use a probability-based sampling method to detect health behaviors regarding COVID-19 to generalize the findings. In addition, studies examining whether the digital divide has an impact on health behaviors are warranted. Although all measurements of this study were drawn from previous studies, validated scales such as different media use should be employed in future research. Besides, it should be noted that the variance explained by our model was relatively small, which suggests potential alternative predictors. Hence, future studies should examine other possible determinants using other relevant theories. Additionally, living alone or with others could also be associated with health behaviors. Future studies should take these into consideration. Moreover, in this paper, we did not examine potential mediators or moderators of the behavioral outcomes as some theories suggest. Finally, future studies should collect longitudinal data to examine the trends of people's engagement

in health-related behaviors as the epidemic develops and the mediators and moderators of both PHBs and IHBs.

Conclusions

This study reveals that, during the early outbreak of COVID-19, Chinese people reported high engagement in preventive and intervening behaviors. Their compliance with the recommended health behaviors by the Chinese government and the WHO has alleviated the serious epidemic and resulted in a controlled situation in March 2019. This study demonstrates the associations of psychosocial factors including the perceived barriers and benefits of health behaviors, self-efficacy, emotional support, patient-centered communication, media literacy and media use for COVID-19 news, and demographic factors such as education, income, and marital status with individuals' adoption of health behaviors. Our findings have practical implications for policy makers and health organizations to design more effective health intervention programs using different media channels.

Acknowledgments

The research reported in this paper was funded by the University of South China COVID-19 epidemic prevention and control scientific research emergency project (2020-2-5), Hunan province 2020 innovative province construction special topic to combat COVID-19 epidemic emergency (2020SK3010), and the National Natural Science Foundation (No. 71802198).

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19) situation reports. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-04-04]
2. WHO Director-General's statement on IHR Emergency Committee on Novel Coronavirus (2019-nCoV). World Health Organization. 2020 Jan 30. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-statement-on-ihr-emergency-committee-on-novel-coronavirus-(2019-ncov) [accessed 2020-04-04]
3. WHO, China leaders discuss next steps in battle against coronavirus outbreak. World Health Organization. 2020 Jan 28. URL: https://www.who.int/news-room/detail/28-01-2020-who-china-leaders-discuss-next-steps-in-battle-against-coronavirus-outbreak [accessed 2020-02-26]
4. James L, Shindo N, Cutter J, Ma S, Chew S. Public health measures implemented during the SARS outbreak in Singapore, 2003. Public Health 2006 Jan;120(1):20-26. [doi: 10.1016/j.puhe.2005.10.005] [Medline: 16297416]
5. Cui B, Wang LD, Liu ZP. Knowledge of H7N9 avian influenza and intention to adopt preventive behaviours among Chinese poultry farmers: a cross-sectional study. Lancet 2016 Oct;388:S28. [doi: 10.1016/s0140-6736(16)31955-9]
6. Lau JT, Griffiths S, Choi K, Lin C. Prevalence of preventive behaviors and associated factors during early phase of the H1N1 influenza epidemic. Am J Infect Control 2010 Jun;38(5):374-380 [FREE Full text] [doi: 10.1016/j.ajic.2010.03.002] [Medline: 20569849]
7. Kasl SV, Cobb S. Health behavior, illness behavior, and sick role behavior. I. Health and illness behavior. Arch Environ Health 1966 Feb;12(2):246-266. [doi: 10.1080/00039896.1966.10664365] [Medline: 5322534]
8. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. J Epidemiol Community Health 2002 Feb;56(2):119-127 [FREE Full text] [doi: 10.1136/jech.56.2.119] [Medline: 11812811]
9. Burton LC, Newsom JT, Schulz R, Hirsch CH, German PS. Preventive health behaviors among spousal caregivers. Prev Med 1997;26(2):162-169. [doi: 10.1006/pmed.1996.0129] [Medline: 9085384]
10. Holt M, Lea T, Mao L, Kolstee J, Zablotska I, Duck T, et al. Community-level changes in condom use and uptake of HIV pre-exposure prophylaxis by gay and bisexual men in Melbourne and Sydney, Australia: results of repeated behavioural surveillance in 2013–17. Lancet HIV 2018 Aug;5(8):e448-e456. [doi: 10.1016/s2352-3018(18)30072-9]
11. Ni Mhurchu C, Te Morenga L, Tupai-Firestone R, Grey J, Jiang Y, Jull A, et al. A co-designed mHealth programme to support healthy lifestyles in Māori and Pasifika peoples in New Zealand (OL@-OR@): a cluster-randomised controlled trial. Lancet Digital Health 2019 Oct;1(6):e298-e307. [doi: 10.1016/s2589-7500(19)30130-x]
12. Seto W, Tsang D, Yung R, Ching T, Ng T, Ho M, et al. Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (SARS). Lancet 2003 May;361(9368):1519-1520. [doi: 10.1016/s0140-6736(03)13168-6]
13. Berkman LF, Glass T, Brissette I, Seeman TE. From social integration to health: Durkheim in the new millennium. Soc Sci Med 2000 Sep;51(6):843-857. [doi: 10.1016/s0277-9536(00)00065-4]
14. Myers R, Ross E, Jepson C, Wolf T, Balshem A, Millner L, et al. Modeling adherence to colorectal cancer screening. Prev Med 1994 Mar;23(2):142-151. [doi: 10.1006/pmed.1994.1020] [Medline: 8047519]
15. Dutta-Bergman MJ. Primary sources of health information: comparisons in the domain of health attitudes, health cognitions, and health behaviors. Health Commun 2004 Jul;16(3):273-288. [doi: 10.1207/s15327027hc1603_1]

16. Randolph W, Viswanath K. Lessons learned from public health mass media campaigns: marketing health in a crowded media world. Annu Rev Public Health 2004;25:419-437. [doi: 10.1146/annurev.publhealth.25.101802.123046] [Medline: 15015928]
17. De Leon E, Fuentes LW, Cohen JE. Characterizing periodic messaging interventions across health behaviors and media: systematic review. J Med Internet Res 2014 Mar 25;16(3):e93 [FREE Full text] [doi: 10.2196/jmir.2837] [Medline: 24667840]
18. Thoman E, Jolls T. Literacy for the 21st century: an overview and orientation guide to media literacy education. Center for Media Literacy. URL: https://www.medialit.org/sites/default/files/01_MLKorientation.pdf [accessed 2020-05-06]
19. Bergsma LJ, Carney ME. Effectiveness of health-promoting media literacy education: a systematic review. Health Educ Res 2008 Jun;23(3):522-542. [doi: 10.1093/her/cym084] [Medline: 18203680]
20. Jeong S, Cho H, Hwang Y. Media literacy interventions: a meta-analytic review. J Commun 2012 Jun 01;62(3):454-472 [FREE Full text] [doi: 10.1111/j.1460-2466.2012.01643.x] [Medline: 22736807]
21. Vahedi Z, Sibalis A, Sutherland JE. Are media literacy interventions effective at changing attitudes and intentions towards risky health behaviors in adolescents? A meta-analytic review. J Adolesc 2018 Aug;67:140-152. [doi: 10.1016/j.adolescence.2018.06.007] [Medline: 29957493]
22. Lau JT, Kim JH, Tsui HY, Griffiths S. Anticipated and current preventive behaviors in response to an anticipated human-to-human H5N1 epidemic in the Hong Kong Chinese general population. BMC Infect Dis 2007 Mar 15;7:18 [FREE Full text] [doi: 10.1186/1471-2334-7-18] [Medline: 17359545]
23. Coronavirus disease (COVID-19) advice for the public. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [accessed 2020-06-27]
24. Gutierrez J, Long JA. Reliability and validity of diabetes specific Health Beliefs Model scales in patients with diabetes and serious mental illness. Diabetes Res Clin Pract 2011 Jun;92(3):342-347 [FREE Full text] [doi: 10.1016/j.diabres.2011.02.018] [Medline: 21411173]
25. Health Information National Trends Survey. URL: http://hints.cancer.gov [accessed 2020-06-27]
26. Epstein R, Street R. Patient-centered communication in cancer care: promoting healing and reducing suffering. Division of Cancer Control and Population Sciences. 2007. URL: https://cancercontrol.cancer.gov/brp/docs/pcc_monograph.pdf [accessed 2020-04-04]
27. Austin EW, Pinkleton BE, Radanielina-Hita ML, Ran W. The role of parents' critical thinking about media in shaping expectancies, efficacy and nutrition behaviors for families. Health Commun 2015;30(12):1256-1268. [doi: 10.1080/10410236.2014.930550] [Medline: 25616579]
28. Howitt D, Cramer D. Introduction to SPSS in Psychology: For Version 16 and Earlier. London, United Kingdom: Pearson Education; 2008.
29. China Bureau of Statistics. The main data bulletin of the 1% national population sampling survey in 2015. National Bureau of Statistics. 2016 Apr. URL: http://www.stats.gov.cn/tjsj/zxfb/201604/t20160420_1346151.html [accessed 2020-04-04]
30. Holtzman D, Bland SD, Lansky A, Mack KA. HIV-related behaviors and perceptions among adults in 25 states: 1997 Behavioral Risk Factor Surveillance System. Am J Public Health 2001 Nov;91(11):1882-1888. [doi: 10.2105/ajph.91.11.1882] [Medline: 11684620]
31. Kanchanachitra C, Lindelow M, Johnston T, Hanvoravongchai P, Lorenzo FM, Huong NL, et al. Human resources for health in southeast Asia: shortages, distributional challenges, and international trade in health services. Lancet 2011 Feb;377(9767):769-781. [doi: 10.1016/s0140-6736(10)62035-1]
32. Schone BS, Weinick RM. Health-related behaviors and the benefits of marriage for elderly persons. Gerontologist 1998 Oct;38(5):618-627. [doi: 10.1093/geront/38.5.618] [Medline: 9803650]
33. Li S, Feng B, Liao W, Pan W. Internet use, risk awareness, and demographic characteristics associated with engagement in preventive behaviors and testing: cross-sectional survey on COVID-19 in the United States. J Med Internet Res 2020 Jun 16;22(6):e19782 [FREE Full text] [doi: 10.2196/19782] [Medline: 32501801]
34. Squiers L, Peinado S, Berkman N, Boudewyns V, McCormack L. The health literacy skills framework. J Health Commun 2012;17 Suppl 3:30-54. [doi: 10.1080/10810730.2012.713442] [Medline: 23030560]
35. Tao Z, Chu G, McGrath C, Hua F, Leung Y, Yang W, et al. Nature and diffusion of COVID-19-related oral health information on Chinese social media: analysis of Tweets on Weibo. J Med Internet Res 2020 Jun 15;22(6):e19981 [FREE Full text] [doi: 10.2196/19981] [Medline: 32501808]
36. Wahbeh A, Nasralah T, Al-Ramahi M, El-Gayar O. Mining physicians' opinions on social media to obtain insights into COVID-19: mixed methods analysis. JMIR Public Health Surveill 2020 Jun 18;6(2):e19276 [FREE Full text] [doi: 10.2196/19276] [Medline: 32421686]

Abbreviations

COVID-19: coronavirus disease
IHB: intervening health behavior
PHB: preventive health behavior
PHM: Preventive Health Model

WHO: World Health Organization

Edited by G Eysenbach; submitted 11.05.20; peer-reviewed by L Kayser, J Sapp, N Guo; comments to author 12.06.20; revised version received 27.06.20; accepted 26.07.20; published 21.08.20.

Please cite as:
Niu Z, Wang T, Hu P, Mei J, Tang Z
Chinese Public's Engagement in Preventive and Intervening Health Behaviors During the Early Breakout of COVID-19: Cross-Sectional Study
J Med Internet Res 2020;22(8):e19995
URL: http://www.jmir.org/2020/8/e19995/
doi:10.2196/19995
PMID:32716897

Original Paper

Dynamics and Development of the COVID-19 Epidemic in the United States: A Compartmental Model Enhanced With Deep Learning Techniques

Qi Deng[1,2,3], BSc, MBA, MSEE, DBA

[1]College of Economics and Management, China-Africa International Business School, Zhejiang Normal University, Jinhua, China

[2]Cofintelligence Financial Technology Ltd, Shanghai, China

[3]Cofintelligence Financial Technology Ltd, Hong Kong, Hong Kong

Corresponding Author:
Qi Deng, BSc, MBA, MSEE, DBA
College of Economics and Management
China-Africa International Business School
Zhejiang Normal University
688 Yingbin Avenue
Jinhua
China
Phone: 86 13811402549
Email: qi.deng@zjnu.edu.cn

Abstract

Background: Compartmental models dominate epidemic modeling. Transmission parameters between compartments are typically estimated through stochastic parameterization processes that depends on detailed statistics of transmission characteristics, which are economically and resource-wise expensive to collect.

Objective: We aim to apply deep learning techniques as a lower data dependency alternative to estimate transmission parameters of a customized compartmental model, for the purpose of simulating the dynamics of the US coronavirus disease (COVID-19) epidemic and projecting its further development.

Methods: We constructed a compartmental model and developed a multistep deep learning methodology to estimate the model's transmission parameters. We then fed the estimated transmission parameters to the model to predict development of the US COVID-19 epidemic for 35 and 42 days. Epidemics are considered suppressed when the basic reproduction number (R_0) is less than 1.

Results: The deep learning–enhanced compartmental model predicts that R_0 will fall to <1 around August 17-19, 2020, at which point the epidemic will effectively start to die out, and that the US "infected" population will peak around August 16-18, 2020, at 3,228,574 to 3,308,911 individual cases. The model also predicted that the number of accumulative confirmed cases will cross the 5 million mark around August 7, 2020.

Conclusions: Current compartmental models require stochastic parameterization to estimate the transmission parameters. These models' effectiveness depends upon detailed statistics on transmission characteristics. As an alternative, deep learning techniques are effective in estimating these stochastic parameters with greatly reduced dependency on data particularity.

(J Med Internet Res 2020;22(8):e21173) doi:10.2196/21173

KEYWORDS

epidemiology; COVID-19; compartmental models; deep learning; model; modeling; transmission; estimation; virus; simulate

Introduction

The coronavirus disease (COVID-19) pathogen that has ravaged China, Europe, and the United States since December 2019 is a member of the coronavirus family, which also includes the severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome–related coronavirus (MERS-CoV). In the United States, as of July 31, 2020, there have been 4,562,038 confirmed cases and 153,314 deaths of COVID-19.

The COVID-19 pandemic is still in progress, and most of the noticeable early research is descriptive in nature, focusing on reported cases to establish the baseline demographic parameters for the disease such as age, gender, health, and medical conditions in addition to the disease's clinical manifestations, in a Chinese context. These studies include reports on demographic characteristics, epidemiological and clinical characteristics, exposure and travel history to the epicenter, and illness timelines of laboratory-confirmed cases [1-5] as well as epidemiological information on patients from social networks and local, national, and international health authorities [6]. The spread of SARS-CoV-2 outside China (eg, Iceland) is also analyzed [7], albeit to a limited extent. Concerned about the worsening situation in New York City, researchers have characterized information on the first 393 consecutive patients with COVID-19 admitted to 2 hospitals in the city [8].

Some stage-specific studies on patients with COVID-19 have also been carried out, including a single-centered, retrospective study on critically ill adult patients in Wuhan, China [9] and a retrospective, multicenter study on adult laboratory-confirmed inpatients (≥18 years of age) from 2 Wuhan hospitals, who have been discharged or have died [10].

The aim of this paper is to establish a class of extended COVID-19 compartmental models, for which the transmission parameters are estimated by a multistep, multivariate deep learning methodology.

Methods

COVID-19 Epidemic Modeling

There have been attempts to model the COVID-19 epidemic dynamics. These studies add a worldwide mobile dimension, reflecting a higher level of mobility and globalization in 2020 than in 2003 (SARS) and even 2013 (MERS). The SEIR (Susceptible–Exposed–Infectious–Recovered) model is used to infer the basic reproduction ratio and simulate the Wuhan epidemic [11]; it considers domestic and international air travel to and from Wuhan to other cities to forecast the national and global spread of the virus. More sophisticated models have also been developed to correlate risk levels of foreign countries with their travel exposure to China [12,13], including a stochastic dual-SEIR approach on both the Wuhan population and international travelers, to estimate how transmission varied over time from Wuhan to international destinations [13]. Simulations on the international spread of the COVID-19 after the start of the travel ban from Wuhan on January 23, 2020, have also been conducted [14], which apply the Global Epidemic and Mobility Model to a multitude of Chinese and international cities, and a SEIR variety (SLIR, Susceptible–Latent–Infectious–Recovered) to project the impact of human-to-human transmissions. To simulate the transmission mechanism itself, a Bats-Hosts-Reservoir-People network is developed to simulate potential transmission from the infection sources (ie, bats) to humans [15].

Since March 2020, with the COVID-19 outbreak winding down in China, researchers have dedicated more efforts to analyzing the effectiveness of containment measures. Mobility and travel history data from Wuhan are used to ascertain the impact of the drastic control measures implemented in China [16]. A study investigated the spread and control of COVID-19 among Chinese cities, using data on human movements and public health interventions [17]. Using contact data for Wuhan and Shanghai and contact tracing information from Hunan Province, a group of researchers built a transmission model to study the impact of social distancing and school closure [18].

Theoretical Foundation

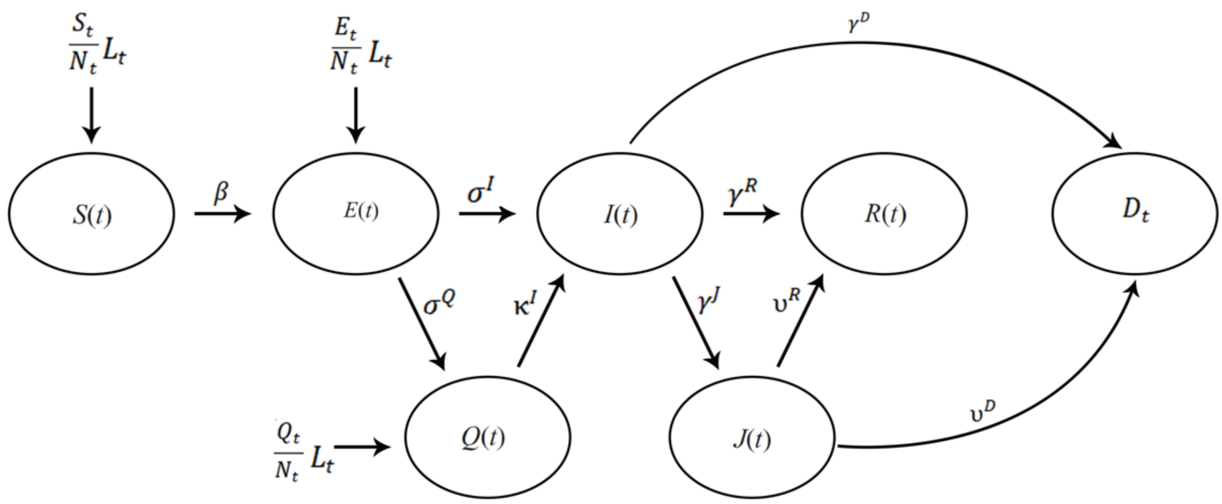
Compartmental models dominate epidemic modeling on COVID-19 epidemics (and previous coronavirus outbreaks), and they require detailed statistics on transmission characteristics to estimate the stochastic transmission parameters between compartments. Essentially, these models correlate factors such as geographic distances and contact intensities among heterogeneous subpopulations with gradient probability decay. Technically, transmission parameterization applies Bayesian inference methods such as Marcov Chain Monte Carlo or Gillespie algorithm [19] simulations to form probability density functions on a cross-section in order to estimate parameters for each timestep of a multivariate time series construct. These detailed statistics on transmission characteristics are economically and resource-wise expensive to collect.

We are particularly interested in extended compartmental models that cover multiple interconnected and heterogeneous subpopulations [10,15,20]. There are also some pure time series analyses on epidemic dynamics outside of mainstream compartmental modeling, for example, the AutoRegressive Integrated Moving Average approach [21] that is typically found in financial applications. Such analyses provide another perspective.

We developed a multistep, multivariate deep learning methodology to estimate the transmission parameters. We then fed these estimated transmission parameters to a customized compartmental model to predict the development of the US COVID-19 epidemic.

We established a SEIR-variety discrete time series on a daily interval as the theoretical foundation for a deep learning–enhanced compartment model. We started with the construction of a so-called SEIRQJD (SEIR-Quarantined-Isolated-Deceased) model (Figure 1).

Figure 1. The SEIRQJD (Susceptible–Exposed–Infectious–Recovered–Quarantined–Isolated–Deceased) Model. E: Exposed; Q: Quarantined; I: Infectious; D: Deceased; S: Susceptible; I: Infectious; J: Isolated; R: Recovered. The transmission parameters (Greek letters) are - β: from Susceptible (S) to Exposed (E) if Exposed (E) is reported directly, or Susceptible (S) to Infectious (I) if Exposed (E) is not reported directly; σ^I, σ^Q: from Exposed (E) to Infectious (I) and Quarantined (Q), respectively; κ^I: from Quarantined (Q) to Infectious (I); γ^J, γ^R, γ^D: from Infectious (I) to Isolated (J), Recovered (R) and Deceased (D), respectively; υ^R, υ^D: from Isolated (J) to Recovered (R) and Deceased (D), respectively.



We used the US COVID-19 epidemic datasets from John Hopkins University Center for Systems Science and Engineering (JHU CSSE) Github COVID-19 data depository, which does not include directly Exposed (E) and Quarantined (Q) data, and therefore, we set all transmission parameters to and from the “E” and “Q” compartments (σ^I, σ^Q, κ^I) to 0. Furthermore, the datasets assume that all deaths arise from the isolated population (J); thus, we also set the transmission parameter from Infectious (I) to Deceased (D), γ^D, to 0. We then simplified the SEIRJD model to a SIRJD (Susceptible–Infectious–Recovered–Isolated–Deceased) construct, in which a population is grouped into 5 compartments:

1. Susceptible (S): The susceptible population arises at a percentage of a net influx of individuals (L_t).
2. Infectious (I): The infectious individuals are symptomatic, come from the Susceptible compartment, and further progress into the Isolated or Recovered compartments.
3. Isolated (J): The isolated individuals have developed clinical symptoms and have been isolated by hospitalization or other means of separation. They come from the Infectious compartment and progress into the Recovered or Deceased compartments
4. Recovered (R): The recovered individuals come from Infectious and Isolated compartments and acquire lasting immunity (there is no contradiction against this assumption yet).
5. Deceased (D): The deceased cases come from the Infectious and Isolated compartments.

The SIRJD model has a daily (Δt=1) multivariate time series construct given by the follow matrix form:

or

The Greek letters in the time series are transmission parameters defined in the state diagram in Figure 1. Essentially, all these parameters are stochastic.

Since we need to estimate the transmission parameters, we can rewrite and rearrange Equations (1) and (2) to the following matrix representation:

or

Data

We collected the following US COVID-19 datasets from the JHU CSSE data depository [22]:

1. Dataset 1: The JHU CSSE updates daily records (confirmed, active, dead, recovered, hospitalized, etc) from April 12, 2020. We used these detailed case data to construct the compartmental model (Multimedia Appendix 1).
2. Dataset 2: The JHU CSSE updates 2 time series on a daily basis. One tracks the confirmed cases and the other tracks the dead cases, both starting from January 22, 2020. We used the confirmed/dead cases as training data for deep learning (Multimedia Appendix 2).

The JHU CSSE dataset has an almost precise period of 7 days (±1 day), indicating that a majority of the reporting agencies in the country choose to update their respective statistics on a

weekly, fixed-calendar interval. We ran a 7-day moving average on the dataset to smooth out this "unnatural" data seasonality.

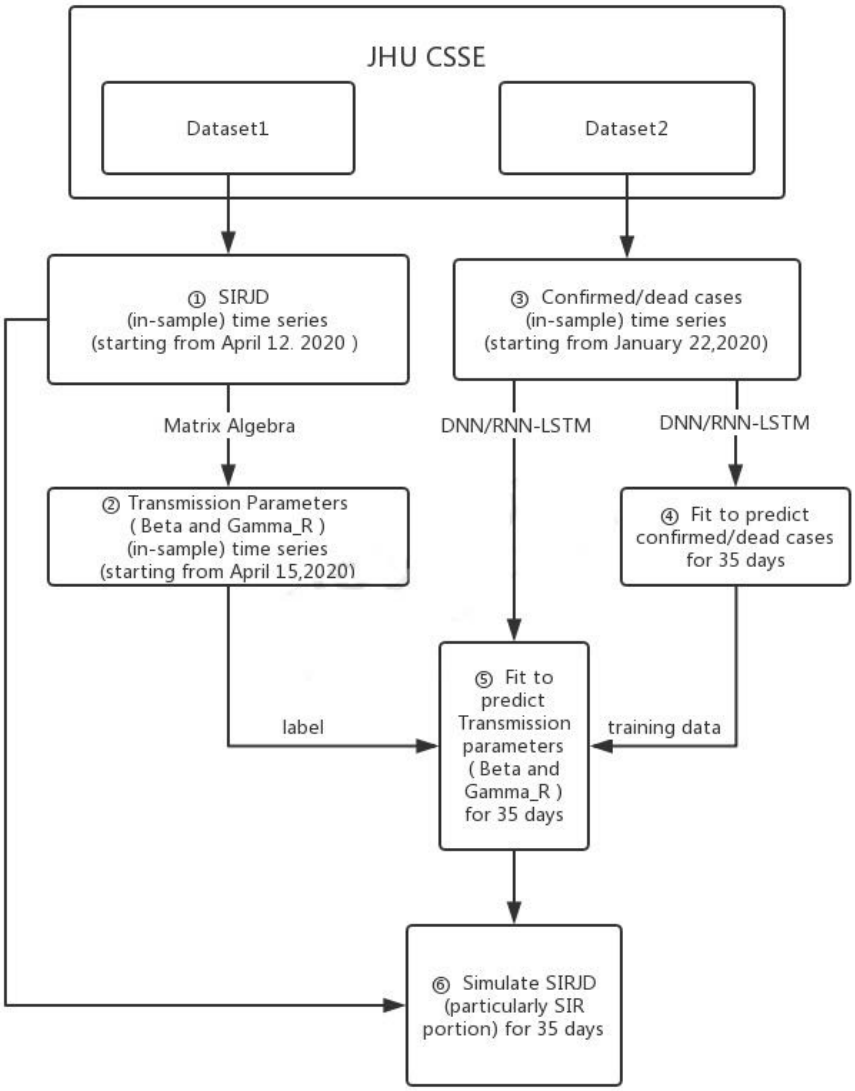
Methodology

We then conducted the following step-by-step operations to model the US epidemic:

1. We constructed an in-sample SIRJD time series starting from April 12, 2020, with Dataset 1.
2. We used the in-sample SIRJD time series constructed in Step 1 to come up with an in-sample time series for the 2 most critical daily transmission parameters (β and γ^R).
3. We constructed a confirmed/dead-case time series starting from January 22, 2020 (in-sample time series), with Dataset 2.
4. We applied 2 deep learning approaches—the standard deep neural networks (DNN) and the advanced recurrent neural networks–long short-term memory (RNN-LSTM)—to fit the confirmed/dead in-sample time series from Step 3 and predict the further development of confirmed/dead cases for 35 and 42 days (out-of-sample time series).
5. We use the confirmed/dead in-sample time series from Step 3 as training data and the in-sample β and γ^R time series from Step 2 as training label. We then applied the DNN and RNN-LSTM techniques to predict β and γ^R for 35 and 42 days (out-of-sample time series).
6. Finally, we used the predicted (out-of-sample) transmission parameters (β and γ^R) from Step 5 to simulate 35- and 42-day progressions (out-of-sample time series) of the SIRJD model (particularly the SIR portion) in a recursive manner, starting with the data point of the last timestep from the in-sample SIRJD time series from Step 1.

Figure 2 presents a flowchart to illustrate the dataset and methodology.

Figure 2. Flowchart of the dataset and methodology. JHU CSSE: John Hopkins University Center for Systems Science and Engineering; DNN: deep neural networks; RNN-LSTM: recurrent neural networks–long short-term memory; SIR: Susceptible–Infectious–Recovered; SIRJD: Susceptible–Infectious–Recovered–Isolated–Deceased.



Results

The results based on data up to July 31, 2020, are illustrated in Figures 3-6 for the 35-day forecast and Figures 7-10 for the 42-day forecast.

In Figure 3 (35-day forecast), the DNN method predicts that on August 19, 2020, the "Infected-to-Recovered" transmission parameter γ^R will rise and stay above the "Susceptible-to-Infected" transmission parameter β. This means that the value of the basic reproduction rate, R_0, will fall to <1 and that the spread of COVID-19 in the United States will effectively end on that day. In Figure 4 (35-day forecast), the RNN-LSTM method gives a slightly more aggressive prediction that γ^R will overtake β on August 17, 2020. Thus, with the 35-day forecast, we predict that the tide of the US epidemic will turn around the August 17-19, 2020, timeframe.

Figure 3. Transmission parameter estimations (deep neural networks) for 35 days. Beta is the "Susceptible-to-Infected" transmission parameter (β) and Gamma_R is the "Infected-to-Recovered" transmission parameter (γ^R) for the in-sample (observed) data. Beta_fo is the forecasted β and Gamma_R_fo is the forecasted γ^R for the out-of-sample (forecasted) data.

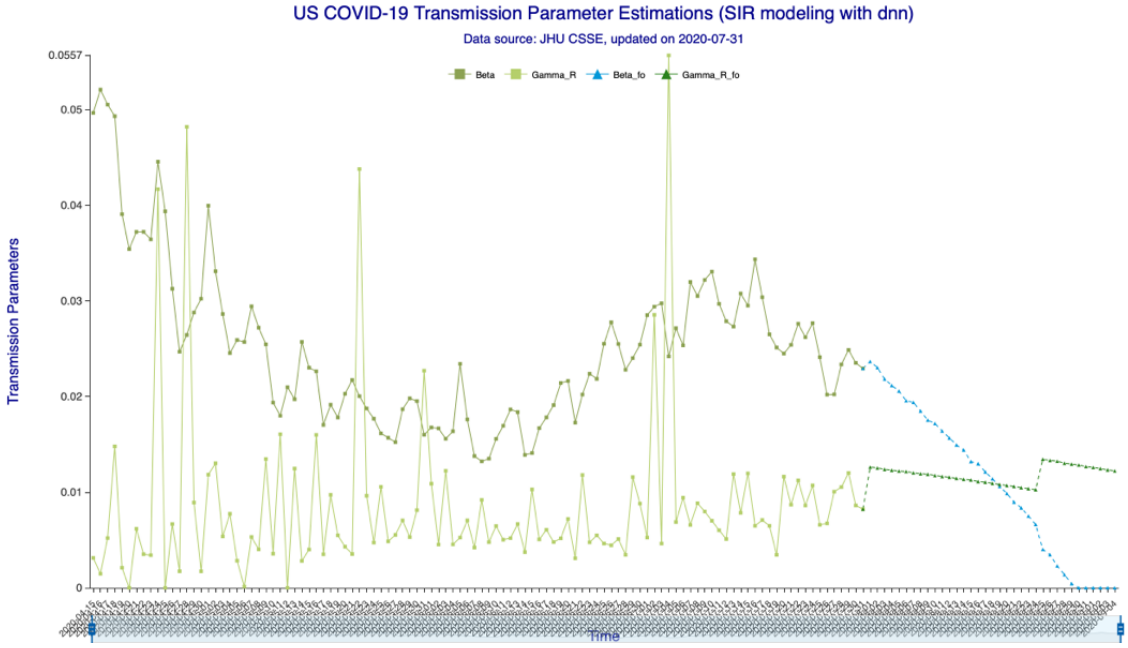
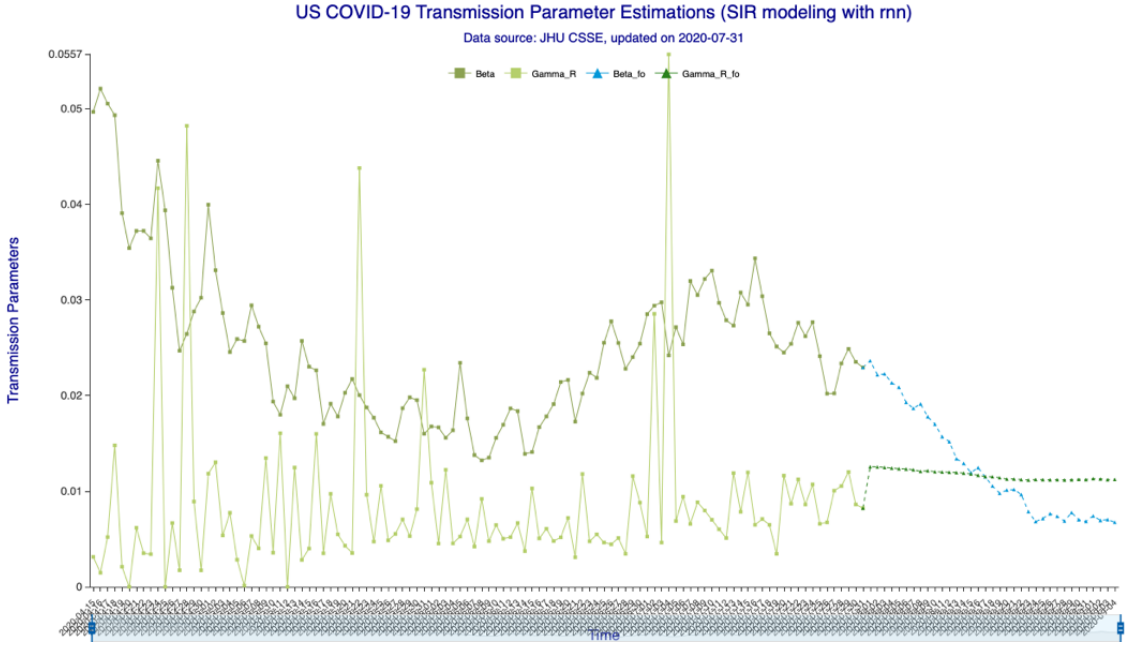


Figure 4. Transmission parameter estimations (recurrent neural networks–long short-term memory) for 35 days. Beta is the "Susceptible-to-Infected" transmission parameter (β) and Gamma_R is the "Infected-to-Recovered" transmission parameter (γ^R) for the in-sample (observed) data. Beta_fo is the forecasted β and Gamma_R_fo is the forecasted γ^R for the out-of-sample (forecasted) data.



In Figure 5 (35-day forecast), the DNN method predicts that the US "Infected" population will peak on August 18, 2020, at 3,267,907 individual cases. In Figure 6 (35-day forecast), the RNN-LSTM method predicts that the US "Infected" population will peak on August 16, 2020, at 3,228,574 individual cases. For the 35-day forecast, the deep learning methods predict that the number of accumulative confirmed cases will cross the 5 million mark on August 7, 2020, at 5,007,479 cases by DNN (Figure 5) and at 5,002,100 cases by RNN-LSTM (Figure 6).

Figure 5. SIR (Susceptible–Infectious–Recovered) model forecasting (deep neural networks) for 35 days. Susceptible, Infected, Recovered, and Dead are in-sample compartmental model data, and Confirmed_cal is the in-sample number of confirmed cases. Susceptible_fo, Infected_fo, Recovered_fo, Dead_fo, and Confirmed_cal_fo are their out-of-sample (forecasted) counterparts. The right y-axis is for Susceptible/Susceptible_fo, while the left y-axis is for all others. The right y-axis is needed for scaling purpose, as Susceptible/Susceptible_fo are derived from the total population.

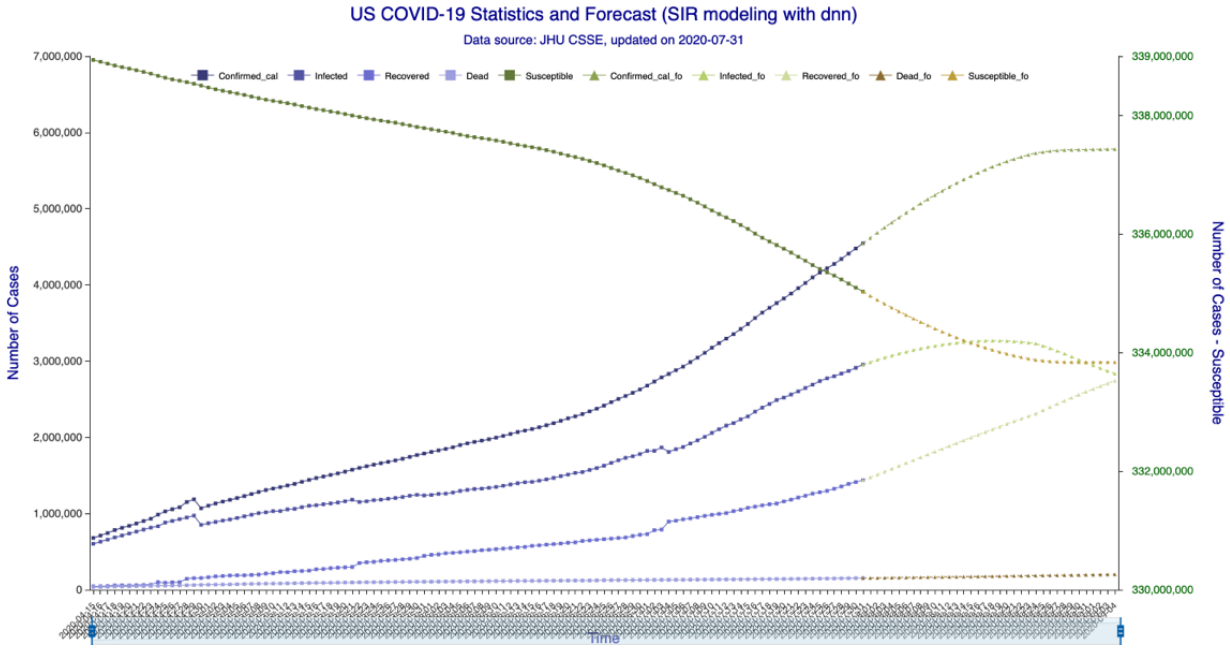
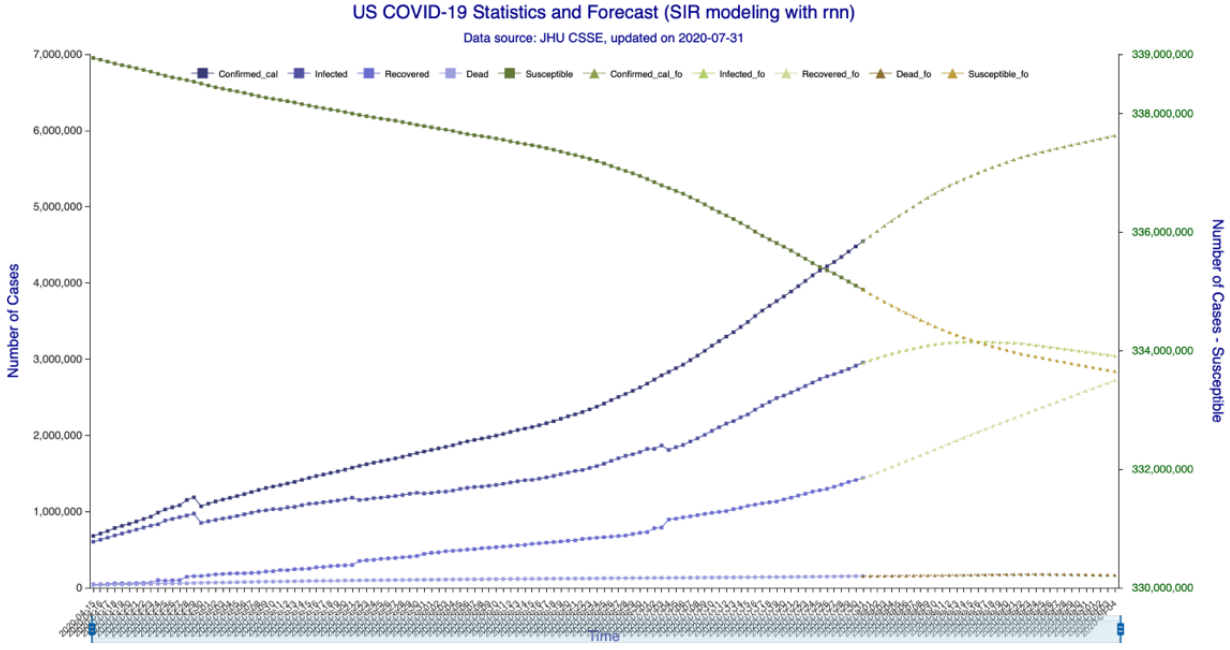


Figure 6. SIR (Susceptible–Infectious–Recovered) model forecasting (recurrent neural networks–long short-term memory) for 35 days. Susceptible, Infected, Recovered, and Dead are in-sample compartmental model data, and Confirmed_cal is the in-sample number of confirmed cases. Susceptible_fo, Infected_fo, Recovered_fo, Dead_fo, and Confirmed_cal_fo are their out-of-sample (forecasted) counterparts. The right y-axis is for Susceptible/Susceptible_fo, while the left y-axis is for all others. The right y-axis is needed for scaling purpose, as Susceptible/Susceptible_fo are derived from the total population.



In Figure 7 (42-day forecast), the DNN method also predicts (same as 35-day forecast) that γ^R will overtake β on August 19, 2020. In Figure 8 (42-day forecast), the RNN-LSTM method gives exactly the same prediction, that R_0 will fall to <1 on August 19, 2020.

Figure 7. Transmission parameter estimations (deep neural networks) for 42 days. Beta is the "Susceptible-to-Infected" transmission parameter (β) and Gamma_R is the "Infected-to-Recovered" transmission parameter (γ^R) for the in-sample (observed) data. Beta_fo is the forecasted β and Gamma_R_fo is the forecasted γ^R for the out-of-sample (forecasted) data.

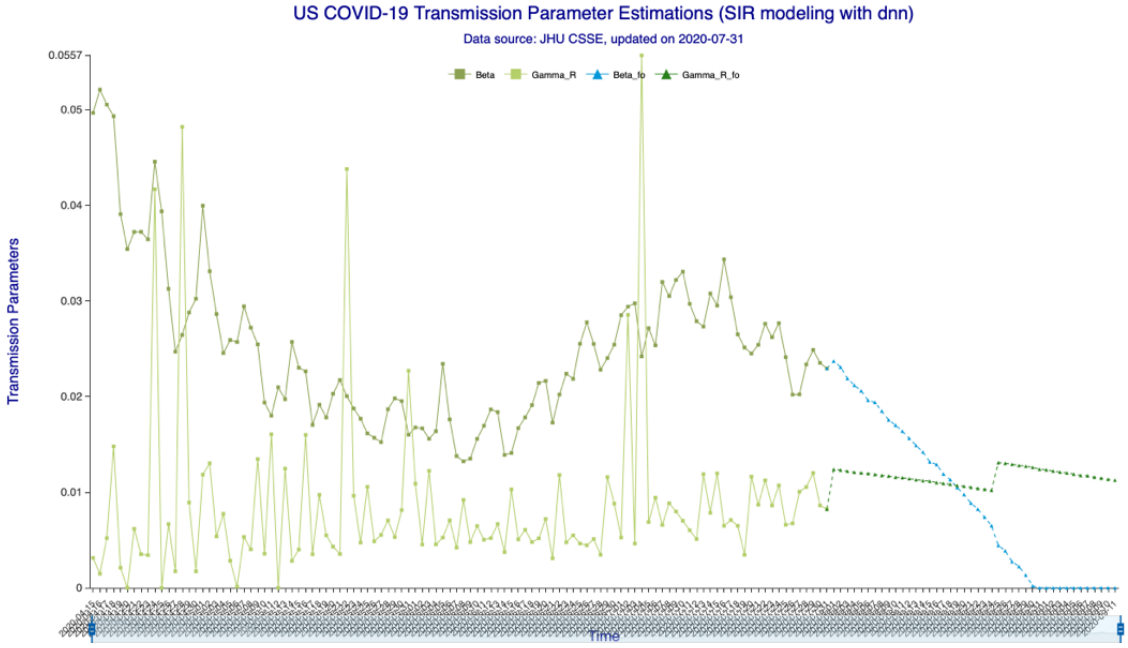
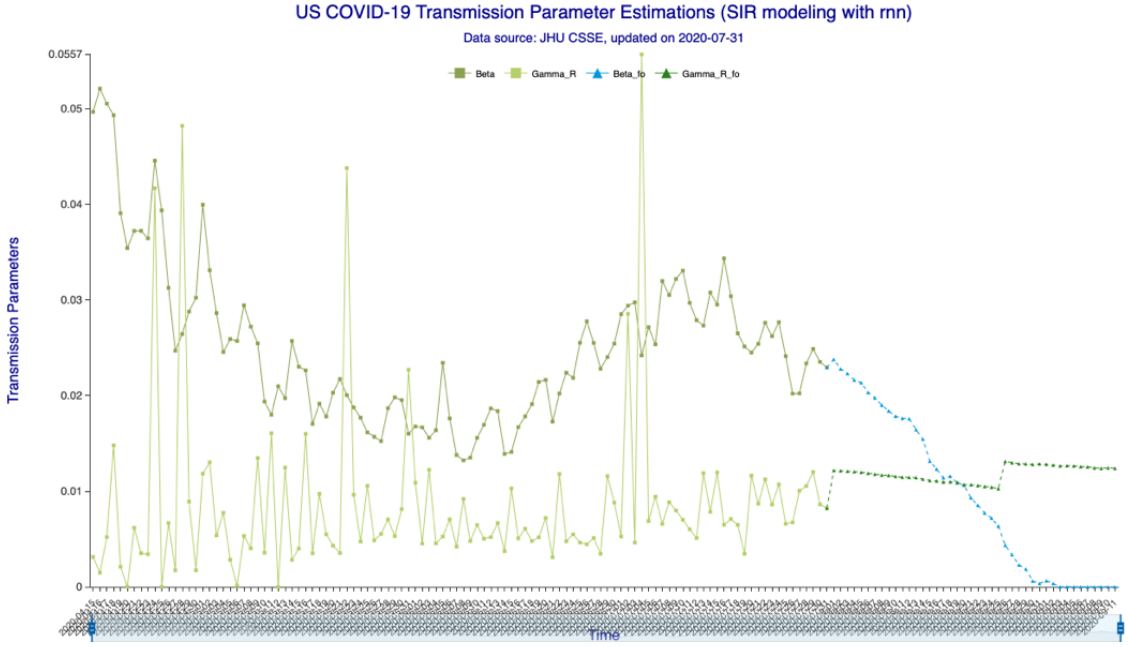


Figure 8. Transmission parameter estimations (recurrent neural networks–long short-term memory) for 42 days. Beta is the "Susceptible-to-Infected" transmission parameter (β) and Gamma_R is the "Infected-to-Recovered" transmission parameter (γ^R) for the in-sample (observed) data. Beta_fo is the forecasted β and Gamma_R_fo is the forecasted γ^R for the out-of-sample (forecasted) data.



In Figure 9 (42-day forecast), the DNN method predicts that the US "Infected" population will peak on August 18, 2020, at 3,275,304 individual cases. In Figure 10 (42-day forecast), the RNN-LSTM method predicts that the US "Infected" population will peak on August 18, 2020, at 3,308,911 individual cases. For the 42-day forecast, the deep learning methods predict that the number of accumulative confirmed cases will cross the 5 million mark on August 7, 2020, at 5,008,504 individual cases by DNN (Figure 9) and 5,014,608 individual cases by RNN-LSTM (Figure 10), which are consistent with the 35-day forecasts.

Figure 9. SIR (Susceptible–Infectious–Recovered) model forecasting (deep neural networks) for 42 days. Susceptible, Infected, Recovered, and Dead are in-sample compartmental model data, and Confirmed_cal is the in-sample number of confirmed cases. Susceptible_fo, Infected_fo, Recovered_fo, Dead_fo, and Confirmed_cal_fo are their out-of-sample (forecasted) counterparts. The right y-axis is for Susceptible/Susceptible_fo, while the left y-axis is for all others. The right y-axis is needed for scaling purpose, as Susceptible/Susceptible_fo are derived from the total population.

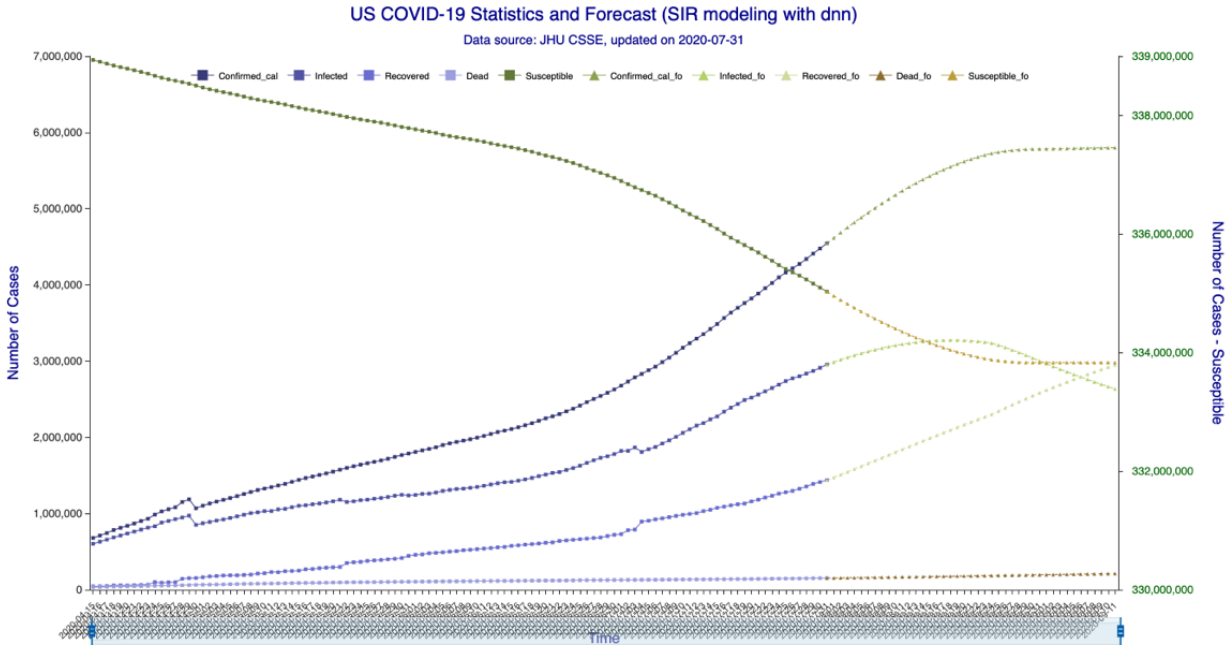
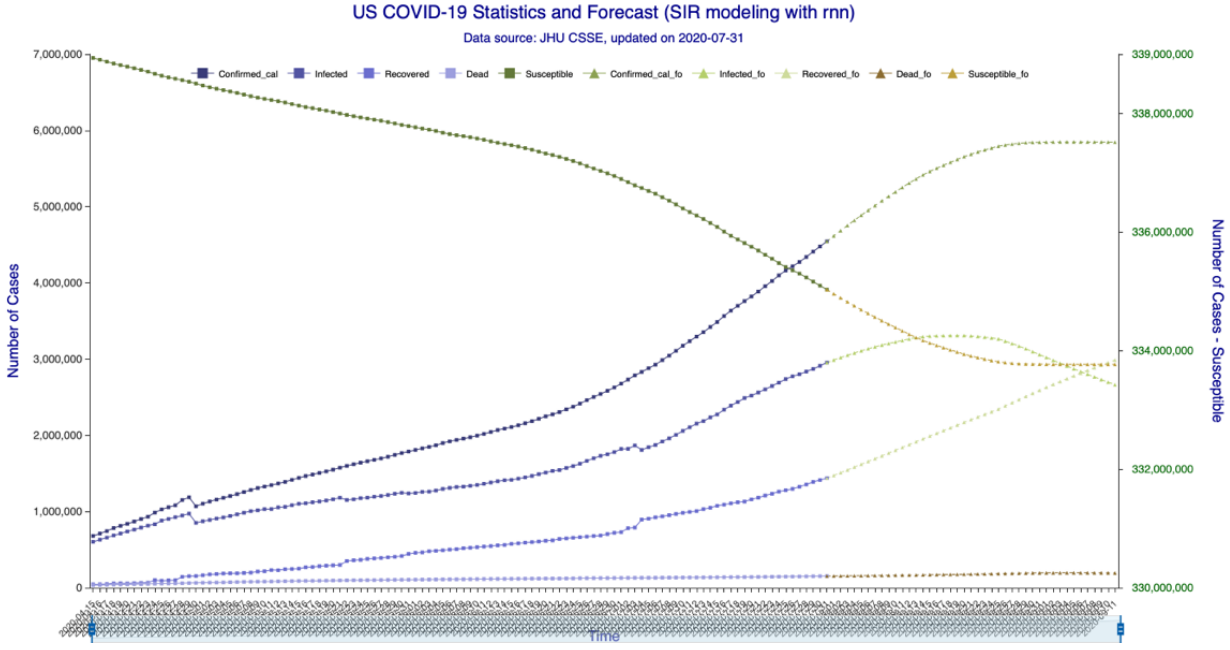


Figure 10. SIR (Susceptible–Infectious–Recovered) model forecasting (recurrent neural networks–long short-term memory) for 42 days. Susceptible, Infected, Recovered, and Dead are in-sample compartmental model data, and Confirmed_cal is the in-sample number of confirmed cases. Susceptible_fo, Infected_fo, Recovered_fo, Dead_fo, and Confirmed_cal_fo are their out-of-sample (forecasted) counterparts. The right y-axis is for Susceptible/Susceptible_fo, while the left y-axis is for all others. The right y-axis is needed for scaling purpose, as Susceptible/Susceptible_fo are derived from the total population.



Discussion

In this study, we applied DNN and RNN-LSTM techniques to estimate the stochastic transmission parameters for an SIRJD model with a discrete time series construct. We then used the SIRJD model to forecast further development of the US COVID-19 epidemic.

We used two US COVID-19 datasets from the JHU CSSE data depository. The first dataset includes detailed daily records (confirmed, active, dead, recovered, hospitalized, etc) starting from April 12, 2020, from which we constructed the SIRJD model. The second dataset includes time series tracked confirmed and dead cases starting from January 22, 2020, which we used to construct training data for deep learning. The JHU CSSE data have an almost precise period of 7 days (±1 day) that masks the true epidemic dynamics; thus, we ran a 7-day

moving average on the dataset to smooth out this data seasonality.

We then applied DNN and RNN-LSTM deep learning techniques to fit the confirmed/dead series to predict the further development of confirmed/dead cases as well as to predict the "Susceptible-to-Infected" and "Infected-to-Recovered" transmission parameters (β and γ^R) for 35 and 42 days. Finally, we used the predicted transmission parameters (β and γ^R) to simulate the epidemic progression for 35 and 42 days.

With data up to July 31, 2020, the deep learning implementations predicted that the basic reproduction rate (R_0) will fall to <1 around August 17-19, 2020, for the 35-day forecast and around August 19, 2020, for the 42-day forecast, at which point the spread of the coronavirus will effectively start to die out.

Implementations for the 35-day forecast predict that the US "Infected" population will peak around August 16-18, 2020, at 3,228,574 to 3,267,907 individual cases. The implementations for the 42-day forecast predict that the peak will occur on August 18, 2020, at 3,275,304 to 3,308,911 individual cases. All implementations indicate that the number of accumulative confirmed cases will cross the 5 million mark around August 7, 2020.

The 42-day forecasts provide a wider range of time and numbers than the 35-day forecasts, because for the same training data size, a longer forecast produces wider probability distributions.

With the introduction of the deep learning–enhanced compartmental model, we provide an effective and easy-to-implement alternative to prevailing stochastic parameterization, which estimates transmission parameters through probability likelihood maximization or Marcov Chain Monte Carlo simulation. The effectiveness of the prevalent approach depends upon detailed statistics on transmission characteristics among heterogeneous subpopulations, and such statistics are economically and resource-wise expensive. On the other hand, deep learning techniques uncover hidden interconnections among seemly less-related data, reducing the prediction's dependency on data particularity. Future research on the usefulness of deep learning in epidemic modeling can further enhance its forecasting power.

Acknowledgments

We thank Ms Liu Chang and Mr Liu Shuigeng from Cofintelligence Fintech Co, Ltd (Hong Kong and Shanghai), for data collection and formatting.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Dataset_1.
[XLSX File (Microsoft Excel File), 24 KB - jmir_v22i8e21173_app1.xlsx]

Multimedia Appendix 2
Dataset_2.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v22i8e21173_app2.xlsx]

References

1. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet 2020 Feb;395(10223):507-513. [doi: 10.1016/S0140-6736(20)30211-7]
2. Eastin C, Eastin T. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr;58(4):711-712. [doi: 10.1016/j.jemermed.2020.04.004]
3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/S0140-6736(20)30183-5]
4. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
5. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. Lancet Infect Dis 2020 Apr;20(4):425-434. [doi: 10.1016/s1473-3099(20)30086-4]
6. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. The Lancet Digit Health 2020 Apr;2(4):e201-e208. [doi: 10.1016/s2589-7500(20)30026-1]
7. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. medRxiv. Preprint posted online 2020 Mar 31. [doi: 10.1101/2020.03.26.20044446]

8. Goyal P, Choi J, Pinheiro L, Schenck EJ, Chen R, Jabri A, et al. Clinical Characteristics of Covid-19 in New York City. N Engl J Med 2020 Jun 11;382(24):2372-2374 [FREE Full text] [doi: 10.1056/NEJMc2010419] [Medline: 32302078]
9. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020 May;8(5):475-481. [doi: 10.1016/s2213-2600(20)30079-5]
10. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. Science 2020 Jun 26;368(6498):1481-1486 [FREE Full text] [doi: 10.1126/science.abb8001] [Medline: 32350060]
11. Wu J, Leung K, Leung G. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet 2020 Feb;395(10225):689-697. [doi: 10.1016/s0140-6736(20)30260-9]
12. Gilbert M, Pullano G, Pinotti F, Valdano E, Poletto C, Boëlle P, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. The Lancet 2020 Mar;395(10227):871-877. [doi: 10.1016/s0140-6736(20)30411-6]
13. Kucharski A, Russell T, Diamond C, Liu Y, CMMID nCoV working group, Edmunds J, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. The Lancet 2020;3099(20):0144. [doi: 10.1101/2020.01.31.20019901]
14. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
15. Chen T, Rui J, Wang Q, Zhao Z, Cui J, Yin L. A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. Infect Dis Poverty 2020 Feb 28;9(1):24 [FREE Full text] [doi: 10.1186/s40249-020-00640-3] [Medline: 32111262]
16. Kraemer MUG, Yang C, Gutierrez B, Wu C, Klein B, Pigott DM, Open COVID-19 Data Working Group, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 2020 May 01;368(6490):493-497 [FREE Full text] [doi: 10.1126/science.abb4218] [Medline: 32213647]
17. Tian H, Liu Y, Li Y, Wu C, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science 2020 May 08;368(6491):638-642 [FREE Full text] [doi: 10.1126/science.abb6105] [Medline: 32234804]
18. Zhang J, Lou J, Ma Z, Wu J. A compartmental model for the analysis of SARS transmission patterns and outbreak control measures in China. Appl Math Comput 2005 Mar 15;162(2):909-924 [FREE Full text] [doi: 10.1016/j.amc.2003.12.131] [Medline: 32287493]
19. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem 1977 Dec;81(25):2340-2361. [doi: 10.1021/j100540a008]
20. Naheed A, Singh M, Lucy D. Numerical study of epidemic model with the inclusion of diffusion in the system. Appl Math Comput 2014 Feb 25;229:480-498 [FREE Full text] [doi: 10.1016/j.amc.2013.12.062] [Medline: 32287498]
21. Lai D. Monitoring the SARS Epidemic in China: A Time Series Analysis. Journal of Data Science 2005;3:293.
22. GitHub. CSSEGISandData / COVID-19. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2020-08-18]

Abbreviations

COVID-19: coronavirus disease
DNN: deep neural networks
JHU CSSE: John Hopkins University Center for Systems Science and Engineering
MERS-CoV: Middle East respiratory syndrome–related coronavirus
RNN-LSTM: recurrent neural networks–long short-term memory
SARS-CoV: severe acute respiratory syndrome coronavirus
SEIR: Susceptible–Exposed–Infectious–Recovered
SEIRQJD: SEIR-Quarantined-Isolated-Deceased
SIRJD: Susceptible–Infectious–Recovered–Isolated–Deceased
SLIR: Susceptible–Latent–Infectious–Recovered

Edited by G Eysenbach; submitted 07.06.20; peer-reviewed by X Zhou, A Louren; comments to author 13.07.20; revised version received 02.08.20; accepted 06.08.20; published 21.08.20.

Please cite as:
Deng Q
Dynamics and Development of the COVID-19 Epidemic in the United States: A Compartmental Model Enhanced With Deep Learning Techniques
J Med Internet Res 2020;22(8):e21173
URL: http://www.jmir.org/2020/8/e21173/
doi:10.2196/21173
PMID:32763892

Original Paper

Effect of an E-Learning Module on Personal Protective Equipment Proficiency Among Prehospital Personnel: Web-Based Randomized Controlled Trial

Laurent Suppan[1], MD; Mohamed Abbas[2*], MD; Loric Stuby[3*], CAS; Philippe Cottet[1], MD; Robert Larribau[1], MD; Eric Golay[1], MAS; Anne Iten[2], MD; Stephan Harbarth[2], MD, MSc; Birgit Gartner[1], MD; Mélanie Suppan[4], MD

[1]Division of Emergency Medicine, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[2]Infection Control Program and WHO Collaborating Centre on Patient Safety, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[3]Genève TEAM Ambulances, Geneva, Switzerland

[4]Division of Anesthesiology, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[*]these authors contributed equally

Corresponding Author:
Laurent Suppan, MD
Division of Emergency Medicine
Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine
University of Geneva Hospitals and Faculty of Medicine
Rue Gabrielle-Perret-Gentil 4
Geneva, 1211
Switzerland
Phone: 41 789151845
Email: laurent.suppan@hcuge.ch

Abstract

Background: To avoid misuse of personal protective equipment (PPE), ensure health care workers' safety, and avoid shortages, effective communication of up-to-date infection control guidelines is essential. As prehospital teams are particularly at risk of contamination given their challenging work environment, a specific gamified electronic learning (e-learning) module targeting this audience might provide significant advantages as it requires neither the presence of learners nor the repetitive use of equipment for demonstration.

Objective: The aim of this study was to evaluate whether a gamified e-learning module could improve the rate of adequate PPE choice by prehospital personnel in the context of the coronavirus disease (COVID-19) pandemic.

Methods: This was an individual-level randomized, controlled, quadruple-blind (investigators, participants, outcome assessors, and data analysts) closed web-based trial. All emergency prehospital personnel working in Geneva, Switzerland, were eligible for inclusion, and were invited to participate by email in April 2020. Participants were informed that the study aim was to assess their knowledge regarding PPE, and that they would be presented with both the guidelines and the e-learning module, though they were unaware that there were two different study paths. All participants first answered a preintervention quiz designed to establish their profile and baseline knowledge. The control group then accessed the guidelines before answering a second set of questions, and were then granted access to the e-learning module. The e-learning group was shown the e-learning module right after the guidelines and before answering the second set of questions.

Results: Of the 291 randomized participants, 176 (60.5%) completed the trial. There was no significant difference in baseline knowledge between groups. Though the baseline proportion of adequate PPE choice was high (75%, IQR 50%-75%), participants' description of the donning sequence was in most cases incorrect. After either intervention, adequate choice of PPE increased significantly in both groups (P<.001). Though the median of the difference in the proportion of correct answers was slightly higher in the e-learning group (17%, IQR 8%-33% versus 8%, IQR 8%-33%), the difference was not statistically significant (P=.27). Confidence in the ability to use PPE was maintained in the e-learning group (P=.27) but significantly decreased in the control group (P=.04).

Conclusions: Among prehospital personnel with an already relatively high knowledge of and experience with PPE use, both web-based study paths increased the rate of adequate choice of PPE. There was no major added value of the gamified e-learning module apart from preserving participants' confidence in their ability to correctly use PPE.

(J Med Internet Res 2020;22(8):e21265) doi:10.2196/21265

KEYWORDS

personal protective equipment; COVID-19; electronic learning; prehospital; randomized controlled trial; protection; equipment; safety; gamified; online learning; communication

Introduction

Background and Importance

Adequate use of personal protective equipment (PPE) is of paramount importance to ensure health care workers' safety and to avoid shortages of such equipment in the context of the coronavirus disease (COVID-19) pandemic [1,2]. Protection guidelines against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection have rapidly evolved following publication of new evidence regarding its mode of transmission, making prompt adaptation of the guidelines both frequent and necessary [3-5]. Prehospital personnel is particularly at risk of contamination as they usually work in a challenging environment and have to stay next to their patients in the narrow space of the ambulance for extended periods of time. This risk is further increased when high-risk procedures such as endotracheal intubation have to be performed [6-8].

To avoid misuse of PPE, effective communication of the corresponding guidelines to frontline health care workers is necessary [9]. However, continuous education has been massively disrupted due to the cancellation of continuous education sessions to limit disease transmission [10]. In this challenging context, electronic learning (e-learning) might provide significant advantages as it requires neither the presence of learners nor the repetitive use of equipment for demonstration as could be the case for live simulation [11,12].

Goal of This Investigation

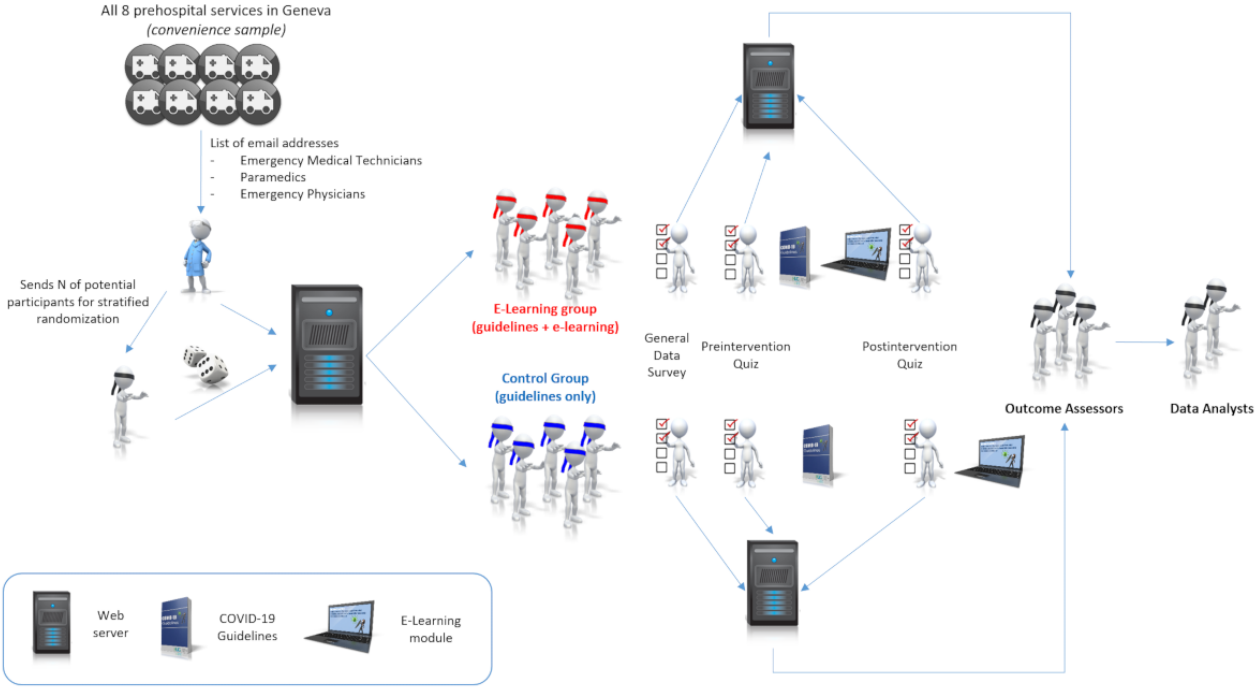
The purpose of this study was to evaluate whether a specifically designed gamified e-learning module [13] could improve the rate of adequate PPE choice by prehospital personnel in the context of the COVID-19 pandemic. Our hypothesis was that knowledge of PPE guidelines would be inconsistent between prehospital personnel, and that an e-learning module may increase and standardize knowledge regarding the use of PPE. This could help limit both underuse and overuse of such equipment.

Methods

Study Design

This was an individual-level, stratified, randomized, controlled, quadruple-blind (investigators, participants, outcome assessors, and data analysts) closed web-based trial (Figure 1) designed following the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth) guidelines [14,15] and incorporating relevant elements from the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) statement [16].

Figure 1. Study design.



The regional ethics committee issued a "Declaration of no objection" (Req-2020-00374) as the population studied was not considered vulnerable according to the Swiss federal law on human research [17]. As the purpose of the study was to examine the effect of two different study paths only on providers' knowledge and attitude toward PPE, registration of the trial was not performed as it is not deemed necessary by the International Committee of Medical Journal Editors [18].

Setting

The study took place in April 2020 in Geneva, Switzerland; the detailed organization of local prehospital emergency services has already been described elsewhere [19]. There are five levels of increasing expertise working in the prehospital field. Ambulance drivers all have a basic life support certificate but are never called upon to deal with emergency situations. Emergency medical technicians (EMTs) are certified after 1 year of training, and can either transport stable patients on their own, or team up with a paramedic to deal with more difficult prehospital situations. The highest level of nonphysician care is provided by paramedics, who complete a 3-year curriculum. Whenever a potentially life-threatening emergency is identified by emergency dispatchers, paramedics are sent in teams of two. A medical reinforcement by way of a light vehicle, Service Mobile d'Urgence et de Réanimation (SMUR), staffed by a paramedic and an emergency physician can either be dispatched at the same time as the ambulance or be called upon by paramedics if specialized medical care is required on site. Emergency physicians are either senior residents or fellows working in one of the following departments: emergency medicine, anesthesiology, or general internal medicine. They can be supervised, either by call or on-site, by a senior specialist emergency physician, which represents the highest level of prehospital care in this setting.

Apart from the physician-staffed prehospital medical service, there are seven different ambulance companies in Geneva, two of which belong to public organizations while the remaining five are privately owned. As each company has its own continuous education structure as well as its own equipment, medical devices, and protocols, one cannot expect all paramedics to share the same knowledge level regarding all aspects of prehospital emergencies.

Online Platform

An online platform [20] developed under the Joomla! 3.9 content management system (Open Source Matters) was created specifically for the purpose of this study. A mailing management component (AcyMailing 5.10, Acyba), a survey component (Community Surveys Pro 5.4.0, CoreJoomla), and a form builder component used to issue completion certificates (BreezingForms Pro 1.9.0, Crosstec) were installed on the website. L Suppan was the only author with access to the platform's administration console. No maintenance or update was planned on the server, platform, or content during the study period.

Study Material

A previously described gamified e-learning module created under Storyline 3 (Articulate Global) was used in this study [13]. The module contains 19 sections and embeds 7 video sequences. Within the module, trigger mechanisms are used to check that the user had accessed and completed all required steps before being allowed to proceed to the following section. The e-learning package is available on the study website where it can be viewed and downloaded freely. The comprehensive prehospital COVID-19 guidelines from the Geneva University Hospitals, version 1.11, was also used in this study [21].

Two quizzes were created by BG and L Suppan: a preintervention quiz designed to establish the participants' baseline knowledge regarding PPE, their use and indication, and a postintervention quiz to assess whether these parameters had changed. Both quizzes contained 10 closed questions, either multiple choice or multiple answer. Questions designed to assess PPE choice were preceded by short clinical scenarios. Each quiz was displayed over 5 pages. The number of questions was limited to reduce attrition. These quizzes were tested and validated by all coinvestigators.

Consistency of specific "free-text" questions, such as age, was checked by means of regular expression ("regex") rules. All answers were automatically checked for completeness by the system before participants were allowed to proceed to the next page. Custom validation messages were displayed to inform users who had not answered a question. Participants were not allowed to correct or review their answers once a page had been completed.

Subjects and Inclusion and Exclusion Criteria

Chief ambulance officers of all services were asked to provide one of the investigators (L Suppan) with a list of all the professional email addresses of their EMTs, paramedics, and emergency physicians. All the email addresses received from these officers were included.

Email addresses of ambulance drivers were excluded as these drivers usually only deal with interhospital transfers and almost never don PPE. Senior specialist emergency physicians were also excluded, as they are few in number and are usually involved in the writing of the guidelines or in the creation of the learning material; in addition, some of them are authors of this study. Finally, the email addresses of the paramedics who participated in the creation of this study or the learning material were also excluded.

Randomization and Allocation Concealment

Before performing a 1:1 randomization, stratification was achieved according to professional status (EMTs, paramedics, and emergency physicians). Email addresses were then sorted according to alphabetical order, and an investigator (MS) who did not have access to the email addresses database was given the number of participants by category and performed the randomization using a computer-generated table. The randomization key was then combined with the list of email addresses and entered in the mailing component by the only investigator who had access to the system (L Suppan). As the list of email addresses and allocations were solely present in an encrypted database, all other investigators were blinded as to group allocation.

Enrolment and Consent

Individual emails that were identical for all participants except for the unique links that pointed to one of the two study paths were sent on April 13, 2020 (Multimedia Appendices 1 and 2). These unique links were automatically created by the survey component but were not recorded in the mailing component other than in generic form (there was no way the individual tokens could be linked to the email addresses). The webmaster could therefore only know whether the email had been opened and the survey started, but was prevented from reconciling the email address with the survey answers, thereby guaranteeing user anonymization. Using unique links served two purposes: they allowed the participants to resume the course in case they were interrupted and also avoided double entries. Given the current circumstances, and as the vast majority of prehospital professionals had had their holidays cancelled and were therefore at work, participants were only given 12 days to complete the study, with email reminders sent on days 3, 7, and 10.

Apart from the survey link, the emails contained information regarding the study length and objectives as well as a short data protection statement. Participants were informed that they would be presented with the most recent version of the prehospital COVID-19 guidelines as well as with an e-learning module, though the order in which these materials would be shown was not explained. They were informed that, by clicking on the survey link, they consented to participate in the study and were provided with the names and electronic addresses of five investigators (BG, EG, L Stuby, L Suppan, and PC), whom they could contact at any time. As collected data were irreversibly anonymized, it was impossible for users to ask for their own answers to be deleted once the survey had been completed.

To improve the response rate, the chief ambulance and medical officers of all companies were asked to encourage their paramedics, EMTs, and emergency physicians to participate in the study. Participation reminders were also sent to all prehospital personnel along with a daily COVID-19 information newsletter.

Participation was not mandatory. No monetary incentive or prize was offered to the survey participants. As the e-learning module was akin to a continuous education session, participants were informed before beginning the survey that they would be able to print a continuous education certificate upon completion. As the certificate component was independent from both the mailing and the survey components, participants were ensured they could generate the certificate without their identity being disclosed.

Study Sequence

After clicking on the survey link, a welcome screen containing detailed information about the study was displayed. This welcome screen was identical for both groups and, similarly to the email messages, did not convey any information regarding the study sequence to ensure the participants were adequately blinded.

After clicking on the start button, participants were asked a series of questions designed to gather demographic-related data. Adaptive questioning was used in this section to avoid displaying irrelevant questions. Participants were then asked a series of general questions related to SARS-CoV-2 and the COVID-19 pandemic.

The control group was then shown the prehospital COVID-19 guidelines. They were then asked to answer a second set of questions before being prompted to evaluate the learning path up to this point. Only then could they access the e-learning module and download their certificate.

The e-learning group followed the same path at first, but accessed the e-learning module immediately after being shown the guidelines. This group then completed the same second set of questions, was asked to evaluate the learning path (which in this case included the e-learning module), and was finally allowed to download the completion certificate.

Outcomes

The main outcome was defined as the difference in the proportion of correct choice of PPE before and after the course, assessed by means of short clinical scenarios.

Secondary outcomes were stratification of the main outcome by profession and by personal history of COVID-19 (whether or not the participant had been infected), accuracy of donning and doffing sequences reconstitutions, differences in the rates of overuse and underuse of PPE, confidence in one's ability to use PPE, perceived usefulness of the course, and satisfaction regarding the course. The latter three outcomes were assessed by means of a 5-point Likert scale.

Data Collection

Data was electronically recorded and securely stored in an encrypted MariaDB database (Version 5.5.5; MariaDB Foundation). At the end of the study, all data was extracted to CSV format by the only investigator who had access (L Suppan). No personally identifiable data (including name, date of birth, email, or IP address) was ever asked for or recorded.

Data Curation and Availability

The extracted data were imported into Stata (StataCorp LLC). Variables were renamed to facilitate their understanding by the blinded data analysts. All data that could have enabled data analysts to identify group allocation were removed, and data fields were sorted accordingly. Incomplete questionnaires were excluded at this stage. The control and e-learning groups were renamed using city names (Moscow and Nairobi), and all relevant data were exported by L Suppan to a Stata .dta file and sent to L Stuby and MA for analysis.

Sample Size

As guidelines differ from region to region, we decided to only include prehospital staff working in the Geneva emergency medical system, thereby using a convenience sample rather than performing a sample size calculation.

Outcomes Assessment

Though most outcomes were electronically recorded and their interpretation therefore was generally independent of subjective human evaluation, comments had to be assessed by outcome

assessors. Two assessors (L Stuby and PC), blinded as to group allocation, were asked to independently assess all comments. The nature of comments were to be rated as "positive," "negative," or "neutral" regarding the study, and as to whether they challenged Infection Prevention and Control (IPC) guidelines (binary, yes versus no). Disagreements were solved by sending the unclear comments to a third outcome assessor (BG), who was blinded to the previous assessments.

Statistical Analysis

Statistical analysis was performed using Stata 15. Continuous independent outcomes were assessed using the Student *t* test or the Mann-Whitney rank-sum test depending on normality. Categorical outcomes were assessed using Fisher exact test. A two-sided $P<.05$ was considered significant. Normality of distribution was first assessed graphically. In case of doubt, the Shapiro-Wilk test was applied.

Continuous paired data were assessed using either the paired Student *t* test or the Wilcoxon matched-pairs signed-rank test depending on normality. The sign test for matched pair was used if symmetry could not be proven. Categorical paired data were analyzed using asymptotic symmetry and marginal homogeneity tests. Binomial paired data were assessed using the McNemar test.

Stratification was defined a priori based on expertise level (EMTs, paramedics, or physicians) and COVID-19 status (negative; confirmed, quarantined; confirmed, cured or unknown). Two post hoc sensitivity analyses were conducted.

Data Availability

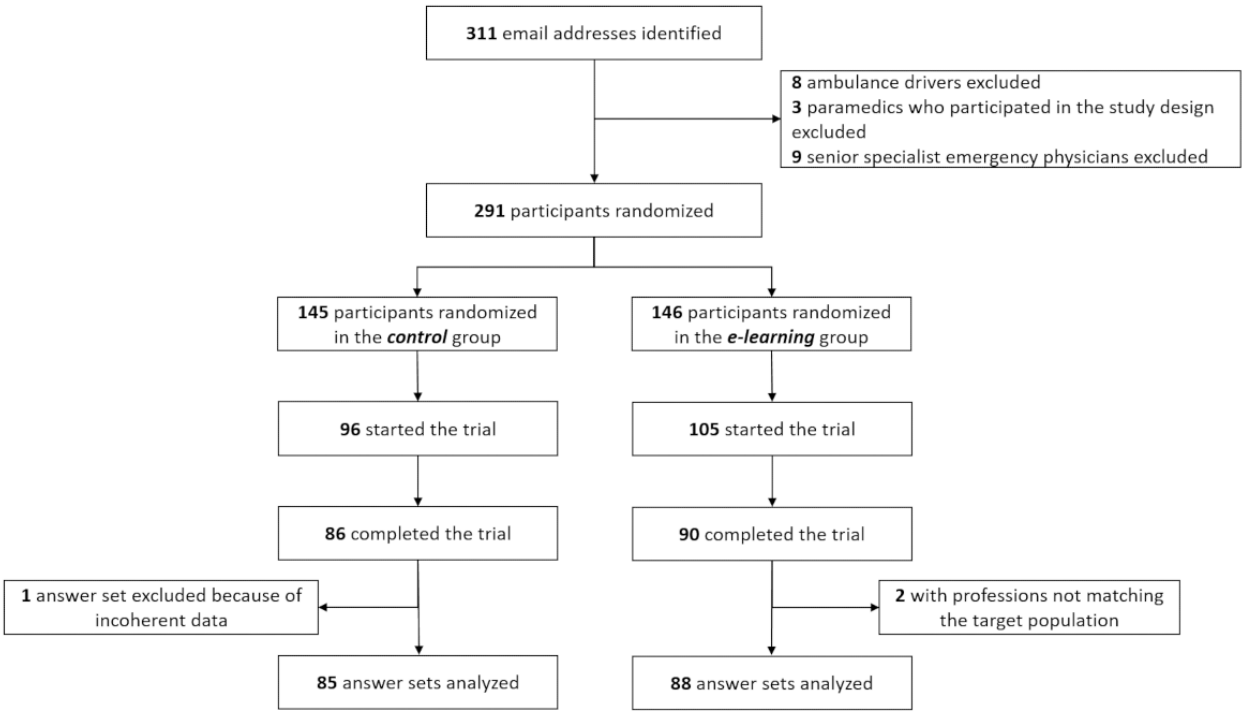
The original data has been deposited to Mendeley Data [22].

Results

Characteristics of Study Subjects

Of the 291 randomized participants, 176 (60.5%) completed the trial (Figure 2). No change was made to the web platform after the first batch of emails was sent. There was no significant downtime (the server was available more than 99% of the time throughout the study period). Just three participants were unable to complete the trial owing to technical problems, the exact nature of which could not be determined: two were not able to access the e-learning module, and one could not access the guideline (in PDF format). The study path was completed in one session by 82.7% of the participants (143/173).

Figure 2. Study flowchart.



The blinded data analysts (L Stuby and MA) excluded two surveys from the e-learning group as the participants' professions did not match the target population (one "ambulance driver" and one "other"). They also excluded one survey from the control group because of incoherent answers (Multimedia Appendix 3). Participant characteristics are described in Table 1.

Table 1. Participant characteristics[a].

Characteristics	Control (n=85)	E-learning (n=88)
Profession, n (%)		
Student paramedic	5 (5.9)	10 (11.4)
Emergency medical technician	11 (12.9)	12 (13.6)
Paramedic	61 (71.8)	60 (68.2)
Physician	8 (9.4)	6 (6.8)
Gender, female, n (%)	32 (37.6)	28 (31.8)
Age (years), median (Q1-Q3)	35 (30-42)	34 (28-40)
Professional experience (years), median (Q1-Q3)	9 (3-15)	7 (2-12)
Prior infection prevention and control course, n (%)		
No/does not remember	73 (85.1)	70 (79.6)
Yes	12 (14.1)	18 (20.4)
Coronavirus disease status, positive, n (%)	7 (8.3)	6 (6.8)
Local guideline seen, yes, n (%)	79 (92.9)	84 (95.4)
Last time guideline seen (days), median (Q1-Q3)	5 (3-10)	5 (2-8)
Specific coronavirus disease course followed, yes, n (%)	28 (32.9)	32 (36.4)

[a]Totals may not equal to 100% due to rounding.

Main Results

There was no significant difference in baseline knowledge between groups. Though the baseline proportion of adequate PPE choice was high (75%, IQR 50%-75%), description of the donning sequence (assessed preintervention) was in most cases incorrect, as only 7 (4%) of the participants were able to reconstitute it accurately, with a similar proportion between groups. The donning sequence initially displayed in the survey was left unchanged by 7% of the participants (12/173, 6 per group).

Adequate choice of PPE was significantly increased in both groups after the intervention (P<.001; Figure 3). Though the median of the difference was higher in the e-learning group (17%, IQR 8%-33% versus 8%, IQR 8%-33%), it did not reach statistical significance (P=.27). This difference was similar regardless of stratification by profession or history of COVID-19 (Table 2). No participant was able to describe the correct doffing sequence, which was assessed postintervention. The doffing sequence originally displayed in the survey was left unchanged by 8% of the participants (14/173, 7 per group).

Figure 3. Change in proportion (%) of adequate choice of personal protective equipment.

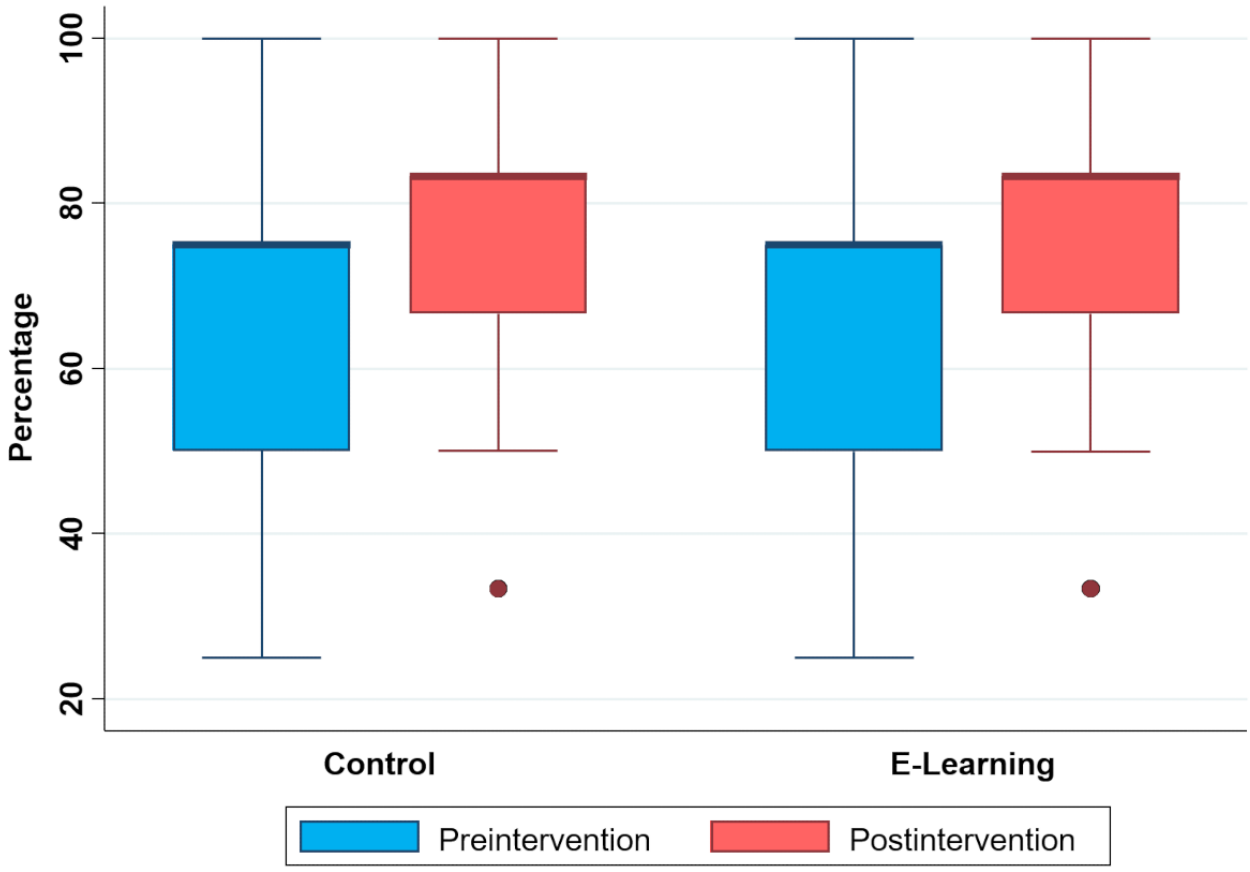


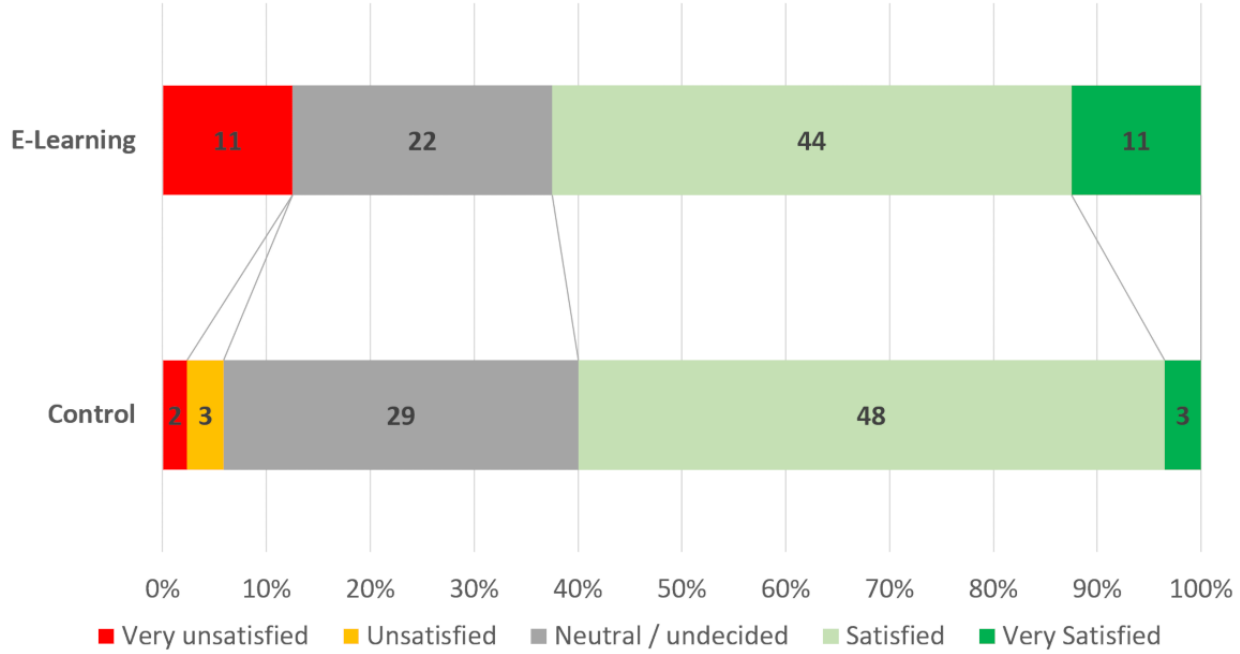
Table 2. Adequate choice of personal protective equipment (main outcome).

Outcomes	Control (n=85)	E-learning (n=88)	*P* value
Main outcome: difference in percentage of correct answers, median (Q1-Q3)	8 (8 to 33)	17 (8 to 33)	.27
Main outcome by profession (%), median (Q1-Q3)			
Paramedic	8 (0 to 25)	17 (8 to 33)	.15
Paramedic student	25 (8 to 33)	29 (25 to 33)	.49
Emergency medical technician	25 (8 to 33)	8 (–8 to 33)	.88
Physician	25 (13 to 38)	13 (–8 to 42)	.30
Main outcome by coronavirus disease status (%), median (Q1-Q3)			
Negative for coronavirus disease	8 (8 to 33)	25 (8 to 33)	.20
Positive for coronavirus disease	8 (–8 to 33)	0 (–8 to 17)	.37

Confidence in the ability of using PPE was identical before and after the course in the e-learning group (*P*=.27), but was significantly lower after the course in the control group (*P*=.04).

While most participants found the course useful (68.5%, 95% CI 61.5%-75.3%), the proportion of participants finding the course “very useful” was significantly higher (*P*=.01) in the e-learning group (33.0%, 95% CI 23.2%-42.8% versus 11.8%, 95% CI 4.9%-18.7%). Participants were generally satisfied regarding the course (60.0%, 95% CI 49.6%-70.4% for the e-learning group versus 62.5%, 95% CI 52.4%-72.6%; *P*=.28; Figure 4).

Figure 4. Participant satisfaction.



The proportion of positive comments was similar between groups (14%, 95% CI 7%-21% in the e-learning group versus 19%, 95% CI 11%-27%; *P*=.16). Participants who expressed disagreements with IPC recommendations were also evenly divided (5.7%, 95% CI 3.2%-8.2% in the e-learning group versus 4.7%, 95% CI 0.2%-9.2%; *P*>0.99).

Two post hoc sensitivity analyses were performed. The first was achieved by excluding all participants who answered they were either "very unsatisfied" or "unsatisfied" with the course. The second was performed by excluding participants whose compliance regarding COVID-19–specific IPC guidelines could be doubted; this included those who did not answer "systematic wearing of a double pair of gloves" to the question "Which of these measures is NOT one of the infection prevention measures?" (40 [45%] in the e-learning group versus 35 [41%]). These analyses did not demonstrate a significant change regarding the main outcome (*P*=.89 and *P*=.29, respectively).

Discussion

Main Results

In this study, the proportion of those making an adequate choice regarding PPE increased significantly after prehospital personnel followed either web-based study path. Though the median of the increase was twice that of the e-learning group, this difference was not statistically significant.

Failure to reach significance might be explained by many different factors. The high baseline proportion of adequate choice of PPE in both groups may have dampened any relative impact of the intervention as there was little room left for improvement. Although this result might seem counterintuitive given the high rate of participants who responded that they did not attend (or did not recall attending) an IPC course prior to this study, more than 90% declared they had consulted the local guidelines recently. Most participants were therefore aware of the local recommendations, and the anxiety generated by the COVID-19 pandemic probably acted as a catalyst regarding their interest in such guidelines [23].

Another factor that might partly explain the lack of a significant difference is the sample size, which limited the power to detect small differences. Nevertheless, as PPE guidelines vary not only between countries, but also between Swiss cantons, it might have been inadequate to draw conclusions from a pooled group of paramedics working in different cantons with different guidelines [24]. Moreover, though increasing the sample size would increase the chances of finding a significant difference, too small an increase would not be clinically relevant regardless of the *P* value obtained [25]. Lack of significance might therefore also be consecutive to lack of effect of the e-learning module regarding knowledge acquisition in this particular population.

Though necessary to assess participants' knowledge and attitudes regarding PPE prior to the interventions, the first set of questions, along with the study title, might have acted as a primer and focused the participants' attention on the specific contents that would be tested postintervention [26]. This effect might have further dampened the potential impact of the e-learning module.

The relatively low level of satisfaction displayed by the participants should also be taken into account. Though e-learning modules and serious games usually increase participant satisfaction when compared to more traditional methods [27,28], such an effect could not be found in this study. Disagreement with IPC guidelines might, at least in part, explain why some participants were dissatisfied [5,29,30]. Indeed, the vast majority of paramedics working in Geneva have had some prior training regarding the use of PPE in situations other than the COVID-19

pandemic. Preparations for high-risk transfers during the 2014 Ebola epidemic [31] and practical exercises in the context of simulated major incidents involving the presence of hazardous materials [32] have presented paramedics with many different PPE guidelines and protocols over the last few years. As the latter situations require more stringent donning and doffing procedures than those described in the COVID-19 context, some paramedics seem to feel that IPC recommendations do not go far enough. There is however a delicate balance between underprotection and overuse of scarcely available resources that must be maintained [33,34]. Moreover, it can hardly be expected to have all paramedics don maximal PPE for all interventions. This would not only considerably increase intervention time, but also make delicate interventions more difficult given the bulk of the PPE and lead to more difficult communications. It could also increase fatigue as such equipment has been shown to be quite uncomfortable [35].

Secondary Outcomes

Though confidence in the ability to use PPE was maintained in the e-learning group, it significantly decreased after reviewing the guidelines in the control group. Participants who initially felt confident in their knowledge might have felt it was challenged after being asked specific questions [36,37]. The e-learning module probably helped restore their altered confidence, as interactive presentations, as well as gamification, have been shown to increase this feeling [38,39]. Nevertheless, most participants were unable, regardless of their assigned intervention, to reconstruct either the donning (assessed preintervention) or the doffing (assessed postintervention) sequences. This potential limitation has already been highlighted in the paper describing the development of the gamified e-learning module used in this study [13], though cut scenes were embedded to provide direct demonstration [40,41].

Though incorrect answers regarding the donning sequence are easily understandable as they might primarily result from a lack of knowledge, an inability to correctly rebuild the doffing sequence is questioning. No less than four different broad categories of causes could be involved: ineffectiveness of the teaching material, inadequate sequence, flawed means of assessing the participants' knowledge, or disagreement with the procedure outlined in the guidelines and in the e-learning module. As this result was unforeseen, the method used in this study was unfortunately ill-suited to the evaluation of the underlying causes. Nevertheless, with less than 10% of participants having left the initially displayed sequences unchanged, a technical flaw can reasonably be ruled out.

Limitations

Apart from the abovementioned limitations, the ever-increasing knowledge regarding SARS-CoV-2 and COVID-19 might render both the guideline and the gamified e-learning module used in this study obsolete. However, current technological tools might mitigate this effect as they allow for a quick adaptation, even of highly interactive content [42].

Another limitation is the relatively small number of questions asked pre- and postintervention. Keeping the total number of questions and the time required to complete either study path relatively low was necessary to limit attrition [16,43]. This strategy was altogether successful as dropouts amounted to less than 15% in either group, a rather lower proportion than generally reported [44]. Attrition was further limited thanks to the use of unique identifiers, which allowed as many as 30 participants (17.3%) to complete their assigned study path in more than one session.

Despite its limitations, this study also has some strengths, among which the quadruple-blind design and the relatively high response rate should be acknowledged. Moreover, as all answers were electronically recorded, there was no risk of an outcome assessment bias. Finally, as neither the control nor the e-learning path requires the physical presence of either participants or instructors, the framework used in this study could serve as the building ground for courses in the context of an epidemic or a pandemic such as the current COVID-19 situation.

Perspectives

The larger impact such a web-based study might have had, regardless of the effect of a specific intervention such as the gamified e-learning module, should be assessed, as a possible change in PPE consumption or infection rate among prehospital providers could ensue.

The potential impact of this gamified e-learning module on less experienced and less primed participants should also be evaluated to confirm (or refute) the theories outlined in this discussion. The gamified e-learning module can thus be freely downloaded from the study website in both a web (HTML5 with Flash fallback) and SCORM (Shareable Content Object Reference Model) format.

Conclusions

Among prehospital personnel with an already relatively high knowledge and experience regarding PPE use, both web-based study paths increased the rate of adequate choice of PPE. There was no major added value of the gamified e-learning module apart from preserving participants' confidence in their ability to correctly use PPE.

Acknowledgments

The authors would like to thank Mr Michel Hofer for his support.

Conflicts of Interest

Most of the authors participated in the development of the gamified e-learning module tested in this trial. Nevertheless, as this module is freely available, the authors deny any financial conflict of interest.

This randomized study was not prospectively registered. As the purpose of the study was to examine the effect of two different study paths only on providers' knowledge and attitude toward PPE, registration of the trial was not performed as it is not deemed necessary by the International Committee of Medical Journal Editors. The editor is therefore following the guidelines suggested by the ICMJE and not mandating prospective registration of this randomized trial. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
First email sent to participants belonging to the control group.
[PDF File (Adobe PDF File), 45 KB - jmir_v22i8e21265_app1.pdf]

Multimedia Appendix 2
First email sent to participants belonging to the e-learning group.
[PDF File (Adobe PDF File), 45 KB - jmir_v22i8e21265_app2.pdf]

Multimedia Appendix 3
Details of the answer set excluded by the blinded data analysts.
[PDF File (Adobe PDF File), 126 KB - jmir_v22i8e21265_app3.pdf]

References

1. Livingston E, Desai A, Berkwits M. Sourcing Personal Protective Equipment During the COVID-19 Pandemic. JAMA 2020 Mar 28:E1-E3. [doi: 10.1001/jama.2020.5317] [Medline: 32221579]
2. Ranney ML, Griffeth V, Jha AK. Critical Supply Shortages - The Need for Ventilators and Personal Protective Equipment during the Covid-19 Pandemic. N Engl J Med 2020 Apr 30;382(18):e41. [doi: 10.1056/NEJMp2006141] [Medline: 32212516]
3. Wax RS, Christian MD. Practical recommendations for critical care and anesthesiology teams caring for novel coronavirus (2019-nCoV) patients. Can J Anaesth 2020 May;67(5):568-576 [FREE Full text] [doi: 10.1007/s12630-020-01591-x] [Medline: 32052373]
4. Guan L, Zhou L, Zhang J, Peng W, Chen R. More awareness is needed for severe acute respiratory syndrome coronavirus 2019 transmission through exhaled air during non-invasive respiratory support: experience from China. Eur Respir J 2020 Mar 20;55(3):2000352 [FREE Full text] [doi: 10.1183/13993003.00352-2020] [Medline: 32198275]
5. Rhee C, Baker MA, Klompas M. The COVID-19 infection control arms race. Infect Control Hosp Epidemiol 2020 May 08:1-3 [FREE Full text] [doi: 10.1017/ice.2020.211] [Medline: 32381140]
6. Teter J, Millin MG, Bissell R. Hand hygiene in emergency medical services. Prehosp Emerg Care 2015 Nov 21;19(2):313-319. [doi: 10.3109/10903127.2014.967427] [Medline: 25415186]
7. Valdez MK, Sexton JD, Lutz EA, Reynolds KA. Spread of infectious microbes during emergency medical response. Am J Infect Control 2015 Jun;43(6):606-611 [FREE Full text] [doi: 10.1016/j.ajic.2015.02.025] [Medline: 26042849]
8. Lockhart SL, Duggan LV, Wax RS, Saad S, Grocott HP. Personal protective equipment (PPE) for both anesthesiologists and other airway managers: principles and practice during the COVID-19 pandemic. Can J Anaesth 2020 Apr 23:1005-1015 [FREE Full text] [doi: 10.1007/s12630-020-01673-w] [Medline: 32329014]
9. Oh HS, Uhm D. Occupational exposure to infection risk and use of personal protective equipment by emergency medical personnel in the Republic of Korea. Am J Infect Control 2016 Jun 01;44(6):647-651 [FREE Full text] [doi: 10.1016/j.ajic.2015.12.022] [Medline: 26874405]
10. Rose S. Medical Student Education in the Time of COVID-19. JAMA 2020 Mar 31;323(21):2131. [doi: 10.1001/jama.2020.5227] [Medline: 32232420]
11. Ahmed H, Allaf M, Elghazaly H. COVID-19 and medical education. Lancet Infect Dis 2020 Jul;20(7):777-778 [FREE Full text] [doi: 10.1016/S1473-3099(20)30226-7] [Medline: 32213335]
12. Wittich C, Agrawal A, Cook D, Halvorsen A, Mandrekar J, Chaudhry S, et al. E-learning in graduate medical education: survey of residency program directors. BMC Med Educ 2017 Jul 11;17(1):114 [FREE Full text] [doi: 10.1186/s12909-017-0953-9] [Medline: 28697744]
13. Suppan M, Gartner B, Golay E, Stuby L, White M, Cottet P, et al. Teaching Adequate Prehospital Use of Personal Protective Equipment During the COVID-19 Pandemic: Development of a Gamified e-Learning Module. JMIR Serious Games 2020 Jun 12;8(2):e20173 [FREE Full text] [doi: 10.2196/20173] [Medline: 32516115]
14. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
15. de Kraker MEA, Abbas M, Huttner B, Harbarth S. Good epidemiological practice: a narrative review of appropriate scientific methods to evaluate the impact of antimicrobial stewardship interventions. Clin Microbiol Infect 2017 Nov;23(11):819-825 [FREE Full text] [doi: 10.1016/j.cmi.2017.05.019] [Medline: 28571767]

16. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
17. CC 810.30 Federal Act of 30 September 2011 on Research involving Human Beings (Human Research Act, HRA). URL: https://www.admin.ch/opc/en/classified-compilation/20061313/index.html [accessed 2020-05-12]
18. Clinical Trials Registration. International Committee of Medical Journal Editors. URL: http://www.icmje.org/about-icmje/faqs/clinical-trials-registration/ [accessed 2020-06-10]
19. Chan M, Fehlmann CA, Pasquier M, Suppan L, Savoldelli GL. Endotracheal Intubation Success Rate in an Urban, Supervised, Resident-Staffed Emergency Mobile System: An 11-Year Retrospective Cohort Study. J Clin Med 2020 Jan 16;9(1):238 [FREE Full text] [doi: 10.3390/jcm9010238] [Medline: 31963162]
20. Gamified E-Learning Module. URL: https://coronavirus.anesth.ch [accessed 2020-08-18]
21. Brigade Sanitaire Cantonale. Coronavirus de la maladie COVID-19: prise en charge préhospitalière. URL: https://smur.hug-ge.ch/sites/smur/files/guideline_covid19_1_11c.pdf [accessed 2020-04-22]
22. Suppan M, Stuby L, Abbas M, Gartner B, Suppan L. COVID-19 - Prehospital PPE Dataset. Mendeley Data. 2020. URL: https://data.mendeley.com/datasets/ffh678bv42/1 [accessed 2020-08-17]
23. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: A systematic review and meta-analysis. Brain Behav Immun 2020 Aug;88:901-907 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.026] [Medline: 32437915]
24. Glass GV. Primary, Secondary, and Meta-Analysis of Research. Educational Researcher 1976 Nov 1;5(10):3-8. [doi: 10.3102/0013189X005010003]
25. Ranganathan P, Pramesh C, Buyse M. Common pitfalls in statistical analysis: Clinical versus statistical significance. Perspect Clin Res 2015;6(3):169-170 [FREE Full text] [doi: 10.4103/2229-3485.159943] [Medline: 26229754]
26. Latimier A, Riegert A, Peyre H, Ly ST, Casati R, Ramus F. Does pre-testing promote better retention than post-testing? NPJ Sci Learn 2019 Sep 24;4(1):15. [doi: 10.1038/s41539-019-0053-1] [Medline: 31583117]
27. Rouleau G, Gagnon M, Côté J, Payne-Gagnon J, Hudson E, Dubois C, et al. Effects of E-Learning in a Continuing Education Context on Nursing Care: Systematic Review of Systematic Qualitative, Quantitative, and Mixed-Studies Reviews. J Med Internet Res 2019 Oct 02;21(10):e15118 [FREE Full text] [doi: 10.2196/15118] [Medline: 31579016]
28. Lozano-Lozano M, Fernández-Lao C, Cantarero-Villanueva I, Noguerol I, Álvarez-Salvago F, Cruz-Fernández M, et al. A Blended Learning System to Improve Motivation, Mood State, and Satisfaction in Undergraduate Students: Randomized Controlled Trial. J Med Internet Res 2020 May 22;22(5):e17101 [FREE Full text] [doi: 10.2196/17101] [Medline: 32441655]
29. Godlee F. Protect our healthcare workers. BMJ 2020 Apr 02:m1324. [doi: 10.1136/bmj.m1324]
30. Houghton C, Meskell P, Delaney H, Smalle M, Glenton C, Booth A, et al. Barriers and facilitators to healthcare workers' adherence with infection prevention and control (IPC) guidelines for respiratory infectious diseases: a rapid qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Apr 21;4:CD013582. [doi: 10.1002/14651858.CD013582] [Medline: 32315451]
31. Schibler M, Vetter P, Cherpillod P, Petty T, Cordey S, Vieille G, et al. Clinical features and viral kinetics in a rapidly cured patient with Ebola virus disease: a case report. The Lancet Infectious Diseases 2015 Sep;15(9):1034-1040. [doi: 10.1016/s1473-3099(15)00229-7]
32. Lettre d'information ERNS 19. Confédération suisse. 2019. URL: https://www.vbs.admin.ch/content/vbs-internet/fr/verschiedene-themen-des-vbs/die-schweizer-sicherheitspolitik/die-sicherheitsverbundsuebung--svu19-/_jcr_content/contentPar/tabs/items/documents/tabPar/downloadlist/downloadItems/30_1567670674944.download/Lettre%20information%20ERNS%2019%20-%20Aout%202019.pdf [accessed 2020-08-18]
33. Rowan NJ, Laffey JG. Challenges and solutions for addressing critical shortage of supply chain for personal and protective equipment (PPE) arising from Coronavirus disease (COVID19) pandemic - Case study from the Republic of Ireland. Sci Total Environ 2020 Jul 10;725:138532 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138532] [Medline: 32304970]
34. Delgado D, Wyss Quintana F, Perez G, Sosa Liprandi A, Ponte-Negretti C, Mendoza I, et al. Personal Safety during the COVID-19 Pandemic: Realities and Perspectives of Healthcare Workers in Latin America. Int J Environ Res Public Health 2020 Apr 18;17(8):2798 [FREE Full text] [doi: 10.3390/ijerph17082798] [Medline: 32325718]
35. Kang J, O'Donnell JM, Colaianne B, Bircher N, Ren D, Smith K. Use of personal protective equipment among health care personnel: Results of clinical observations and simulations. Am J Infect Control 2017 Jan 01;45(1):17-23. [doi: 10.1016/j.ajic.2016.08.011] [Medline: 28065328]
36. Serra M, DeMarree K. Unskilled and unaware in the classroom: College students' desired grades predict their biased grade predictions. Mem Cognit 2016 Oct;44(7):1127-1137. [doi: 10.3758/s13421-016-0624-9] [Medline: 27270923]
37. Osterhage JL, Usher EL, Douin TA, Bailey WM. Opportunities for Self-Evaluation Increase Student Calibration in an Introductory Biology Course. CBE Life Sci Educ 2019 Jun;18(2):ar16 [FREE Full text] [doi: 10.1187/cbe.18-10-0202] [Medline: 31025916]
38. Tan AJQ, Lee CCS, Lin PY, Cooper S, Lau LST, Chua WL, et al. Designing and evaluating the effectiveness of a serious game for safe administration of blood transfusion: A randomized controlled trial. Nurse Educ Today 2017 Aug;55:38-44. [doi: 10.1016/j.nedt.2017.04.027] [Medline: 28521248]

39. Dankbaar MEW, Richters O, Kalkman CJ, Prins G, Ten Cate OTJ, van Merrienboer JJG, et al. Comparative effectiveness of a serious game and an e-module to support patient safety knowledge and awareness. BMC Med Educ 2017 Feb 02;17(1):30 [FREE Full text] [doi: 10.1186/s12909-016-0836-5] [Medline: 28148296]
40. Koka A, Suppan L, Cottet P, Carrera E, Stuby L, Suppan M. Teaching the National Institutes of Health Stroke Scale to Paramedics (E-Learning vs Video): Randomized Controlled Trial. J Med Internet Res 2020 Jun 09;22(6):e18358 [FREE Full text] [doi: 10.2196/18358] [Medline: 32299792]
41. Brame CJ. Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. CBE Life Sci Educ 2016;15(4):1-6 [FREE Full text] [doi: 10.1187/cbe.16-03-0125] [Medline: 27789532]
42. Kwag K, González-Lorenzo M, Banzi R, Bonovas S, Moja L. Providing Doctors With High-Quality Information: An Updated Evaluation of Web-Based Point-of-Care Information Summaries. J Med Internet Res 2016 Jan 19;18(1):e15 [FREE Full text] [doi: 10.2196/jmir.5234] [Medline: 26786976]
43. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
44. Hochheimer CJ, Sabo RT, Krist AH, Day T, Cyrus J, Woolf SH. Methods for Evaluating Respondent Attrition in Web-Based Surveys. J Med Internet Res 2016 Nov 22;18(11):e301 [FREE Full text] [doi: 10.2196/jmir.6342] [Medline: 27876687]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth
COVID-19: coronavirus disease
E-learning: electronic learning
EMT: emergency medical technician
IPC: infection prevention and control
PPE: personal protective equipment
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
SCORM: Shareable Content Object Reference Model
SMUR: Service Mobile d'Urgences et de Réanimation

Edited by G Eysenbach; submitted 13.06.20; peer-reviewed by C Miranda, H Schröder; comments to author 31.07.20; revised version received 02.08.20; accepted 03.08.20; published 21.08.20.

Please cite as:
Suppan L, Abbas M, Stuby L, Cottet P, Larribau R, Golay E, Iten A, Harbarth S, Gartner B, Suppan M
Effect of an E-Learning Module on Personal Protective Equipment Proficiency Among Prehospital Personnel: Web-Based Randomized Controlled Trial
J Med Internet Res 2020;22(8):e21265
URL: http://www.jmir.org/2020/8/e21265/
doi:10.2196/21265
PMID:32747329

Original Paper

Clinical Characteristics and Outcomes of Childbearing-Age Women With COVID-19 in Wuhan: Retrospective, Single-Center Study

Lijie Wei[1*], PhD; Xuan Gao[1*], MD; Suhua Chen[1], PhD; Wanjiang Zeng[1], PhD; Jianli Wu[1], PhD; Xingguang Lin[1], PhD; Huiting Zhang[1], MD; Lali Mwamaka Sharifu[1], PhD; Ling Chen[2], PhD; Ling Feng[1], PhD; Shaoshuai Wang[1], PhD

[1]Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430030, China

[2]Department of Pediatrics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[*]these authors contributed equally

Corresponding Author:
Shaoshuai Wang, PhD
Department of Obstetrics and Gynecology
Tongji Hospital
Tongji Medical College, Huazhong University of Science and Technology
No. 1095, Jiefang Avenue
Wuhan, Hubei 430030
China
Phone: 86 83663812
Email: colombo2008@sina.com

Abstract

Background: Since December 2019, an outbreak of the coronavirus disease (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread rapidly in Wuhan and worldwide. However, previous studies on pregnant patients were limited.

Objective: The aim of this study is to evaluate the clinical characteristics and outcomes of pregnant and nonpregnant women with COVID-19.

Methods: This study retrospectively collected epidemiological, clinical, laboratory, imaging, management, and outcome data of 43 childbearing-age women patients (including 17 pregnant and 26 nonpregnant patients) who presented with laboratory-confirmed COVID-19 in Tongji Hospital, Wuhan, China from January 19 to March 2, 2020. Clinical outcomes were followed up to March 28, 2020.

Results: Of the 43 childbearing-age women in this study, none developed a severe adverse illness or died. The median ages of pregnant and nonpregnant women were 33.0 and 33.5 years, respectively. Pregnant women had a markedly higher proportion of history exposure to hospitals within 2 weeks before onset compared to nonpregnant women (9/17, 53% vs 5/26, 19%, *P*=.02) and a lower proportion of other family members affected (4/17, 24% vs 19/26, 73%, *P*=.004). Fever (8/17, 47% vs 18/26, 69%) and cough (9/17, 53% vs 12/26, 46%) were common onsets of symptoms for the two groups. Abdominal pain (n=4, 24%), vaginal bleeding (n=1, 6%), reduced fetal movement (n=1, 6%), and increased fetal movement (n=2, 13%) were observed at onset in the 17 pregnant patients. Higher neutrophil and lower lymphocyte percent were observed in the pregnant group compared to the nonpregnant group (79% vs 56%, *P*<.001; 15% vs 33%, *P*<.001, respectively). In both groups, we observed an elevated concentration of high-sensitivity C-reactive protein, erythrocyte sedimentation rate, aminotransferase, and lactate dehydrogenase. Concentrations of alkaline phosphatase and D-dimer in the pregnant group were significantly higher than those of the nonpregnant group (119.0 vs 48.0 U/L, *P*<.001; 2.1 vs 0.3μg/mL, *P*<.001, respectively). Both pregnant (4/10, 40%) and nonpregnant (8/15, 53%) women tested positive for influenza A virus. A majority of pregnant and nonpregnant groups received antiviral (13/17, 76% vs 25/26, 96%) and antibiotic (13/17, 76% vs 23/26, 88%) therapy. Additionally, both pregnant (2/11, 18%) and nonpregnant (2/19, 11%) recovered women redetected positive for SARS-CoV-2 after discharge.

Conclusions: The epidemiology and clinical and laboratory features of pregnant women with COVID-19 were diverse and atypical, which increased the difficulty of diagnosis. Most pregnant women with COVID-19 were mild and moderate, and rarely developed severe pneumonia or severe adverse outcomes.

(J Med Internet Res 2020;22(8):e19642) doi:10.2196/19642

KEYWORDS

COVID-19; SARS-CoV-2; childbearing age; pregnancy; clinical characteristics; outcomes; women; health information; epidemiology; diagnosis; symptom

Introduction

In December 2019, a cluster of cases of pneumonia of unknown etiology were identified in Wuhan, China [1]. Further investigation revealed these cases were caused by a novel coronavirus, which was termed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Pneumonia caused by SARS-CoV-2 was termed the coronavirus disease (COVID-19) [2,3]. In the past 2 decades, two human coronaviruses including severe acute respiratory syndrome–related coronavirus (SARS-CoV) and Middle East respiratory syndrome–related coronavirus (MERS-CoV) can cause severe lower respiratory tract infections [4,5]. SARS-CoV-2 is similar to SARS-CoV as both of them belong to the beta coronavirus genus, and SARS-CoV-2 shares more than 79.6% sequence identity with SARS-CoV [6]. As of April 18, 2020, the cumulative number of confirmed cases of COVID-19 infection in China had exceeded 86,700, and the death toll is more than 4600. The cumulative total number of confirmed cases has globally exceeded 2,350,000 and continues to increase [7,8]. The World Health Organization (WHO) has designated the COVID-19 pandemic a Public Health Emergency of International Concern.

Pregnant women have been hypothesized to be susceptible to respiratory pathogens and severe adverse outcomes of pneumonia due to the normal physiological changes during pregnancy, including altered cell-mediated immunity and changes in pulmonary [9,10]. Previous studies reported that pregnant women infected with SARS-CoV or MERS-CoV were more susceptible to severe adverse outcomes including maternal morbidity and death. The case fatality rate (CFR) for pregnant women infected with SARS-CoV reached 25%-30%, much higher than that of the general population [11,12]. Data for pregnant women infected with MERS-CoV is scarce. A case series of 5 pregnant women with MERS reported that the CFR reached 40% [13]. Unfortunately, there is limited experience on COVID-19 infections during pregnancy, and all current studies are single-center trials. Two studies with a small sample size reported none of the pregnant women with COVID-19 have died yet [14,15]. However, currently there is no vaccine or specific treatment for COVID-19 infection.

In this study, we describe the clinical, laboratory, imaging findings, and clinical outcomes of 43 childbearing-age women patients (including 17 pregnant and 26 nonpregnant women) in Wuhan infected with SARS-CoV-2. This will provide an insight into the prevention and treatment of pregnant women with COVID-19.

Methods

Recruitment

This study retrospectively recruited patients from January 19 to March 2, 2020, at Tongji Hospital, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, Hubei, China. According to the arrangements put in place by the Chinese Government, pregnant and nonpregnant women patients were admitted to the designated hospitals in Wuhan for managing COVID-19 without selectivity. All patients were diagnosed with COVID-19 according to “Diagnosis and Treatment Protocol for COVID-19 (Sixth Trial Edition)” released by the National Health Commission of the People’s Republic of China [16].

There were 17 pregnant and 32 nonpregnant women’s throat swabs that tested positive for SARS-CoV-2 RNA from January 19 to March 2, 2020; among them, 6 nonpregnant women with comorbidities were excluded (2 had hypertension, 1 had diabetes, 1 had a history of kidney transplantation, 1 had lymphoma, and 1 had connective tissue disease). The remaining 17 pregnant women and 26 nonpregnant women did not have any underlying comorbidities due to a chronic disease such as hypertension, diabetes, or heart disease. Two groups were matched with respect to age, gender, timing of contact with COVID-19, and the proportion of health care workers. Additionally, all patients recruited were Chinese residents and lived in Wuhan with no exposure to the Huanan seafood market in Wuhan.

This study was reviewed and approved by the Ethics Committee of Tongji Hospital, Tongji Medical College of Huazhong University of Science and Technology (TJ-IRB20200222). Informed consent for this retrospective study was waived. The anonymous data was collected and analyzed to facilitate better clinical decisions and treatment.

Data Collection

We retrospectively collected epidemiological, clinical, laboratory, imaging, management, and outcome data for all the patients with COVID-19 in the two groups. Clinical outcomes were followed up to March 28, 2020. Two researchers (LW and XG) evaluated the participants and reviewed the data independently, with disagreements resolved by consensus.

Throat swab specimens for all patients were tested for SARS-CoV-2 at Tongji Hospital. SARS-CoV-2 was confirmed following the WHO guidelines for quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR) [17]. Throat swab specimens from the upper respiratory tract that were obtained from all patients on admission were maintained in a viral-transport medium. Other pneumonia-related respiratory pathogens including influenza A virus, influenza B virus, respiratory syncytial virus, adenovirus, parainfluenza viruses, legionella pneumophila, mycoplasma pneumoniae, and chlamydia pneumoniae were tested by enzyme-linked immunosorbent assay (ELISA). qRT-PCR and ELISA detection reagents were provided by Tongji Hospital. Additionally, except for 1 pregnant woman who did not consent, all the remaining patients took a chest computed tomography (CT).

Statistical Analysis

Statistical analysis was performed with SPSS version 23.0 (IBM Corp). Continuous variables were presented as median (interquartile range). Categorical variables were expressed as number and proportion (%). The Mann-Whitney U test was applied for comparing two groups of continuous variables. The chi-square test or Fisher exact test were applied for discrete variables of two groups. A *P* value with a two-tailed test <.05 was considered as statistically significant.

Results

Demographics and Clinical Characteristics of Pregnant Women and Nonpregnant Women

A total of 17 pregnant and 26 nonpregnant women with COVID-19 were included in this study. Among the 17 pregnant women, 1 was in her first trimester, 3 were in their second trimester, and 13 were in their third trimester. None of them had a history of exposure to the Huanan seafood market. Of the sample, 18% (3/17) of pregnant and 19% (5/26) of nonpregnant women were health care workers. Pregnant women had a higher proportion of history exposure to hospitals within 2 weeks before onset compared to nonpregnant women (9/17, 53% vs 5/26, 19%, *P*=.02) and a lower proportion of other family members infected with COVID-19 (4/17, 24% vs 19/26, 73%, *P*=.004) than nonpregnant women. The median ages of pregnant and nonpregnant women were 33.0 and 33.5 years, respectively. The median time from symptom onset to hospital presentation in the pregnant and nonpregnant groups were 2.0 and 4.0 days, respectively. Of the patients, 2 of 17 (12%) pregnant women and 3 of 26 (11%) nonpregnant women were diagnosed with a severe type on admission. None of the patients developed critical illness (Table 1).

The symptoms at onset of pregnant women with COVID-19 were similar to nonpregnant women. The most common symptoms at onset of pregnant and nonpregnant women were fever (8/17, 47% vs 18/26, 69%) and cough (9/17, 53% vs 12/26, 46%). Other pneumonia-related symptoms at onset including fatigue, expectoration, chest tightness, and shortness of breath were less common. Chills and rigors, headache, and myalgia had not been observed in pregnant women prior to the infection. Both pregnant (1/17, 6%) and nonpregnant (4/26, 15%) groups had diarrhea at onset. Of the 17 pregnant women, 2 (12%) asymptomatic pregnant women were diagnosed during hospitalization routine tests as a requirement before delivery. Of the 26 nonpregnant women, 2 (8%) asymptomatic nonpregnant women were diagnosed by testing for SARS-CoV-2 of throat swabs because they had a history of contact with an infected person. Additionally, pregnancy-related symptoms were also observed in pregnant women, including abdominal pain (n=4, 24%), vaginal bleeding (n=1, 6%), reduced fetal movement (n=1, 6%), and increased fetal movement (n=2, 13%). There were 2 pregnant women who only had pregnancy-related symptoms until being diagnosed (Table 1).

Table 1. Epidemiological and clinical features of pregnant and nonpregnant women with the coronavirus disease.

Variables	Total (N=43)	Pregnancy (n=17)	Nonpregnancy (n=26)	*P* value
Age (years), median (IQR)	33.0 (30.0-37.0)	33.0 (30.0-35.0)	33.5 (31.0-38.0)	.28
Gestational age on admission, n (%)				
First trimester	N/A[a]	1 (6)	N/A	N/A
Second trimester	N/A	3 (18)	N/A	N/A
Third trimester	N/A	13 (76)	N/A	N/A
Health care workers, n (%)	8 (19)	3 (18)	5 (19)	.77
Hospital exposure within 2 weeks before onset, n (%)	14 (33)	9 (53)	5 (19)	.02
Other family members affected, n (%)	23 (53)	4 (24)	19 (73)	.004
Time from onset of symptom to first outpatient visit (days), median (IQR)	3.5 (1.0-7.0)	2.0 (0.9-10.8)	4.0 (1.0-7.0)	.75
Clinical classification, n (%)				.54
Mild	3 (7)	2 (12)	1 (4)	
Moderate	35 (81)	13 (76)	22 (85)	
Severe	5 (12)	2 (12)	3 (12)	
Critical	0 (0)	0 (0)	0 (0)	
Symptoms at onset, n (%)				
Fever	26 (60)	8 (47)	18 (69)	.15
Chills and rigors	2 (5)	0 (0)	2 (8)	.67
Headache	1 (2)	0 (0)	1 (4)	.83
Dizziness	1 (2)	1 (6)	0 (0)	.83
Fatigue	5 (12)	1 (6)	4 (15)	.93
Cough	21 (49)	9 (53)	12 (46)	.66
Expectoration	9 (21)	3 (18)	6 (23)	.96
Chest tightness	5 (12)	2 (12)	3 (12)	.64
Shortness of breath	2 (5)	1 (6)	1 (4)	.67
Myalgia	1 (2)	0 (0)	1 (4)	.83
Diarrhea	5 (12)	1 (6)	4 (15)	.64
Asymptomatic	4 (9)	2 (12)	2 (8)	.93
Abdominal pain	N/A	4 (24)	N/A	N/A
Vaginal bleeding	N/A	1 (6)	N/A	N/A
Reduced fetal movements	N/A	1 (6)	N/A	N/A
Increased fetal movement	N/A	2 (13)	N/A	N/A

[a]N/A: not applicable.

Laboratory and Imaging Characteristics of Pregnant Women and Nonpregnant Women

On admission, the median white blood cell count of patients in the pregnant group with COVID-19 was significantly higher than the nonpregnant group (7.8 vs 3.8×10^9/L, *P*<.001). There were 4/17 (24%) pregnant women and 0/26 (0%) nonpregnant women that developed leukocytosis (white blood cell count>10.0×10^9/L). Neutrophil percentage and neutrophil count were higher in pregnant women compared to nonpregnant women (79% vs 56%, *P*<.001; 6.7 vs 2.3×10^9/L, *P*<.001, respectively). Lymphopenia (lymphocyte count<1.0×10^9/L) occurred in 7 (41%) of the 17 pregnant women and 10 (38%) of the 26 nonpregnant women. There was no statistical difference in hemoglobin concentration and platelet count between the two groups (Table 2).

Table 2. Laboratory and imaging features of pregnant and nonpregnant women with the coronavirus disease.

Variables	Total (N=43)	Pregnancy (n=17)	Nonpregnancy (n=26)	*P* value
Routine blood test				
White blood cell count (×10^9/L), median (IQR)	5.2 (3.8-7.6)	7.8 (6.6-10.2)	3.8 (3.6-5.1)	<.001
<4.0, n (%)	12 (28)	0 (0)	12 (46)	.003
>10.0, n (%)	4 (9)	4 (24)	0 (0)	.04
Neutrophil percent (%), median (IQR)	64.4 (55.9-79.4)	80.5 (72.2-85.2)	58.0 (49.4-63.0)	<.001
>75, n (%)	12 (28)	12 (71)	0 (0)	<.001
Neutrophil count (×10^9/L), median (IQR)	3.3 (2.1-5.5)	6.7 (5.3-8.2)	2.3 (1.9-2.9)	<.001
<1.5, n (%)	1 (2)	0 (0)	1 (3)	.83
Lymphocyte percent (%), median (IQR)	24.9(14.4-35.9)	13.0 (11.6-20.1)	32.7 (26.4-39.7)	<.001
<20, n (%)	14 (33)	13 (77)	1 (4)	<.001
Lymphocyte count (×10^9/L), median (IQR)	1.4 (1.0-1.8)	1.1 (0.9-1.6)	1.4 (1.0-2.0)	.21
<1.0, n (%)	17 (40)	7 (42)	10 (38)	.86
Hemoglobin (g/L), median (IQR)	122.5 (113.8-128.5)	117.0 (111.0-132.0)	123.0 (117.0-127.0)	.86
<115, n (%)	12 (28)	6 (35)	6 (23)	.38
Platelet count (×10^9/L), median (IQR)	209.0 (160.0-242.0)	198.0 (138.0-227.3)	210.0 (171.0-250.3)	.24
<150, n (%)	10 (23)	6 (35)	4 (15)	.25
Other laboratory features				
High sensitivity C-reactive protein (mg/L), median (IQR)	6.7 (0.7-25.3)	16.7 (7.1-47.6)	1.6 (0.4-13.0)	.07
≥10, n/N (%)	14/31 (45)	7 /10 (70)	7/21 (33)	.12
Procalcitonin (ng/mL), median (IQR)	0.04 (0.03-0.05)	0.05 (0.03-0.17)	0.04 (0.03-0.05)	.16
≥0.05, n/N (%)	0/23 (0)	0/9 (0)	0/14 (0)	N/A[a]
Erythrocyte sedimentation rate (mm/h), median (IQR)	26.0 (12.0-41.0)	36.5 (26.3-82.0)	24.0 (7.0-38.0)	.08
>20, n/N (%)	13/20 (65)	5/5 (100)	8/15 (53)	.11
Alanine aminotransferase (U/L), median (IQR)	16.5 (9.0-26.0)	13.0 (9.0-28.0)	23.0 (9.0-26.5)	.72
≥45, n/N (%)	6/42 (14)	3/17 (18)	3/25 (12)	.95
Aspartate aminotransferase (U/L), median (IQR)	17.0 (13.0-28.3)	20.0 (14.0-42.5)	15.0 (10.5-25.0)	.047
≥35, n/N (%)	9/42 (21)	5/17 (29)	4/25 (16)	.51
Lactate dehydrogenase (U/L), median (IQR)	204.0 (172.0-286.0)	235.0 (182.0-309.0)	193.0 (161.0-277.0)	.13
≥250, n/N (%)	13/38 (34)	6/15 (40)	7/23 (30)	.73
Alkaline phosphatase (U/L), median (IQR)	57.5 (46.5-111.3)	119.0 (77.0-142.0)	48.0 (42.0-57.0)	<.001
≥100, n/N (%)	11/38 (29)	10/15 (67)	1/23 (4)	<.001
Creatinine (μmol/L), median (IQR)	52.5 (46.0-61.0)	50.0 (43.2-59.5)	53.0 (48.0-62.5)	.21
≥106, n/N (%)	0/38 (0)	0/14 (0)	0/24 (0)	N/A
Creatine kinase (U/L), median (IQR)	51.5 (35.8-70.8)	81.0 (29.0-147.5)	48.5 (37.3-61.0)	.34
≥140, n/N (%)	1/20 (5)	1/6 (17)	0/14 (0)	.30
D-dimer (μg/mL), median (IQR)	0.7 (0.3-2.0)	2.1 (1.7-3.1)	0.3 (0.2-0.7)	<.001
≥0.5, n/N (%)	19/34 (56)	11/12 (92)	8/22 (36)	.003
Pneumonia-associated pathogens, n/N (%)				
Respiratory syncytial virus	0/24 (0)	0/10 (0)	0/14 (0)	N/A
Adenovirus	0/24 (0)	0/10 (0)	0/14 (0)	N/A
Influenza A virus	12/25 (48)	4/10 (40)	8/15 (53)	.69

Variables	Total (N=43)	Pregnancy (n=17)	Nonpregnancy (n=26)	*P* value
Influenza B virus	0/25 (0)	0/10 (0)	0/15 (0)	N/A
Parainfluenza viruses	0/24 (0)	0/10 (0)	0/14 (0)	N/A
Legionella pneumophila	1/24 (4)	1/10 (10)	0/14 (0)	.42
Mycoplasma pneumoniae	2/22 (9)	1/10 (10)	1/12 (8)	>.99
Chlamydia pneumoniae	0/23 (0)	0/10 (0)	0/13 (0)	N/A
Chest computed tomographic findings, n/N (%)				.67
Normal	1/42 (2)	0/16 (0)	1/26 (4)	
Unilateral pneumonia	9/42 (21)	3/16 (19)	6/26 (23)	
Bilateral pneumonia	32/42 (76)	13/16 (81)	19/26 (73)	

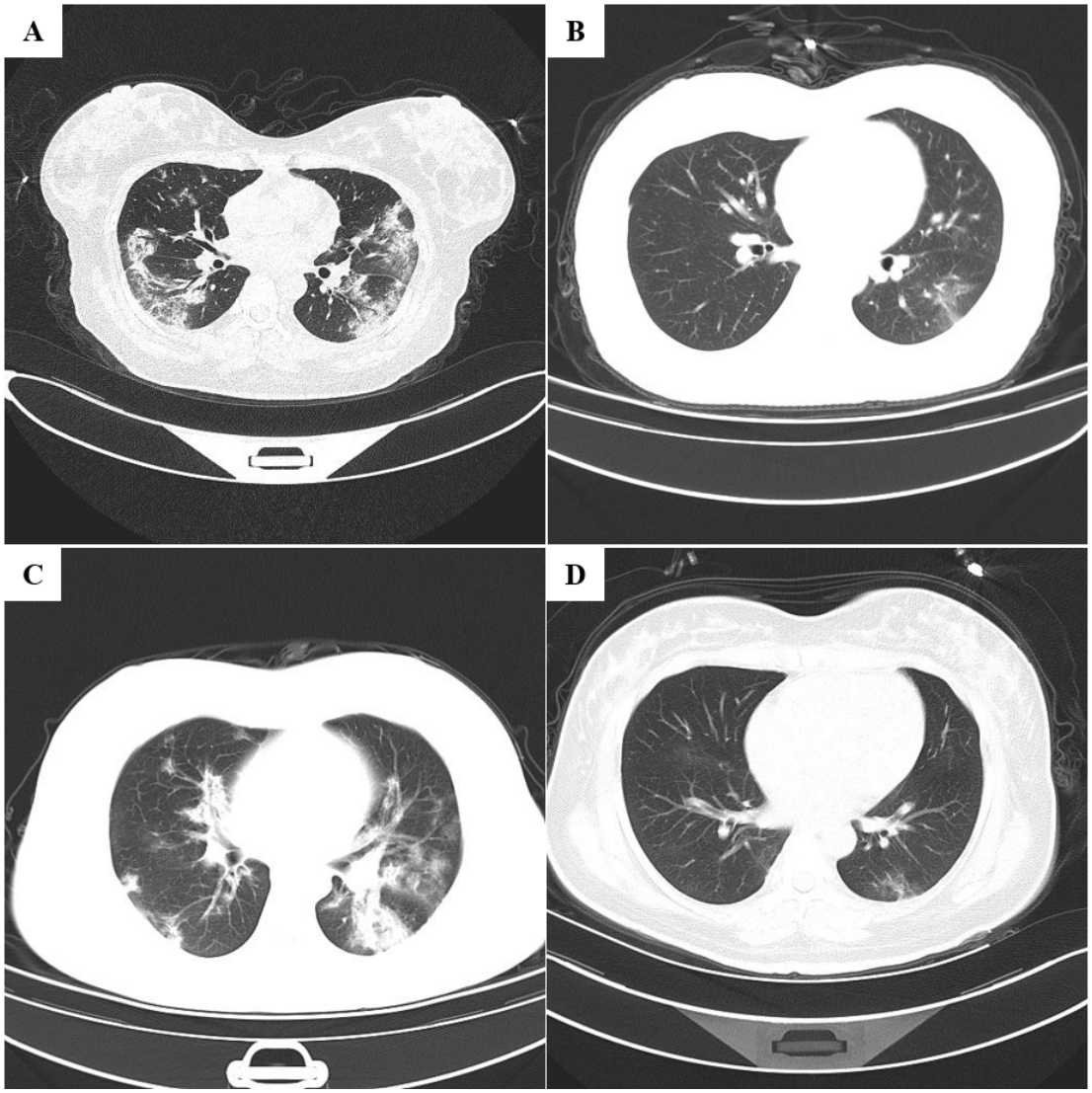
[a]N/A: not applicable.

In both pregnant and nonpregnant groups, we observed elevated high-sensitivity C-reactive protein (hs-CRP; ≥10 mg/L: 7/10, 70% vs 7/21, 33%) and erythrocyte sedimentation rate (>20 mm/hr: 5/5 100% vs 8/15, 53%). The mean concentrations of alanine aminotransferase (ALT) or aspartate aminotransferase (AST) in the pregnant group were above the normal range, while the nonpregnant group was normal. One patient in the pregnant group had ALT of up to 882 U/L and AST of up to 783 U/L. Concentration of lactate dehydrogenase (LDH) and alkaline phosphatase in the pregnant group were observed as higher than in the nonpregnant group (235.0 vs 193.0 U/L; 119.0 vs 48.0 U/L, respectively). Additionally, 92% (11/12) of pregnant women were observed with an elevated D-dimer level, which was significantly higher than nonpregnant women (2.1 vs 0.3 μg/mL, *P*<.001; Table 2).

Serological examination of pneumonia-associated pathogens was performed in pregnant and nonpregnant patients with COVID-19. There were 4/10 (40%) pregnant women and 8/15 (53%) nonpregnant women that tested positive for influenza A virus immunoglobulin M. Other respiratory viruses had not been observed. Of 10 pregnant women, 1 (10%) tested positive for legionella pneumophila and 1 (10%) tested positive for mycoplasma pneumoniae. Except for 1 pregnant woman who refused to undergo a chest CT scan, all patients accepted chest CT examinations. Of the 42 patients, 41 (98%) displayed typical findings of pneumonia, in which 9 (21%) patients had unilateral pneumonia and 32 (76%) patients had bilateral pneumonia (Table 2 and Figure 1).

In Figure 1, A and B are chest CTs showing the axial view lung window of 2 pregnant women with COVID-19. A is the chest CT from a 34-year-old woman who was 38 weeks and 4 days pregnant, showing multiple bilateral ground-glass opacities. B is a chest CT from a 30-year-old woman who was 39 weeks and 1 day pregnant, showing left-sided ground-glass opacity. C and D are chest CTs showing the axial view lung window of 2 nonpregnant women with COVID-19. C is a chest CT from a 30-year-old woman showing multiple bilateral ground-glass opacities. D is a chest CT from a 33-year-old woman showing left-sided ground-glass opacity.

Figure 1. Chest computed tomography scans of 4 patients with the coronavirus disease.



Management and Clinical Outcomes of Pregnant Women and Nonpregnant Women

A majority of the 17 pregnant and 26 nonpregnant patients with COVID-19 received antiviral (n=13, 76% vs n=25, 96%) and antibiotic (n=13, 76% vs n=23, 88%) therapy. There were 4 (24%) pregnant and 5 (19%) nonpregnant women that received glucocorticoid therapy, and 1 (6%) of the pregnant and 3 (12%) of the nonpregnant women received immunoglobulins therapy. Compared with the pregnant group, the proportion of patients that received antitussive therapy in the nonpregnant group significantly increased (18/26, 69% vs 6/17, 35%, *P*=.03). Additionally, oxygen support was administered in 35% (6/17) of pregnant and 54% (14/26) of the nonpregnant women with COVID-19. None of the patients underwent mechanical ventilation, continuous renal replacement therapy, or extracorporeal membrane oxygenation (Table 3).

None of the patients were lost in the follow-up during the study. None of the patients in the two groups were admitted to the intensive care unit (ICU), and none developed acute respiratory distress syndrome, disseminated intravascular coagulation (DIC), renal failure, heart failure, secondary bacterial pneumonia, or sepsis. In addition, none of the patients died (Table 3).

Table 3. Clinical treatment and outcomes of pregnant and nonpregnant women with the coronavirus disease.

Variables	Total (N=43)	Pregnancy (n=17)	Nonpregnancy (n=26)	*P* value
Management, n (%)				
Antiviral therapy	38 (88)	13 (76)	25 (96)	.14
Antibiotic therapy	36 (84)	13 (76)	23 (88)	.54
Glucocorticoid therapy	9 (21)	4 (24)	5 (19)	.96
Immunoglobulin	4 (9)	1 (6)	3 (12)	.93
Cough-suppressant therapy	24 (56)	6 (35)	18 (70)	.03
Oxygen support (nasal cannula)	20 (47)	6 (35)	14 (54)	.23
Mechanical ventilation	0 (0)	0 (0)	0 (0)	N/A[a]
Noninvasive	0 (0)	0 (0)	0 (0)	N/A
Invasive	0 (0)	0 (0)	0 (0)	N/A
Continuous renal replacement therapy	0 (0)	0 (0)	0 (0)	N/A
Extracorporeal membrane oxygenation	0 (0)	0 (0)	0 (0)	N/A
Clinical outcomes				
Intensive care unit admission, n (%)	0 (0)	0 (0)	0 (0)	N/A
Acute respiratory distress syndrome, n (%)	0 (0)	0 (0)	0 (0)	N/A
Disseminated intravascular coagulation, n (%)	0 (0)	0 (0)	0 (0)	N/A
Renal failure, n (%)	0 (0)	0 (0)	0 (0)	N/A
Heart failure, n (%)	0 (0)	0 (0)	0 (0)	N/A
Secondary bacterial pneumonia, n (%)	0 (0)	0 (0)	0 (0)	N/A
Sepsis, n (%)	0 (0)	0 (0)	0 (0)	N/A
Death, n (%)	0 (0)	0 (0)	0 (0)	N/A
Time of hospitalization (days), median (IQR)	22.0 (14.0-28.0)	17.0 (11.0-28.0)	22.0 (15.5-26.5)	.53
Time from onset to diagnosis (days), median (IQR)	9.5 (6.3-17.0)	4.0 (2.0-17.0)	10.0 (7.5-17.0)	.09
Time of viral shedding after onset of symptom (days), median (IQR)	25.0 (19.0-29.0)	24.0 (14.0-26.0)	26.0 (20.0-29.0)	.21
Redetected positive for discharged patients, n/N (%)	2/30 (7)	2/11 (18)	2/19 (11)	.61

[a]N/A: not applicable.

Of the pregnant women, 2 were classified with severe illness on admission, neither progressed to critical illness. In addition, no miscarriage was observed in the pregnant women. Out of 11 pregnant women, 10 underwent cesarean sections (2 had preterm birth; Multimedia Appendix 1).

The median length of hospitalization for the pregnant and nonpregnant groups was 17.0 and 22.0 days, respectively. In addition, the median interval from onset to diagnosis of SARS-CoV-2 were 4.0 and 10.0 days, respectively. The median duration of viral shedding after COVID-19 onset was 24.0 and 26.0 days, respectively. All patients who recovered from COVID-19 were placed in an isolation center for quarantine for a period of 2 weeks. SARS-CoV-2 was redetected in 11 pregnant and 19 nonpregnant women after discharge. SARS-CoV-2 was tested positive in 2 (18%) pregnant women and 2 (11%) nonpregnant women, and all were readmitted at hospitals for COVID-19 treatment (Table 3).

Discussion

Principal Results

This study retrospectively analyzes the epidemiological, clinical, laboratory, and imaging characteristics, and clinical outcomes of 43 women of childbearing age infected with COVID-19, including 17 pregnant women and 26 nonpregnant women. As of March 28, 2020, none of the patients involved in this study developed severe pneumonia or died. Based on our findings, there is currently no evidence indicating that pregnant women are more susceptible to the occurrence and severe adverse outcomes of COVID-19 than the general population.

A woman's body is in a highly immunosuppressive state after pregnancy, and the anatomy, physiology, and biochemistry will always change. For example, the immunity of T lymphocyte changes, the oxygen consumption increases, and the diaphragm elevates, which increases the risk of respiratory infection of pregnant women [9,10]. Studies during the outbreak of influenza virus and SARS-CoV have demonstrated that pregnant women

are more susceptible to severe illness. In the outbreak of the "Spanish flu" in 1918, 675,000 people died, with an overall mortality rate of 1%-2%, while 27% of pregnant women died, and the mortality rate of pregnant patients reached 50% or higher when complicated with secondary bacterial pneumonia [18,19]. In the outbreak of SARS in 2003, among 12 pregnant women diagnosed with SARS, 6 (50%) needed to be admitted to the ICU, 6 (50%) underwent mechanical ventilation, and the mortality rate was 25% [11]. Another study reported that 6 of 10 (60%) pregnant women with SARS were admitted to the ICU, 4 (40%) underwent mechanical ventilation, 3 (30%) progressed to renal failure, 2 (20%) progressed to secondary sepsis, 2 (20%) progressed to secondary DIC, and the mortality rate reached 30% [12]. During the COVID-19 outbreak in 2019, one study reported that none of the 9 pregnant patients progressed to critical illness or death [14]. Of the 16 cases of pregnant women with COVID-19, one was classified as severe but did not develop severe adverse outcomes in the later stage [20]. This is consistent with our findings that none of the pregnant women with COVID-19 developed severe adverse outcomes. Although critical pneumonia and death have not been reported in pregnant women, we should still be alert to the possibility of pregnant women developing severe adverse outcomes considering the high similarity of genomic sequence between SARS-CoV and SARS-CoV-2 [6].

Comparison With Prior Work

In this study, none of the 17 pregnant women had a history of exposure to the Huanan Seafood Market, 53% (n=9) had routine prenatal care within 2 weeks before onset, and 24% (n=4) had a family cluster of COVID-19. Therefore, during the epidemic, it was recommended that pregnant women delay their routine prenatal care for safety, unless it was necessary, or to use an online clinic to reduce the risks of nosocomial infection. Similar to previous studies, common symptoms at the onset of COVID-19 were fever and cough, and less common symptoms were expectoration, chest tightness, and diarrhea [14,15,21,22]. Notably, the onset of symptoms for several pregnant women were atypical, given that they had no fever or cough before diagnosis and only symptoms related to pregnancy were observed, including abnormal pain, vaginal bleeding, and increased or reduced fetal movement, which indicated that attention should be paid to the occurrence of atypical symptoms in pregnant women. Laboratory findings were significantly different in hematological parameters between the two groups. Leukocytosis featured prominently in pregnant patients [14,15], while leukopenia featured prominently in nonpregnant patients [21,22]. Lymphopenia is likely to occur in both groups. Elevated concentration of hs-CRP, D-dimer, and liver enzymes (including ALT, AST, LDH, and ALP) in pregnant patients with COVID-19 were observed; none of them developed liver failure or coagulation disorders. Recently, a study of 274 cases of patients with COVID-19 found that patients who were deceased generally had markedly higher level of CRP and LDH than recovered patients [22]. Therefore, the possibility that pregnant women with COVID-19 develop severe adverse outcomes cannot be eliminated. Additionally, a certain proportion of SARS-CoV-2 and influenza A virus coinfection were shown in the two groups. Given the similar clinical manifestations caused by the two viruses and a relatively low positive rate for the SARS-CoV-2 RNA test, it is recommended that a comprehensive assessment including epidemiological exposure, symptoms, laboratory, and imaging tests are necessary to the diagnosis of COVID-19.

Currently, vaccine or specific treatment for COVID-19 infection is absent. The majority of patients received antivirals such as arbidol and oseltamivir, and empirical antibiotics treatment, while few patients received glucocorticoid and immunoglobulin therapy. Arbidol is an antiviral agent with a unique mechanism of action targeting the S protein/angiotensin-converting enzyme 2 interaction and inhibiting membrane fusion of the viral envelope [23]. In vitro data suggested its activity against SARS [24]. In addition, a nonrandomized study of 67 patients with COVID-19 reported that, compared with arbidol-untreated patients, arbidol-treated patients with a treatment for a median time of 9 days showed a lower mortality rates (0% vs 16%) and higher discharge rates (33% vs 19%) [25]. However, limited data are available on the safety of medications used during pregnancy. Oseltamivir is a neuraminidase inhibitor approved for the treatment of influenza, but it has no documented in vitro activity against SARS-CoV-2. Antibiotics were used routinely after operation to prevent secondary bacterial infections. Routinely systemic corticosteroids for treatment of COVID-19 is not recommended [3]. A large proportion of nonpregnant women used antitussive drugs in this study, which was related to higher proportions of cough (77%) during disease progression (Multimedia Appendix 1). Supportive therapy and oxygen therapy are important for the management of COVID-19 [3].

No significant difference in the length of hospitalization for patients with COVID-19 patients was observed in the two groups. Notably, both pregnant and nonpregnant recovered patients tested positive for SARS-CoV-2 RNA during isolation. Fortunately, none of them experienced symptoms again or developed severe pneumonia. In a case series including 4 patients with COVID-19 who had 3 repeated qRT-PCR after discharge or discontinuation of quarantine, 4 (100%) redetected positive for SARS-CoV-2 RNA. All of them did not have contact with patients with suspected or confirmed COVID-19, and no family member was infected [26]. Thus, at least a proportion of recovered patients may still be virus carriers, and quarantine is still indispensable even after the patient with COVID-19 is discharged.

Limitations

Our study has some notable limitations. First, this study is limited by its small sample size. More cases of infection from COVID-19 should be used for analysis. Second, only 1 pregnant woman in her first trimester and 3 in their second trimester were included in this study. The effect of COVID-19 on mother and fetus in early pregnancy still needs to be clarified. Third, this is a retrospective study; the uncertainty of the exact dates and related information on exposure (recall bias) might have an inevitable impact on assessment. Fourth, this study only included pregnant women and nonpregnant women; another group of healthy pregnant women should be included to assess pregnant outcomes of mother and fetus, and intrauterine vertical transmission potential of COVID-19.

Conclusion

In this study, the clinical outcomes of pregnant women with COVID-19 appeared good, and none of the patients developed severe adverse outcomes. Additionally, the epidemiology of pregnant women with COVID-19 was complicated, and nosocomial infection cannot be underestimated. Fever and cough were the most common onset of symptoms in pregnant women. Notably, pregnancy-related symptoms (ie, abdominal pain, vaginal bleeding, increased or decreased fetal movement) might be the specific onset of symptoms for pregnant women with COVID-19. Quarantine is still needed after hospital discharge, as a small proportion of recovered patients may still be virus carriers. In conclusion, early detection and active management effectively helps in the risk of developing severe pneumonia and death in pregnant women with COVID-19.

Acknowledgments

We thank all the patients involved in this study and the nurses and clinical staff who provided care for the patients. This work was supported by the research grants from the National Key Research and Development Program of China (grant numbers 2018YFC1002900, 2020YFC0846300) and the Fundamental Research Funds for the Central Universities (grant number 2020kfyXGYJ00).

Authors' Contributions

SW and LF made substantial contributions to the study concept and design. LW and XG were in charge of the manuscript draft. SC, WZ, and JW were responsible for obtaining written consent from patients, obtaining ethical approval, collecting data, and confirming the data accuracy. XL, HZ, and LMS did the analysis and interpretation. LC was the pediatrician in charge of treatment of the newborn babies. All authors critically revised the manuscript for important intellectual content and gave final approval for the version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 19 KB - jmir_v22i8e19642_app1.docx]

References

1. Report of clustering pneumonia of unknown etiology in Wuhan City. Wuhan Municipal Health Commission. 2019. URL: http://wjw.hubei.gov.cn/fbjd/dtyw/201912/t20191231_1822343.shtml [accessed 2020-04-26]
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus InvestigatingResearch Team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
3. Clinical management of severe acute respiratory infection (SARI) when COVID-19 disease is suspected Interim guidance. World Health Organization. 2020 Mar 13. URL: https://www.who.int/docs/default-source/coronaviruse/clinical-management-of-novel-cov.pdf?sfvrsn=bc7da517_2 [accessed 2020-04-19]
4. Drosten C, Günther S, Preiser W, van der Werf S, Brodt H, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 2003 May 15;348(20):1967-1976. [doi: 10.1056/NEJMoa030747] [Medline: 12690091]
5. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 2012 Nov 08;367(19):1814-1820. [doi: 10.1056/NEJMoa1211721] [Medline: 23075143]
6. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020 Mar;579(7798):270-273 [FREE Full text] [doi: 10.1038/s41586-020-2012-7] [Medline: 32015507]
7. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. 2020. URL: https://who.sprinklr.com/ [accessed 2020-04-26]
8. Office of Health Emergency. National Health Commission of the People's Republic of China. 2020. URL: http://www.nhc.gov.cn/yjb/s7860/202004/2d391a171acc4624a50a1188c8de7361.shtml [accessed 2020-04-26]
9. Weinberger SE, Weiss ST, Cohen WR, Weiss JW, Johnson TS. Pregnancy and the lung. Am Rev Respir Dis 1980 Mar;121(3):559-581. [doi: 10.1164/arrd.1980.121.3.559] [Medline: 6998334]
10. Jamieson DJ, Theiler RN, Rasmussen SA. Emerging infections and pregnancy. Emerg Infect Dis 2006 Nov;12(11):1638-1643 [FREE Full text] [doi: 10.3201/eid1211.060152] [Medline: 17283611]

11. Wong SF, Chow KM, Leung TN, Ng WF, Ng TK, Shek CC, et al. Pregnancy and perinatal outcomes of women with severe acute respiratory syndrome. Am J Obstet Gynecol 2004 Jul;191(1):292-297 [FREE Full text] [doi: 10.1016/j.ajog.2003.11.019] [Medline: 15295381]
12. Lam CM, Wong SF, Leung TN, Chow KM, Yu WC, Wong TY, et al. A case-controlled study comparing clinical course and outcomes of pregnant and non-pregnant women with severe acute respiratory syndrome. BJOG 2004 Aug;111(8):771-774 [FREE Full text] [doi: 10.1111/j.1471-0528.2004.00199.x] [Medline: 15270922]
13. Assiri A, Abedi GR, Al Masri M, Bin Saeed A, Gerber SI, Watson JT. Middle East respiratory syndrome coronavirus infection during pregnancy: a report of 5 cases from Saudi Arabia. Clin Infect Dis 2016 Oct 01;63(7):951-953 [FREE Full text] [doi: 10.1093/cid/ciw412] [Medline: 27358348]
14. Chen H, Guo J, Wang C, Luo F, Yu X, Zhang W, et al. Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. Lancet 2020 Mar 07;395(10226):809-815 [FREE Full text] [doi: 10.1016/S0140-6736(20)30360-3] [Medline: 32151335]
15. Yu N, Li W, Kang Q, Xiong Z, Wang S, Lin X, et al. Clinical features and obstetric and neonatal outcomes of pregnant patients with COVID-19 in Wuhan, China: a retrospective, single-centre, descriptive study. Lancet Infect Dis 2020 May;20(5):559-564 [FREE Full text] [doi: 10.1016/S1473-3099(20)30176-6] [Medline: 32220284]
16. Diagnosis and treatment protocol for COVID-19 (sixth trial edition). General Office of the National Health Commission of China, General Office of the State Administration of Traditional Chinese Medicine. 2020. URL: http://www.nhc.gov.cn/yzygj/s7653p/202002/8334a8326dd94d329df351d7da8aefc2.shtml [accessed 2020-04-26]
17. Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases: interim guidance. World Health Organization. 2020 Mar 19. URL: https://www.who.int/publications-detail/laboratory-testing-for-2019-novel-coronavirus-in-suspected-human-cases-20200117 [accessed 2020-04-26]
18. Harris JW. Influenza occurring in pregnant women. JAMA 1919 Apr 05;72(14):978 [FREE Full text] [doi: 10.1001/jama.1919.02610140008002]
19. Taubenberger JK, Morens DM. 1918 Influenza: the mother of all pandemics. Emerg Infect Dis 2006 Jan;12(1):15-22 [FREE Full text] [doi: 10.3201/eid1201.050979] [Medline: 16494711]
20. Zhang L, Jiang Y, Wei M, Cheng BH, Zhou XC, Li J, et al. [Analysis of the pregnancy outcomes in pregnant women with COVID-19 in Hubei Province]. Zhonghua Fu Chan Ke Za Zhi 2020 Mar 25;55(3):166-171. [doi: 10.3760/cma.j.cn112141-20200218-00111] [Medline: 32145714]
21. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
22. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. BMJ 2020 Mar 31;368:m1295 [FREE Full text] [doi: 10.1136/bmj.m1295] [Medline: 32234718]
23. Kadam RU, Wilson IA. Structural basis of influenza virus fusion inhibition by the antiviral drug Arbidol. Proc Natl Acad Sci U S A 2017 Jan 10;114(2):206-214 [FREE Full text] [doi: 10.1073/pnas.1617020114] [Medline: 28003465]
24. Khamitov RA, Loginova SI, Shchukina VN, Borisevich SV, Maksimov VA, Shuster AM. [Antiviral activity of arbidol and its derivatives against the pathogen of severe acute respiratory syndrome in the cell cultures]. Vopr Virusol 2008;53(4):9-13. [Medline: 18756809]
25. Wang Z, Yang B, Li Q, Wen L, Zhang R. Clinical features of 69 cases with coronavirus disease 2019 in Wuhan, China. Clin Infect Dis 2020 Jul 28;71(15):769-777 [FREE Full text] [doi: 10.1093/cid/ciaa272] [Medline: 32176772]
26. Lan L, Xu D, Ye G, Xia C, Wang S, Li Y, et al. Positive RT-PCR test results in patients recovered from COVID-19. JAMA 2020 Feb 27:1502-1503 [FREE Full text] [doi: 10.1001/jama.2020.2783] [Medline: 32105304]

Abbreviations

ALT: alanine aminotransferase
AST: aspartate aminotransferase
CFR: case fatality rate
COVID-19: coronavirus disease
CT: computed tomography
DIC: disseminated intravascular coagulation
ELISA: enzyme-linked immunosorbent assay
hs-CRP: high-sensitivity C-reactive protein
ICU: intensive care unit
LDH: lactate dehydrogenase
MERS-CoV: Middle East respiratory syndrome–related coronavirus
qRT-PCR: quantitative real-time reverse transcription polymerase chain reaction
SARS-CoV: severe acute respiratory syndrome–related coronavirus

SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach; submitted 26.04.20; peer-reviewed by M Schoen, YH Yaw; comments to author 10.06.20; revised version received 03.07.20; accepted 26.07.20; published 24.08.20.

Please cite as:
Wei L, Gao X, Chen S, Zeng W, Wu J, Lin X, Zhang H, Mwamaka Sharifu L, Chen L, Feng L, Wang S
Clinical Characteristics and Outcomes of Childbearing-Age Women With COVID-19 in Wuhan: Retrospective, Single-Center Study
J Med Internet Res 2020;22(8):e19642
URL: http://www.jmir.org/2020/8/e19642/
doi:10.2196/19642
PMID:32750000

Original Paper

Grappling With the COVID-19 Health Crisis: Content Analysis of Communication Strategies and Their Effects on Public Engagement on Social Media

Cindy Sing Bik Ngai[1], BA, MPhil, PhD; Rita Gill Singh[2], MSc, MEd, DSocSci; Wenze Lu[1], BA, MA; Alex Chun Koon[3], BSc, PhD

[1]The Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, Hong Kong (China)

[2]The Language Centre, Hong Kong Baptist University, Hong Kong, Hong Kong (China)

[3]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong (China)

Corresponding Author:
Alex Chun Koon, BSc, PhD
School of Biomedical Sciences
Li Ka Shing Faculty of Medicine
The University of Hong Kong
21 Sassoon Road, Laboratory Block
Pok Fu Lam
Hong Kong
Hong Kong (China)
Phone: 852 39179500
Email: alexkoon@hku.hk

Abstract

Background: The coronavirus disease (COVID-19) has posed an unprecedented challenge to governments worldwide. Effective government communication of COVID-19 information with the public is of crucial importance.

Objective: We investigate how the most-read state-owned newspaper in China, People's Daily, used an online social networking site, Sina Weibo, to communicate about COVID-19 and whether this could engage the public. The objective of this study is to develop an integrated framework to examine the content, message style, and interactive features of COVID-19–related posts and determine their effects on public engagement in the largest social media network in China.

Methods: Content analysis was employed to scrutinize 608 COVID-19 posts, and coding was performed on three main dimensions: content, message style, and interactive features. The content dimension was coded into six subdimensions: action, new evidence, reassurance, disease prevention, health care services, and uncertainty, and the style dimension was coded into the subdimensions of narrative and nonnarrative. As for interactive features, they were coded into links to external sources, use of hashtags, use of questions to solicit feedback, and use of multimedia. Public engagement was measured in the form of the number of shares, comments, and likes on the People's Daily's Sina Weibo account from January 20, 2020, to March 11, 2020, to reveal the association between different levels of public engagement and communication strategies. A one-way analysis of variance followed by a post-hoc Tukey test and negative binomial regression analysis were employed to generate the results.

Results: We found that although the content frames of action, new evidence, and reassurance delivered in a nonnarrative style were predominant in COVID-19 communication by the government, posts related to new evidence and a nonnarrative style were strong negative predictors of the number of shares. In terms of generating a high number of shares, it was found that disease prevention posts delivered in a narrative style were able to achieve this purpose. Additionally, an interaction effect was found between content and style. The use of a narrative style in disease prevention posts had a significant positive effect on generating comments and likes by the Chinese public, while links to external sources fostered sharing.

Conclusions: These results have implications for governments, health organizations, medical professionals, the media, and researchers on their epidemic communication to engage the public. Selecting suitable communication strategies may foster active liking and sharing of posts on social media, which in turn, might raise the public's awareness of COVID-19 and motivate them to take preventive measures. The sharing of COVID-19 posts is particularly important because this action can reach out to a large audience, potentially helping to contain the spread of the virus.

(J Med Internet Res 2020;22(8):e21360) doi:10.2196/21360

KEYWORDS

COVID-19; communication; public engagement; social media; infodemiology; infodemic; message style; health content frames; interactive features; framework; content analysis

Introduction

Background

The first known coronavirus disease (COVID-19) case was reported in China on November 17, 2019 [1], and on January 23, 2020, the government in China imposed a strict lockdown in Wuhan, the epicenter of the virus. Despite a massive containment effort, by late February, 80,000 cases had emerged [2]. By March, COVID-19 was confirmed in many countries worldwide and the World Health Organization (WHO) declared COVID-19 a pandemic on March 11, 2020 [3,4]. Pandemics in the past such as the 2003 severe acute respiratory syndrome and H1N1 have had significant impacts on people's lives, socioeconomic activities, and population movement [5]. COVID-19 also presented similar impacts, but its spread was even faster [6]. A pandemic requires large-scale immediate actions by the government to connect with the public and a change in behavior of the public to combat the rapid spread of the disease [7]. For a new disease such as COVID-19, effective epidemic communication is crucial to inform the public about the latest updates of the disease, motivate them to adopt preventive measures to minimize the transmission of the disease, and reassure them that the government is capable of handling the situation [8-11]. Many studies on epidemic and pandemic communication exist on traditional media [8,12,13], suggesting that the public learns about the health risks associated with the pandemic from the media [14,15], which affects how they respond to the epidemic or pandemic [16]. In recent years, social media has played an increasingly important role in promoting health risk communication during an epidemic [17,18]. Research on the use of social media to investigate public attention to new epidemics has been conducted, such as with H7N9 [19-21], the Ebola outbreak [9], and the H1N1 pandemic in 2009 [22]. However, there are few studies that have adopted social media analysis in examining government media communication with the public and the public's response to the new COVID-19 epidemic [10,11]. Because timely public action is needed to contain the spread of the new disease, it is of urgent importance to investigate how the government media communication engages the public. This information can provide insights on what the media, health organizations, and government can further do to disseminate information to the public so that the latter can take appropriate measures to stem the spread of the virus.

In terms of what organizations emphasize in their epidemic or pandemic communication, a prior study [22] found that most corporate and government organizations in the United States relied on the content frames of health crises, health issues, and disasters in communicating messages about the 2009 H1N1 flu pandemic with the public. Government organizations were more likely than corporate organizations to frame the H1N1 pandemic as a general health issue and emphasized uncertainty, disease detection, and preventive measures [22]. The style of communication can also have an impact on public engagement in that a narrative style has a positive effect on preventive and detection health behaviors [23] and arguments and facts may be used too [23]. Researchers have pointed out that narratives promote health behavior change [24,25], yet there is a lack of research on the use of narratives in pandemics for effective health communication, apart from Sandell et al [13] who revealed that positive narratives were effective in raising the public's awareness of health risks and the preventive measures to curb the spread of the 2009 H1N1 pandemic [13]. Additionally, interactions on social media can affect health behavior and attitudes [26], and thus, the creation of the dialogic loop via the use of interactive features [27,28] is important on social media. This can be done by allowing the public to post questions and receive feedback [27,29] and using interactive features such as multimedia and hashtags [30]. A previous study has found a positive relationship between chief executive officers' (CEOs) use of hashtags and public engagement with respect to likes, shares, and comments on social media [28]. However, a research gap exists in understanding the interactive features used by the government in its communication with the public on social media with regard to a pandemic.

Synthesizing this literature, our study was guided by the observation that there is scant research on the use of social media to disseminate information about COVID-19 and public engagement with this information [10,11]. In particular, there is a research gap in understanding the content frames employed by the government's media in the Chinese context, its style of communication, and the use of interactive features in its communication with the public regarding a new epidemic. Therefore, in this study, we investigated how the most-read government-owned newspaper in China, *People's Daily*, serving as the main vehicle for the government's dissemination of information to the public, employed a social networking site, Sina Weibo, to communicate and possibly engage the public with COVID-19.

As of 2014, China had 649 million internet users [31]. To use the power of the internet, the main Chinese state-owned media such as *People's Daily* and *CCTV News* have shifted the paradigm of media coverage by placing more emphasis on communication with the public via social media [32]. They have also switched to a more interactive style to better connect with the public [32]. In China, where Facebook is blocked, Weibo, a social media platform under *People's Daily* introduced by the Chinese commercial corporation Sina, has taken over and become the largest social media network [33]. In 2018, Weibo had over 462 million active users [34] and was used by approximately 200 million people every day [35].

With years of the government's continued efforts, the reputation of the Chinese state-owned media has improved significantly in the eyes of the Chinese public [36]. State-owned media such

as *People's Daily* now maintain a strong web presence and a user-friendly image rather than an authoritative image [37]. *People's Daily* encourages its audience to participate in discussions and demonstrates a strong tendency to adopt positive and persuasive messages [38]. For example, on a topic of haze-related issues, instead of providing pictures of haze with a negative valence, *People's Daily* posted positive images that encouraged the public to appreciate the beauty of nature, accompanied by persuasive messages that suggested substantial improvements to be made in the future [38]. This is vastly different from how *China Daily* handled the same topic, which displayed a cartoon of Santa Claus hitting a tree due to haze [38]. This example demonstrates that the state-owned media in China, *People's Daily*, and its online platform, Sina Weibo, have actively adapted their styles of interactive communication to better engage the public.

In terms of health emergency communication, previous studies have found that social media platforms such as Twitter and even the photograph-based Instagram played a significant role in guiding the public during the Zika virus outbreak in 2016 [18,39]. For China, Sina Weibo performs a similar role during pandemics since the government, news media, and the public heavily relied on it as an online platform for communicating information during the current COVID-19 outbreak [11]. Sina Weibo serves as a pivotal communication platform for the government to interact with the public and disseminate information about COVID-19, such as its symptoms, preventive measures, and adopted health policies [11]. Therefore, we contend that *People's Daily* would also communicate information about COVID-19 and interact with the public on its social media platform, Sina Weibo. In our study, we integrated factors, including health crises framing in the media context [22,40-42], message style in health communication [13,23], and interactive features [27,28] to examine epidemic communication and public engagement in China. We then developed an integrated framework to investigate the relationship between these factors and the levels of public engagement. Since our study also investigated public engagement in the form of likes, comments, and shares, it might offer insights on how effectively social media platforms such as Weibo can be used for epidemic communication.

Developing an Integrated Framework

The WHO has advised governments to take proactive steps to communicate with the public about epidemics, as the sharing of critical information about the epidemic can minimize the spread of the disease and foster the public's collaboration with the government [10,43,44]. Social media serves as a major communication platform for the government and public health authorities to provide timely health information to the public [11,22,45-47]. The contribution of this study is that we incorporated three key dimensions in health emergency communication on social media, namely, the framing of health crises and issues [22,40-42], message style [13,23], and the interactive loop [27,28] to examine COVID-19 communication by the government-owned media and public engagement in China. Our findings shed light on how responses to the epidemic are framed by the media and what encourages the public to engage with such communication and take appropriate actions to slow the spread of the virus. In the following, we explain the three dimensions adopted in our study: content frame, message style, and interactive features.

Content Frame Dimension

Communication related to health risks depends on persuasion for the framing of the message that informs the public about important information and motivates them to act [48]. Framing refers to how a text or message defines an issue and provides the necessary context [49,50]. Entman [51] pointed out that "to frame is to select some aspects of a perceived reality and make them more salient in a communicating text." Drawing on framing analysis, one can identify how organizations and the government frame their messages pertaining to critical issues for the public [52], thereby impacting the effectiveness of the information disseminated [53].

In the management of a health crisis, the media and government tend to employ six frames in message delivery: conflict (aspects of crises that bring tensions between parties), action (past or current crisis response actions), consequence (the effects or severity of the crisis), new evidence (discovery of new evidence that contributes to the crisis understanding), uncertainty (aspects such as the spread of the epidemic, treatment, and what is unknown), and reassurance (reassuring the public) [22,41]. When handling communication of health issues, five frames in the delivery of health messages are noted, namely, disease detection (symptoms to indicate how the disease is spreading), disease prevention (taking preventive measures), health care services (the actions that the health care system is taking), scientific advances (discovery of new evidence showing how the disease is spread), and lifestyle risk factors (personal habits that are likely to lead to the disease) [22,40,42]. In the application of these frames, Liu and Kim [22] noted that most corporate and government organizations in the United States used the frames of health crises and health issues much more via traditional media than social media in disseminating messages about the 2009 H1N1 flu pandemic [22]. Yet corporate organizations framed the pandemic as a health crisis rather than as a general health issue, meaning that they did not emphasize the long-term actions that could prevent the health issue from arising in future. In addition to this, Liu and Kim [22] noted that government organizations were more likely to use uncertainty subsumed under the health crisis frame whereas corporate organizations tended to use the conflict indicator [22]. In another study, Shih et al [41] noted that the frames of governmental action and consequence were predominantly used by journalists to craft stories about epidemics including mad cow disease, West Nile virus, and avian flu in the print version of New York Times [41].

Given that COVID-19 was a health crisis and health issue emerging in China and required immediate action from the public, we contend that framing this epidemic using the health crisis frames of action, new evidence, uncertainty, and reassurance, would be of relevance to communication with the public, while the frame of health issues, namely, disease prevention and health care services, are of salient importance too since information is lacking on the details and duration of the epidemic. As highlighted by Shih et al [41], the government

may attempt to minimize loss by reassuring the public with actions and new evidence via its influence on the media and its frames. Therefore, the previously mentioned frames could be effectively used in the media coverage of the epidemic. For a new epidemic, vaccines and medicine are not available to the public, so disease detection and scientific advances are tasks that only medical professionals can undertake and, thus, may not be able to engage the public. Disease prevention is vital and includes information about what preventive measures the public should adopt to reduce the risk of infection [12,43]. A prior study [22] found that government organizations in the United States were more likely to incorporate uncertainty into their crisis responses to the H1N1 pandemic, and with the implications of their results, we incorporated uncertainty into our framework, since the newspaper we examined is the main vehicle used by the government in China to communicate with the public. Uncertainty is useful because by indicating what is unknown, more transparency of information is provided, possibly generating trust [14,43]. Conflict was primarily used by corporate organizations as opposed to government organizations for the H1N1 pandemic in the United States [22] and, thus, not deemed of specific value in our framework.

The frames that we employed are in line with the information that the WHO recommends that the media should provide to the public: offering accurate and transparent information to the public; encouraging appropriate attitudes, actions, and behaviors; and helping prevent unnecessary fear [44]. As a result, we combined the eleven frames of health crises and issues into six frames for the investigation of COVID-19 content frames in social media posts, namely, *(1) action*, *(2) new evidence*, *(3) reassurance*, *(4) disease prevention*, *(5) health care services*, and *(6) uncertainty*. Since these frames are all content-related, we termed them *subdimensions* under the *content frame* dimension.

Message Style Dimension

Since a key objective of epidemic communication is to persuade the public to change their behavior to limit the spread of the disease [11] while the public has a need for real-time information [47], effective messages need to be designed, requiring some form of appeal. In this regard, the effectiveness of narratives in health communication on disease detection and prevention has been explored [54-56]. Narratives refer to stories that people use and tell, and consist of anecdotes and personal stories with plots [23,24]. Narratives engage the public because they make them concentrate on the story events instead of disputing the presented information while eliciting emotional reactions and being both entertaining and informative [23,46,55,57,58]. On the other hand, nonnarrative messages depend on the use of arguments and facts presented logically and are considered as informative [23].

Studies on the effectiveness of narratives in brand advertisement connection with customers and in the area of health communication have been conducted [58-60]. For example, a narrative film was effectively employed to communicate the need for vaccination against the human papillomavirus [25]. Scholars have increasingly recognized the role of narratives in promoting health behavior change [24], but studies on the use of narratives in pandemics for effective health communication are scarce with the notable exception of Sandell et al [13], who found that positive narratives were powerful in raising the public's awareness of health risks and preventive measures for the 2009 H1N1 pandemic [13]. Based on this, we categorized *narrative* and *nonnarrative* as subdimensions under the *message style* dimension.

Interactive Features Dimension

The interaction (ie, one-to-one or one-to-many) on social media sites can influence health behavior and attitudes [26], and consequently, the promotion of the dialogic loop with interactive features [27,28] is crucial on such sites. An interactive dialogic loop allows the public to post questions and receive feedback as well as post comments and share them [20,27]. A wide range of interactive features are available on sites, including multimedia (eg, videos, audio, photos, podcasts), stay-up-to-date tools such as hashtags, and comments on content [30]. Hashtags enable users to find relevant shares on an issue [61] and facilitate in making synchronous conversations on Twitter, thereby fostering engagement [62], with a study noting a positive relationship between CEOs' use of hashtags and engagement in terms of likes, shares, and comments [28]. To encourage users to return to the site, an attractive site and relevant links are necessary. Regarding conservation of visitors, the site should include useful external links [27]. In health-related communication, it is known that social media posts with interactive features leave a deep impression on the public when compared with posts in plain text [63]. Hence, we assigned "interactive features" as the third dimension, comprising the four subdimensions of *(1) links to external sources*, *(2) use of hashtags*, *(3) use of questions to solicit feedback*, and *(4) use of multimedia*.

Although prior studies have recognized the importance of the content frames [22,40-42], message style [13,23], and interactive features [27,28] in health-related communication, the question as to whether these three dimensions can facilitate the communication of COVID-19 on the government's social media platform remains unclear. Therefore, our first research question (RQ) is derived:

- RQ1: How frequently did the official social media employ the subdimensions of content frames, message style, and interactive features in its communication of COVID-19?

A clearer indication of the public's awareness of the information communicated by the government can be revealed through their actions of liking, sharing, and commenting on the government's posts. Therefore, it is pertinent to investigate the effects of the content frames, message style, and interactive features on different levels of public engagement [62,64,65]. Social media users may use "likes" to indicate their interest in a health issue [66,67], and by commenting and sharing, the public can let others know that the issue is important, thereby serving as disseminators of the original message posted [9]. To investigate differences in public engagement with health information posted by the government in response to COVID-19, our second RQ is posed:

- RQ2: Did the subdimensions of content frames, message style, and interactive features have different levels of impact on public engagement?

Different dimensions may function synergistically to impact public engagement. As has been found in a study, an interaction effect between content and style of communication on public engagement in brand social media communication was observed [68]. It is, therefore, likely that interaction effects might exist between some of the dimensions or subdimensions on public engagement in COVID-19 communication. Thus, our third RQ is as follows:

- RQ3: Could the dimensions (ie, content frames, style, and interactive features) or subdimensions interact synergistically to increase or decrease the levels of public engagement with the government's communication of COVID-19?

By examining the impact of content frames, message style, and interactive features on public engagement in COVID-19 communication, our study aims to provide meaningful and critical information for governments, health organizations, communication professionals, and researchers regarding the health emergency communication strategies employed and their effectiveness in raising the public's awareness of and urgent need for taking preventive measures against COVID-19.

Methods

Data Collection

We selected the government-owned social media platform *People's Daily*'s Sina Weibo account for data collection. *People's Daily* is the official newspaper of the Central Committee of the Communist Party of China [69] for disseminating government information to the Chinese public [70]. It is the most influential and authoritative newspaper in China, having a circulation of 3 million, and is ranked as one of the world's top 10 newspapers [71]. With 117 million followers, Sina Weibo of *People's Daily* is also one of the top followed and most visited Sina Weibo sites in China. Due to the prominent use of Sina Weibo for social media communication in China [34] with 462 million active online users in 2018, we captured all posts and the public's responses communicated between the government and public on COVID-19 from *People's Daily* for the investigation of government communication of COVID-19 and its interaction with the public.

Sample Period

A text corpus containing all posts on Sina Weibo of *People's Daily* pertaining to COVID-19 from January 20, 2020, to March 11, 2020, was constructed. The sampled period began on January 20, 2020, when the Chinese State Council officially announced the management of COVID-19 as a public health emergency issue and the corresponding preventive measures were launched to tackle COVID-19 [72,73]. The sampled period ended on March 11, 2020, when the WHO declared the COVID-19 outbreak a pandemic, meaning that the regional epidemic had become a global public health emergency [4]. Subsequently, all online posts related to COVID-19 were manually extracted from Sina Weibo's account of *People's Daily*, and in total, 3255 posts were collected.

Sample Size and Sample Data Collection

To generalize a sample size to represent the target population (3255 posts), we employed the sample size calculator developed by the Australian Statistics Bureau to estimate a sample size of 620, giving a confidence level of 95%, a confident interval of 0.035, and a standard error of 0.018. A random sampling method was employed. The 620 posts and their corresponding public responses (ie, number of shares, comments, and likes) on *People's Daily*'s Sina Weibo account from January 20, 2020, to March 11, 2020, were harnessed for quantitative content analysis. To systematically detect statistically valid outliers, we employed z score to quantify the unusualness in the observations [74]. There were 12 posts (2%) identified as outliers and removed from the data pool. These outliers included posts that were significantly longer or shorter, which would have otherwise caused problems during content analysis, as the length of the posts would affect the number of counts in content themes, style, and interactive features. Consequently, 608 posts and the related public responses were included in the corpus for content analysis.

Content Analysis and Coding Scheme

Content analysis was employed to examine COVID-19 communication in the 608 posts of *People's Daily*'s Sina Weibo. Content analysis is a widely employed method in the study of technical and media communication [75]. It is concerned with the context in which the occurrences of words, phrases, signs, and sentences are recorded and analyzed to provide an in-depth understanding [75]. Researchers can design a variety of categories based on their interactions with the data to develop an integrated framework for quantitative studies [76]. Content analysis can be applied to "virtually any form of linguistic communication to answer the classic questions of who says what to whom, why, how, and with what effect" [77]. Therefore, it is well-suited to a coding operation involving a developed framework in the media communication context [78]. Through an in-depth analysis of mainstream media communication, we were able to reveal and establish the relationship between the variables in the proposed conceptual framework.

First, to answer RQ1, we drew on previous studies of epidemic communication, health crisis communication, and public relations studies [11,40] to code the topics of the content dimension exhibited in the government's COVID-19 communication into the following six subdimensions on a sentence basis: *(1) action*, *(2) new evidence*, *(3) reassurance*, *(4) disease prevention*, *(5) health care services*, and *(6) uncertainty*. Second, to examine the communication styles of COVID-19 posts from *People's Daily*, we built on prior studies [55,58] and coded the two message styles in the style dimension into the subdimensions of *(1) narrative* and *(2) nonnarrative* on a sentence basis. To determine whether the narrative style of communication was employed, we examined if the post had a temporal or spatial sequence and revealed the writer's feelings or thoughts. Last, we built on prior public relations studies [27,38,68] and coded the number of interactive features used to facilitate the creation of the interactive dialogic loop. These

interactive features included *(1) links to external sources, (2) use of hashtags, (3) use of questions to solicit feedback*, and *(4) use of multimedia* (see Multimedia Appendix 1 for the exemplifications of coding items and examples extracted from the collected posts).

Regarding RQ2 and 3, we recoded the dimensions of content, style, and interactive loop using the dominant category for performing the analysis of variance (ANOVA) tests on content, style, and interactive loop on public engagement. For example, we found 43% of the sentences in post number 128 belonging to *action* and 29% to *disease prevention*; 57% of sentences employed a narrative style of communication while 43% were nonnarrative; 1 link provided an external source, 2 pairs of hashtags, and 1 multimedia feature. We then recorded the content as *action* based on the dominant content topic, style as *narrative* based on the dominant use of narrative sentences, and interactive loop as *use of hashtags* based on the dominant use of hashtags. If the count of sentences or interactive features was the same, the primary coder checked the title, topic sentences, and context of the post to determine the dominant category.

To address RQ2 and 3, we recorded the number of shares, comments, and likes of the sampled posts to investigate the relationship between *People's Daily* communication and its impact on public engagement. Regarding the negative binomial regression (NB2) analysis, the coding results of RQ1 were adopted to investigate the effect of all subdimensions on public engagement. The number of shares, comments, and likes in RQ2 and 3 were also included for statistical analyses.

Intercoder Reliability

The coding was conducted by the third author (the primary coder) and a well-trained coder who all possess a postgraduate degree in communication. To ensure intercoder reliability on the coding of dimensions, subdimensions, and public engagement, the coder was repeatedly trained on the coding scheme. Any disagreement between the author and coder was discussed in the coding process. The measure of intercoder reliability was based on the co-coding of 120 posts from the data pool of 608 posts (19% of the total number of posts sampled) [75]. For all categories, the average agreement was higher than 0.83, and the average Cohen kappa was greater than 0.8, indicating an almost perfect agreement [76] (see Multimedia Appendix 2 for intercoder checking results of all categories).

Statistical Analyses

To analyze the differences in the frequencies of the use of each subdimension in the communication of COVID-19–related news by the official social media (RQ1), we coded the presence of subdimensions in each of the 608 posts and then calculated the mean counts for each of the 12 subdimensions. We then employed the one-way ANOVA and the post-hoc Tukey test in SPSS (IBM Corp) to reveal the differences in the use of content, style, and interactive features in COVID-19 social media communication (RQ1) and the difference in the number of shares, comments, and likes in relation to the subdimensions of content, style, and interactive dimensions (RQ2). The two-way ANOVA was performed to examine the interaction effect of content and style on public engagement in the form of shares, comments, and likes (RQ3).

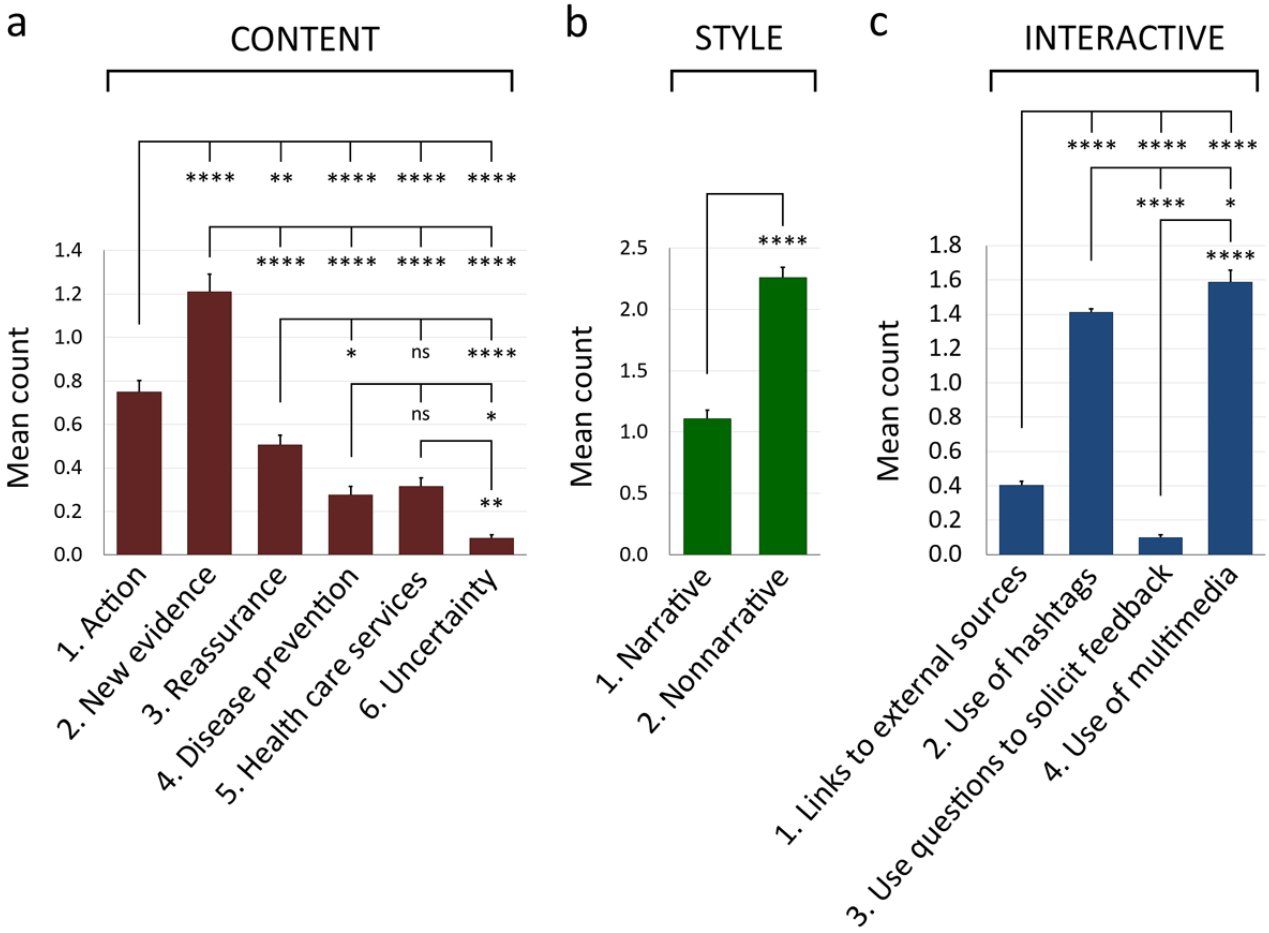
To test the assumptions of normality in ANOVA, we performed the Kolmogorov-Smirnov and Shapiro-Wilk tests on the normality of the variables. Most variables were not normally distributed, but we decided to continue using ANOVA as it has proven to be robust and valid in testing the difference between independent variables, even if the normality assumption is violated [77]. In addition, we conducted the test of homogeneity of variances when performing ANOVA. When the assumption of homogeneity of variances was violated, the ANOVA results were replaced with those of the Welch ANOVA.

As for RQ2, which involved examining the relationship between the 12 subdimensions (independent variables) and the public's responses in terms of the count number of shares, comments, and likes (dependent variables), we first employed Poisson regression, a count regression model in SPSS [78,79]. However, real-world data sets are commonly known to violate the assumption in the Poisson regression with respect to overdispersion of outcome variables [80]. As expected, such a violation was detected in our data set, and thus, we followed the common practice of replacing Poisson regression with the NB2 [80] to improve the goodness of fit, especially Akaike information criterion and bayesian information criterion. NB2 is effective in fitting various types of data arising in technical and communication research [81], and NB2 is a more general model that relaxes the strong assumption that the underlying rate of the outcome is the same for each included participant [81].

Results

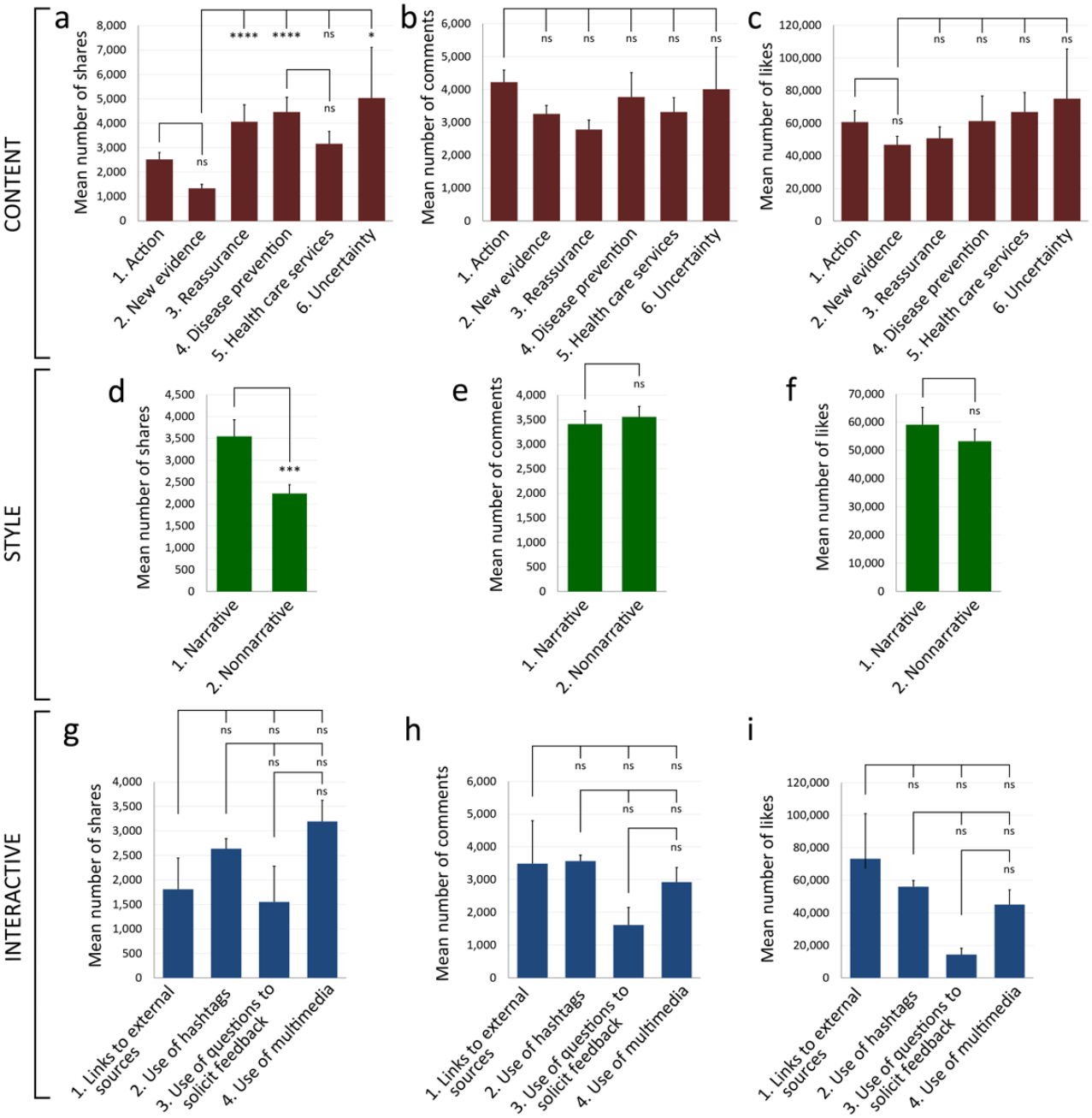
In response to RQ1 regarding the differences in the frequencies of each subdimension's use in the communication of COVID-19–related news by the official social media, we found that *new evidence* in the content dimension was the most used subdimension (mean 0.749, standard error of the mean [SEM] 0.05) and significantly used much more than any other subdimensions (Figure 1a). *Action* was the second most prevalent subdimension (mean 1.210, SEM 0.08), and *reassurance* was the third most frequently used one (mean 0.506, SEM 0.05). *Disease prevention* (mean 0.276, SEM 0.04) and *health care services* (mean 0.315, SEM 0.04) ranked fourth and fifth respectively, and *uncertainty* was the least used subdimension (mean 0.077, SEM 0.02; Figure 1a). In relation to the style dimension, the *nonnarrative* (mean 2.259, SEM 0.09) style was used approximately twice as much as the *narrative* (mean 1.110, SEM 0.07) one (Figure 1b). Concerning the interactive dimension, the *use of multimedia* (mean 1.586, SEM 0.08) and *use of hashtags* (mean 1.411, SEM 0.02) were the most prevalent subdimensions, with the *use of multimedia* being slightly but significantly higher than that of the *use of hashtags* (Figure 1c). By contrast, both *links to external sources* (mean 0.402, SEM 0.02) and *use of questions to solicit feedback* (mean 0.097, SEM 0.02) were used infrequently, with the *use of questions to solicit feedback* being used significantly less in comparison to all other subdimensions (Figure 1c).

Figure 1. Comparison of the mean counts of subdimensions within each of the three dimensions. (a) Mean counts of subdimensions under the content dimension. (b) Mean counts of subdimensions under the style dimension. (c) Mean counts of subdimensions under the interactive dimension. *P<.05, **P<.01, ***P<.001, ****P<.0001. All histograms depict mean and standard error of the mean.



Regarding the levels of impact on public engagement from individual subdimensions in COVID-19 social media posts (RQ2), our results showed that posts of *new evidence* generated the least number of shares of all six subdimensions (mean 1327.81, SEM 165.90). Posts of *new evidence* had significantly fewer shares than posts of *reassurance* (mean 4065.32, SEM 689.88), *disease prevention* (mean 4455.71, SEM 604.95), and *uncertainty* (mean 5033.35, SEM 2242.13; Figure 2a). However, the six subdimensions of content did not show differences in their impact on comments and likes (Figure 2b, c). For the style dimension, *narrative* posts generated significantly more shares than *nonnarrative* posts (*narrative*: mean 3544.03, SEM 379.80 vs *nonnarrative*: mean 2237.06, SEM 204.18; Figure 2d). Similar to content, the message style did not exert any impact on the number of comments and likes (Figure 2e, f). As for the interactive dimension, no significant differences were observed among the four subdimensions in terms of shares, comments, and likes (Figure 2g-i). Surprisingly, although they were the most frequently used subdimensions (Figure 1a, b), *new evidence* and the *nonnarrative style* had the least impact on the number of shares in their own dimensions (Figure 2a, d).

Figure 2. Comparison of the mean number of shares, comments, and likes for posts of each subdimension. Mean number of (a) shares, (b) comments, and (c) likes of the subdimensions of the content dimension. Mean number of (d) shares, (e) comments, and (f) likes of the subdimensions of the style dimension. Mean number of (g) shares, (h) comments, and (i) likes of the subdimensions of the interactive dimension. *P<.05, **P<.01, ***P<.001, ****P<.0001. All histograms depict mean and standard error of the mean.



To determine which of the twelve subdimensions were effective positive or negative predictors of public engagement in COVID-19 communication (RQ3), we fitted the share, comment, and like count data to a NB2 model. Our results in Figures 1 and 2 indicated that, although *new evidence* was the most used content subdimension (Figure 1a), its posts received the lowest number of shares (Figure 2a), suggesting a negative correlation between *new evidence* and the number of shares. In our NB2 analysis, we confirmed that *new evidence* was a strong negative predictor of the number of shares (β=–.253, SE 0.068, P<.001; Table 1). Similarly, *nonnarrative* was the most frequently used style (Figure 1b) but generated a lower number of shares as opposed to the *narrative* one (Figure 2d). Again, we confirmed that the *nonnarrative* style was indeed a strong negative predictor of the number of shares (β=–.223, SE 0.068, P<.001; Table 1). By contrast, the *narrative* style was found to be a strong positive predictor of the number of shares (β=.283, SE 0.064, P<.001; Table 1). For the interactive dimension, *links to external sources* was a strong positive predictor of the number of shares (β=.319, SE 0.087, P<.001), whereas the *use of multimedia* was a weak positive predictor of the number of shares (β=.057, SE 0.023, P=.02; Table 1). Finally, we noted that the *use of questions to solicit feedback* was a strong negative predictor of the number of comments (β=–.177, SE 0.087, P=.04) and likes (β=–.290, SE 0.111, P=.01; Table 1).

Subdimensions are likely to function synergistically in affecting public engagement. To examine whether there was an interaction

among the dimensions on public engagement (RQ3), we performed a two-way ANOVA analysis on the mean number of shares, comments, and likes of the dimensions. Our results confirmed a significant interaction effect between content and style on the number of likes (Table 2). However, there was neither any interaction effect between content and the interactive dimension itself nor between style and the designated interactive dimension (Table 2).

Table 1. Identification of positive and negative predictors of the number of shares, comments, and likes using a negative binomial regression model.

Dimensions and subdimensions	Shares				Comments				Likes			
	β	SE	95% CI	*P* value	β	SE	95% CI	*P* value	β	SE	95% CI	*P* value
Content												
Action	–.071	0.068	0.816-1.063	.29	.096	0.063	0.973-1.244	.13	.049	0.080	0.899-1.228	.53
New evidence	–.253	0.064	0.685-0.881	*<.001* [a]	.003	0.060	0.893-1.128	.95	–.053	0.077	0.816-1.102	.49
Reassurance	–.053	0.072	0.824-1.092	.46	–.080	0.066	0.812-1.049	.22	–.059	0.086	0.797-1.116	.495
Disease prevention	.057	0.078	0.909-1.234	.46	–.019	0.074	0.848-1.134	.79	–.054	0.097	0.783-1.145	.57
Health care services	–.097	0.077	0.780-1.055	.21	.029	0.067	0.902-1.174	.67	.070	0.089	0.902-1.276	.43
Uncertainty	–.090	0.138	0.697-1.197	.51	–.012	0.115	0.788-1.239	.92	.019	0.150	0.759-1.368	.90
Style												
Narrative	.283	0.064	1.170-1.506	*<.001*	.069	0.055	0.961-1.194	.21	.129	0.074	0.984-1.316	.08
Nonnarrative	–.223	0.068	1.094-1.427	*.001*	.037	0.061	0.921-1.169	.54	.108	0.078	0.955-1.299	.17
Interactive												
Links to external sources	.319	0.087	1.160-1.633	*< .001*	.022	0.071	0.889-1.175	.76	.088	0.090	0.915-1.303	.33
Use of hashtags	.079	0.081	0.923-1.268	.33	.059	0.070	0.925-1.216	.40	.016	0.092	0.848-1.217	.87
Use of questions to solicit feedback	–.121	0.106	0.720-1.092	.26	–.321	0.090	0.608-0.865	*<.001*	–.463	0.116	0.501-0.790	*<.001*
Use of multimedia	.057	0.023	1.011-1.108	*.02*	–.010	0.022	0.948-1.033	.64	–.046	0.029	0.903-1.011	.11

[a]Italics indicate a significant relationship.

Table 2. Test of interaction effect between content, style, and the interactive dimension.

Interactions	Shares					Comments					Likes				
	df	MS[a]	*F* test	*P* value	η_p^2	*Df*	MS	*F* test	*P* value	η_p^2	*df*	MS	*F* test	*P* value	η_p^2
Content x style															
Content	5	74,362,796.0	3.627	.003	0.0295	5	13,679,259.3	0.804	.55	0.0067	5	1,600,144,996.5	0.213	.96	0.0018
Style	1	2,955,797.2	0.144	.70	0.0002	1	644,487.0	0.038	0.85	0.0001	1	856,248,211.1	0.114	.74	0.0002
Content x style	5	14,294,815.4	0.697	.63	0.0058	5	35,061,851.2	2.060	.07	0.0170	5	17,759,392,993.4	2.367	*.04* [b]	0.0194
Error	597	20,504,947.6	N/A[c]	N/A	N/A	597	17,018,927.7	N/A	N/A	N/A	597	7,502,954,304.2	N/A	N/A	N/A
Content x interactive															
Content	5	14,588,523.9	0.706	.62	0.0059	5	7,622,337.7	0.452	.81	0.0038	5	2,892,590,073.7	0.384	.86	0.0032
Interactive	3	12,299,580.7	0.595	.61	0.0030	3	15,422,984.7	0.914	.43	0.0046	3	8,720,670,460.3	1.158	.33	0.0059
Content x interactive	9	7,690,008.8	0.372	.95	0.0056	9	12,698,088.8	0.753	.66	0.0114	9	6,061,808,678.1	0.805	.61	0.0121
Error	590	20,657,294.6	N/A	N/A	N/A	590	16,865,817.6	N/A	N/A	N/A	590	7,532,498,681.0	N/A	N/A	N/A
Style x interactive															
Style	1	4,166,367.8	0.195	.66	0.0003	1	543,790.1	0.032	.86	0.0001	1	454,343,258.8	0.060	.81	0.0001
Interactive	3	8,001,481.9	0.374	.77	0.0019	3	13,205,355.1	0.775	.51	0.0039	3	5,564,405,021.7	0.736	.53	0.0037
Style x interactive	3	11,004,271.3	0.514	.67	0.0026	3	4,937,789.1	0.290	83	0.0014	3	882,821,621.1	0.117	.95	0.0006
Error	600	21,411,044.7	N/A	N/A	N/A	600	17,039,291.1	N/A	N/A	N/A	600	7,560,858,471.6	N/A	N/A	N/A

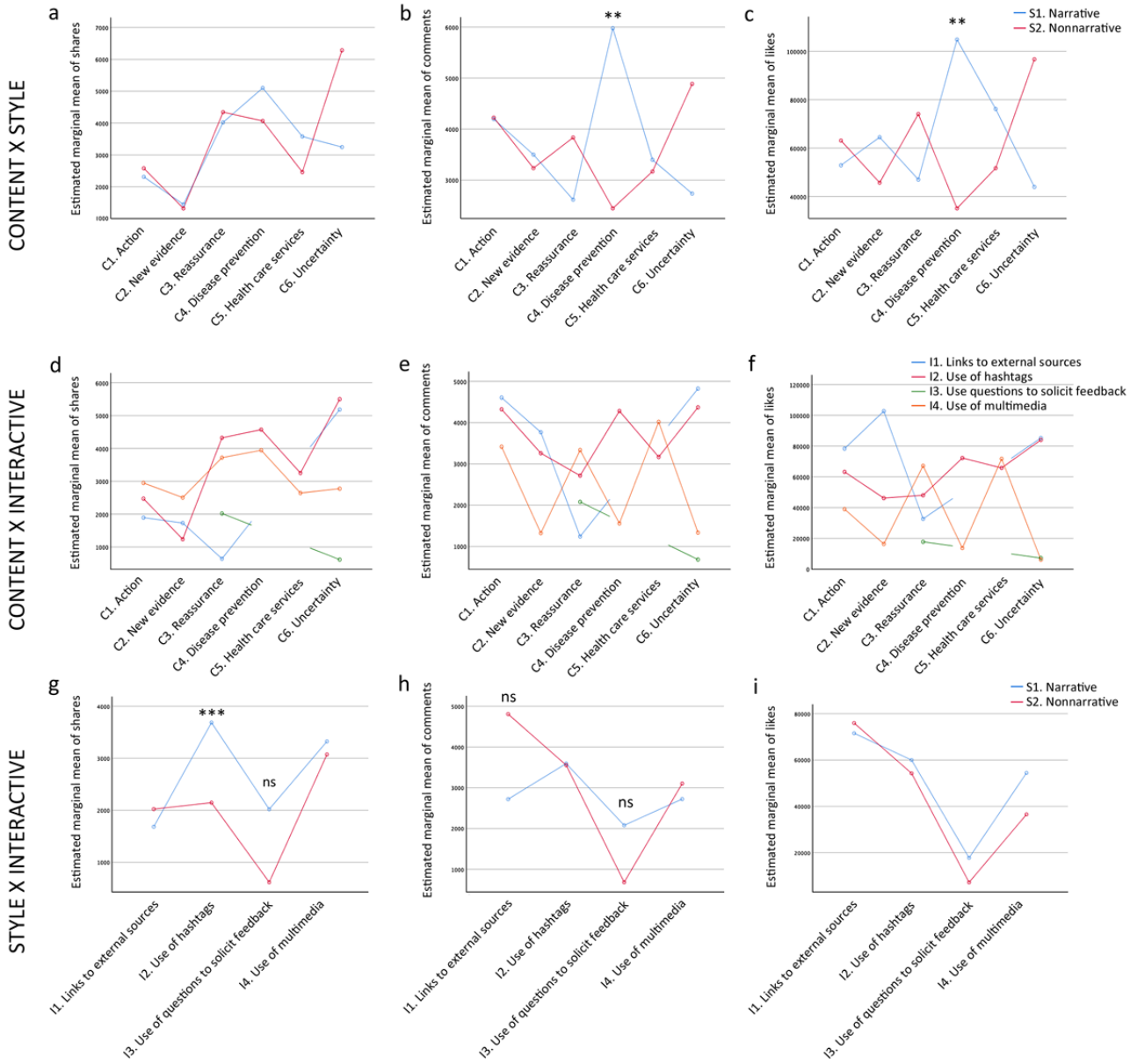
[a]MS: mean square.

[b]Italics indicate a significant interaction effect.

[c]N/A: not applicable.

To investigate the interactions between specific subdimensions, we performed simple main effect analyses to examine the interactions between specific subdimensions on the number of shares, comments, and likes. Between content and style, the different content subdimensions showed no significant differences in the number of shares between *narrative* and *nonnarrative* styles (Figure 3a). However, for the number of comments, the *narrative* style was significantly higher than that of *nonnarrative* in *disease prevention* posts (*narrative*: mean 5978.83, SEM 972.37 vs *nonnarrative:* mean 2446.33, SEM 753.19; $F_{1,597}$=8.249, *P*=<.001, η_p^2=0.014; Figure 3b). Likewise, a higher number of likes was observed for the *narrative* style as opposed to *nonnarrative* in *disease prevention* posts (*narrative*: mean 104,881.00, SEM 20,416.43 vs *nonnarrative*: mean 35,092.87 SEM 15,814.50; $F_{1,597}$=7.303, *P*=.01, η_p^2=0.012; Figure 3c). These results indicate that the pairing of *disease prevention* content with a *narrative* style generated a higher number of comments and likes.

Figure 3. Simple main effects between the sub-dimensions on the number of shares, comments, and likes. Simple main effects between the sub-dimensions of content and style on the number of shares, comments, and likes (a-c). Simple main effects between the sub-dimensions of content and interactive loop on the number of shares, comments, and likes (d-f). Simple main effects between the sub-dimensions of style and interactive loop on the number of shares, comments, and likes (g-i). ** *P*<.01, *** *P*<.001.



Between the content and the interactive dimensions, no significant differences were observed in the number of shares, comments, or likes for the *narrative* and *nonnarrative* styles (Figure 3d-f). Between the style and the interactive dimensions, the *narrative* style received significantly more shares than the *nonnarrative* one on the *use of hashtag* posts (*narrative*: mean 3685.92, SEM 359.02 vs *nonnarrative*: mean 2145.31, SEM 245.16; $F_{1,601}$=12.558, *P*<.001, η_p^2=0.021; Figure 3g), highlighting that the pairing of *narrative* style with the *use of hashtags* generated a higher number of shares. For the number of comments and likes, no significant differences were found (Figure 3h, i).

Discussion

Principal Results

Our results showed that a range of content frames, message styles, and interactive features was employed by the government to communicate about COVID-19 with the public on social media with a view to handling the health crisis. Yet different levels of engagement were revealed. In particular, *new evidence* and the *nonnarrative* style had the least impact on the number of shares (Figure 2a, d), although they were the most frequently used subdimensions (Figure 1a, b). Additionally, our NB2 results confirmed that *new evidence* and *nonnarrative* style were strong negative predictors of the number of shares (Table 2). On the other hand, the two-way ANOVA indicated that the pairing of *disease prevention* posts with a *narrative* style generated a higher number of comments and likes (Figure 3b, c), while the

NB2 results confirmed that the *narrative* style was a strong positive predictor of the number of shares (Table 2). As found in an earlier study [47], posts on preventive and safety measures related to COVID-19 were most frequently employed by public health organizations in Singapore, the United States, and England, and our results on disease prevention posts were consistent with this study. In line with previous studies, our results also revealed the strong effect of the narrative style on public engagement [23,46,55,57,58]. A narrative style of communication fosters the public's identification and emotional involvement through the character's sharing in a story event [54,58]. Through this, health narratives can possibly raise the public's awareness of health risks and encourage them to take actions to curb the spread of the disease [13,23,25,55,60].

A previous study has demonstrated an interaction effect between content and style [68], and therefore, we expected an interaction between these two dimensions. Indeed, our data showed a significant interaction between content and style on the number of likes (Table 2). With respect to the interaction between the subdimensions, our results showed that more shares were generated for posts related to *disease prevention*, *reassurance*, and *uncertainty* (Figure 2a) delivered in a *narrative* style (Figure 2d). *Links to external sources* and *use of multimedia* were also positive predictors of the number of shares (Table 2). A "share" indicates a high engagement level because it involves a cognitive action of disseminating the post to others, which can potentially reach a large audience [82-84]. Disease prevention is fundamental in a new epidemic [40-42] and uncertainty needs to be addressed because, by indicating what is unknown, more transparency of information is provided, thereby helping to build trust [14,43]. The public has a tendency to rely on social media during crises as the sites offer emotional support [85-87], indicating that the communication of uncertainty and reassurance might have served the purposes of offering emotional support and allaying anxiety. Our novel findings regarding the interaction between the subdimensions provide important insights for enhancing public engagement in epidemic communication on social media.

Implications, Limitations, and Future Work

This study contributes to the understanding of what drives the public to be engaged with COVID-19 communication by the government and adds to the body of knowledge on public engagement with epidemic communication on social media. First, our integrated, comprehensive framework of public engagement with government health communication regarding COVID-19 in China was empirically tested. *People's Daily* currently has 117 million followers, but Sina Weibo on its own is widely used for social media communication in China [34], with 462 million active online users in 2018. Existing followers of *People's Daily* can influence other Weibo users through sharing the posts, fostering a sense of community with them, and potentially helping to contain the spread of COVID-19. Both "comments" and "likes" were noted for disease prevention posts delivered in a narrative style (Figure 3b, c). A "comment" is indicative of a high engagement level because it requires the user to read a post and respond to it [88] but the interpretation of a "like" is subject to change depending on the context. For example, one study suggested that a "like" is indicative of a lower engagement level [82], although within the context of epidemic communication, a "like" might be perceived as a user's approval of the post's importance. In view of this, both "likes" and "comments" are regarded as good indicators of health risk communication.

Second, *People's Daily*'s approach of predominantly employing new evidence posts disseminated in a nonnarrative style in COVID-19 communication was not perceived as the ideal strategy to engage the public. We have gained insights into the subdimensions that can effectively enhance public engagement with epidemic communication; for instance, disease prevention posts delivered in a narrative style are viewed favorably. It is imperative for health organizations, governments, and researchers to use the public's preferred subdimensions to increase the number of shares, comments, and likes with a view to effectively disseminating new epidemic information.

One of the limitations of this study pertains to the sampling period. Because we only captured the posts from a certain period of time, the results might vary in different time periods of an evolving epidemic. Our developed framework on COVID-19 communication with the public can be further empirically tested to assess the strength of the three dimensions and applied to other cultural contexts. As social media are frequently accessed by young people while there are demographics that still use traditional mass media in different ways, COVID-19 communication can be examined in terms of impact through other channels of behavior or practice too. An investigation into the use of other popular social media platforms such as WeChat in China to disseminate COVID-19 information can be conducted to gain more insights into this topic.

Conclusions

In summary, this study presents a novel, comprehensive framework of the factors that engage the public in COVID-19 communication by the government on social media through empirically testing the measures of health content frames, style of messages, and interactive features. By drawing on this knowledge and harnessing the power of social media, governments and health organizations can determine which aspects to emphasize in an attempt to reduce the spread of the new disease.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Exemplifications of the content dimension, style dimension, interactive dialogic loop dimension, and relevant expressions identified from the corpus.
[DOCX File , 58 KB - jmir_v22i8e21360_app1.docx]

Multimedia Appendix 2
Summary of intercoder reliability statistics.
[DOCX File , 17 KB - jmir_v22i8e21360_app2.docx]

References

1. Ma J. Coronavirus: China?s first confirmed Covid-19 case traced back to November 17. South China Morning Post. 2020 Mar 13. URL: https://www.scmp.com/news/china/society/article/3074991/coronavirus-chinas-first-confirmed-covid-19-case-traced-back [accessed 2020-04-21]
2. Rocklöv J, Sjödin H, Wilder-Smith A. COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures. J Travel Med 2020 May 18;27(3):taaa030 [FREE Full text] [doi: 10.1093/jtm/taaa030] [Medline: 32109273]
3. Kommenda N, Gutiérrez P, Adolphe J. Coronavirus world map: which countries have the most cases and deaths? The Guardian. 2020 Jul 15. URL: https://www.theguardian.com/world/ng-interactive/2020/jul/15/covid-19-coronavirus-us-map-latest-cases-state-by-state [accessed 2020-04-21]
4. WHO announces COVID-19 outbreak a pandemic. World Health Organization. 2020 Mar 12. URL: http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic [accessed 2020-04-21]
5. Xiao Y, Tang S, Wu J. Media impact switching surface during an infectious disease outbreak. Sci Rep 2015 Jan 16;5:7838. [doi: 10.1038/srep07838] [Medline: 25592757]
6. Wilder-Smith A, Chiew C, Lee V. Can we contain the COVID-19 outbreak with the same measures as for SARS? Lancet Infect Dis 2020 May;20(5):e102-e107. [doi: 10.1016/s1473-3099(20)30129-8]
7. Cummings L. Emerging infectious diseases: coping with uncertainty. Argumentation 2008 Sep 24;23(2):171-188 [FREE Full text] [doi: 10.1007/s10503-008-9116-9]
8. Jones S, Waters L, Holland O, Bevins J, Iverson D. Developing pandemic communication strategies: preparation without panic. J Business Res 2010 Feb;63(2):126-132. [doi: 10.1016/j.jbusres.2009.02.009]
9. Strekalova Y. Health risk information engagement and amplification on social media. Health Educ Behav 2017 Apr;44(2):332-339. [doi: 10.1177/1090198116660310] [Medline: 27413028]
10. de Vries MS. Framing crises: response patterns to explosions in fireworks factories. Adm Soc 2016 Jul 26;36(5):594-614. [doi: 10.1177/0095399704268622]
11. Zhao Y, Cheng S, Yu X, Xu H. Chinese public's attention to the COVID-19 epidemic on social media: observational descriptive study. J Med Internet Res 2020 May 04;22(5):e18825 [FREE Full text] [doi: 10.2196/18825] [Medline: 32314976]
12. Ophir Y. Coverage of epidemics in American newspapers through the lens of the crisis and emergency risk communication framework. Health Secur 2018;16(3):147-157. [doi: 10.1089/hs.2017.0106] [Medline: 29870279]
13. Sandell T, Sebar B, Harris N. Framing risk: communication messages in the Australian and Swedish print media surrounding the 2009 H1N1 pandemic. Scand J Public Health 2013 Dec;41(8):860-865. [doi: 10.1177/1403494813498158] [Medline: 23873631]
14. Frewer L, Miles S, Brennan M, Kuznesof S, Ness M, Ritson C. Public preferences for informed choice under conditions of risk uncertainty: the need for effective risk communication. Public Underst Sci 2016 Dec 21;11(4):363-372. [doi: 10.1088/0963-6625/11/4/304]
15. Roche J, Muskavitch M. Limited precision in print media communication of West Nile virus risks. Sci Commun 2016 Aug 17;24(3):353-365. [doi: 10.1177/1075547002250300]
16. Hall L. Swine flu cases rise to 1157. The Sydney Morning Herald. 2009 Jul 04. URL: https://www.smh.com.au/national/swine-flu-cases-rise-to-1157-20090704-d86j.html [accessed 2020-04-21]
17. Househ M. Communicating Ebola through social media and electronic news media outlets: a cross-sectional study. Health Informatics J 2016 Sep;22(3):470-478. [doi: 10.1177/1460458214568037] [Medline: 25656678]
18. Gui X, Wang Y, Kou Y, Reynolds TL, Chen Y, Mei Q, et al. Understanding the patterns of health information dissemination on social media during the Zika outbreak. AMIA Annu Symp Proc 2017;2017:820-829 [FREE Full text] [Medline: 29854148]
19. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, et al. Importance of Internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. J Med Internet Res 2014 Jan 17;16(1):e20 [FREE Full text] [doi: 10.2196/jmir.2911] [Medline: 24440770]
20. Chen Y, Zhang Y, Xu Z, Wang X, Lu J, Hu W. Avian Influenza A (H7N9) and related internet search query data in China. Sci Rep 2019 Jul 18;9(1):10434. [doi: 10.1038/s41598-019-46898-y] [Medline: 31320681]

21. Xie T, Yang Z, Yang S, Wu N, Li L. Correlation between reported human infection with avian influenza A H7N9 virus and cyber user awareness: what can we learn from digital epidemiology? Int J Infect Dis 2014 May;22:1-3 [FREE Full text] [doi: 10.1016/j.ijid.2013.11.013] [Medline: 24576842]
22. Liu B, Kim S. How organizations framed the 2009 H1N1 pandemic via social and traditional media: implications for U.S. health communicators. Public Relations Rev 2011 Sep;37(3):233-244. [doi: 10.1016/j.pubrev.2011.03.005]
23. Shen F, Sheer V, Li R. Impact of narratives on persuasion in health communication: a meta-analysis. J Advertising 2015 May 05:A. [doi: 10.1080/00913367.2015.1018467]
24. Hinyard L, Kreuter M. Using narrative communication as a tool for health behavior change: a conceptual, theoretical, and empirical overview. Health Educ Behav 2007 Oct;34(5):777-792. [doi: 10.1177/1090198106291963] [Medline: 17200094]
25. Murphy S, Frank L, Chatterjee J, Baezconde-Garbanati L. Narrative versus non-narrative: the role of identification, transportation and emotion in reducing health disparities. J Commun 2013 Feb;63(1):116-137 [FREE Full text] [doi: 10.1111/jcom.12007] [Medline: 24347679]
26. Paek H, Hove T, Jung Y, Cole R. Engagement across three social media platforms: an exploratory study of a cause-related PR campaign. Public Relations Rev 2013 Dec;39(5):526-533. [doi: 10.1016/j.pubrev.2013.09.013]
27. Kent M, Taylor M. Building dialogic relationships through the world wide web. Public Relations Rev 1998 Sep;24(3):321-334. [doi: 10.1016/s0363-8111(99)80143-x]
28. Men L, Tsai W, Chen Z, Ji Y. Social presence and digital dialogic communication: engagement lessons from top social CEOs. J Public Relations Res 2018 Jul 30;30(3):83-99. [doi: 10.1080/1062726x.2018.1498341]
29. Rybalko S, Seltzer T. Dialogic communication in 140 characters or less: how Fortune 500 companies engage stakeholders using Twitter. Public Relations Rev 2010 Nov;36(4):336-341. [doi: 10.1016/j.pubrev.2010.08.004]
30. Freeman B, Chapman S. Measuring interactivity on tobacco control websites. J Health Commun 2012 Aug;17(7):857-865. [doi: 10.1080/10810730.2011.650827] [Medline: 22574732]
31. The 35th survey report. CNNIC. 2015. URL: https://cnnic.com.cn/IDR/ReportDownloads/201507/P020150720486421654597.pdf [accessed 2020-07-01]
32. Chao W. How to shoot a virus (with) a message: a study on the usage and effect of coronavirus messages on Weibo. Uppsala Universitet. 2020. URL: https://www.diva-portal.org/smash/get/diva2:1445305/FULLTEXT01.pdf [accessed 2020-07-01]
33. Han Y. Framing media discourse: a content analysis of Chinese weibo posts on Chinese territorial disputes. Indiana University, ProQuest Dissertations Publishing. 2013. URL: https://search.proquest.com/openview/bbbf62c479b889133f47073ddffb2f80/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2020-04-21]
34. Wang N, She J, Chen J. How "big vs" dominate chinese microblog: a comparison of verified and unverified users on sina weibo. In: Proceedings of the 2014 ACM conference on Web science. 2014 Jun Presented at: WebSci '14: ACM Web Science Conference; June 2014; Bloomington, IN p. 182-186. [doi: 10.1145/2615569.2615683]
35. Sina microblog user development report in Chinese. Sina Microblog Data Center. 2018. URL: https://data.weibo.com/report/reportDetail?id=433 [accessed 2020-07-01]
36. Chang T, Wang J, Chen C. News as social knowledge in China: the changing worldview of Chinese national media. J Commun 1994 Aug:52. [doi: 10.1111/j.1460-2466.1994.tb00688.x]
37. Wang Q. Who sets the news agenda on “Chinese Twitter”? The relationships between the media and opinion leaders on Weibo. Dissertation Presented to the Faculty of the Graduate School of The University of Texas at Austin. 2016. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/40314/WANG-DISSERTATION-2016.pdf?sequence=1&isAllowed=y [accessed 2020-07-01]
38. Yang F, Wendorf Muhamad J, Yang Q. Exploring environmental health on Weibo: a textual analysis of framing haze-related stories on Chinese social media. Int J Environ Res Public Health 2019 Jul 04;16(13):2374 [FREE Full text] [doi: 10.3390/ijerph16132374] [Medline: 31277378]
39. Seltzer E, Horst-Martz E, Lu M, Merchant R. Public sentiment and discourse about Zika virus on Instagram. Public Health 2017 Sep;150:170-175. [doi: 10.1016/j.puhe.2017.07.015] [Medline: 28806618]
40. Higgins J, Naylor P, Berry T, O'Connor B, McLean D. The health buck stops where? Thematic framing of health discourse to understand the context for CVD prevention. J Health Commun 2006;11(3):343-358. [doi: 10.1080/10810730600614110] [Medline: 16624798]
41. Shih T, Wijaya R, Brossard D. Media coverage of public health epidemics: linking framing and issue attention cycle toward an integrated theory of print news coverage of epidemics. Mass Commun Soc 2008 Apr 07;11(2):141-160. [doi: 10.1080/15205430701668121]
42. Umphrey L. The effects of message framing and message processing on testicular self‐examination attitudes and perceived susceptibility. Commun Res Rep 2003 Mar;20(2):97-105. [doi: 10.1080/08824090309388805]
43. WHO outbreak communication guidelines. World Health Organization. 2005. URL: https://www.who.int/csr/resources/publications/WHO_CDS_2005_28/en/ [accessed 2020-03-01]
44. Effective media communication during public health emergencies. World Health Organization. 2005. URL: https://www.who.int/csr/resources/publications/WHO_CDS_2005_31/en/ [accessed 2020-03-01]

45. Ho K, Peter Wall Workshop Participants. Harnessing the social web for health and wellness: issues for research and knowledge translation. J Med Internet Res 2014 Feb 11;16(2):e34 [FREE Full text] [doi: 10.2196/jmir.2969] [Medline: 24518432]
46. Sutter JD. Swine flu creates controversy on Twitter. CNN. 2009. URL: http://www.cnn.com/2009/TECH/04/27/swine.flu.twitter/ [accessed 2020-03-01]
47. Sesagiri Raamkumar A, Tan S, Wee H. Measuring the outreach efforts of public health authorities and the public response on Facebook during the COVID-19 pandemic in early 2020: cross-country comparison. J Med Internet Res 2020 May 19;22(5):e19334 [FREE Full text] [doi: 10.2196/19334] [Medline: 32401219]
48. Lundgren R, McMakin A. Risk Communication: A Handbook for Communicating Environmental, Safety, and Health Risks. Columbus, OH: Battelle Press; 2004:A.
49. Entman RM. Cascading activation: contesting the White House's frame after 9/11. Political Commun 2003 Oct;20(4):415-432. [doi: 10.1080/10584600390244176]
50. Tankard JW. The empirical approach to the study of media framing. In: Reese SD, Gandy Jr OH, Grant AE, editors. Framing Public Life: Perspectives on Media and Our Understanding of the Social World. London: Routledge; 2001:95-106.
51. Entman RM. Framing: toward clarification of a fractured paradigm. J Commun 1993 Aug;43(4):51 [FREE Full text]
52. Murphree V, Reber B, Blevens F. Superhero, instructor, optimist: FEMA and the frames of disaster in Hurricanes Katrina and Rita. J Public Relations Res 2009 May 29;21(3):273-294. [doi: 10.1080/10627260802640732]
53. Avery E, Kim S. Anticipating or precipitating crisis? Health agencies may not be heeding best practice advice in avian flu press releases. J Public Relations Res 2009 Apr 14;21(2):187-197. [doi: 10.1080/10627260802557449]
54. Green MC. Narratives and cancer communication. J Commun 2006 Aug 04:163-183. [doi: 10.1111/j.1460-2466.2006.00288.x]
55. Kreuter M, Green MC, Cappella JN, Slater MD, Wise ME, Storey D, et al. Narrative communication in cancer prevention and control: a framework to guide research and application. Ann Behav Med 2007 Sep;33(3):221-235. [doi: 10.1007/bf02879904]
56. Wang J, Calder B. Media engagement and advertising: transportation, matching, transference and intrusion. J Consumer Psychol 2009 Jul;19(3):546-555. [doi: 10.1016/j.jcps.2009.05.005]
57. Dal CS, Zanna M, Fong G. Narrative persuasion and overcoming resistance. In: Knowles ES, Linn JA, editors. Resistance and Persuasion. New York: Psychology Press; 2004:175-191.
58. Escalas JE. Narrative processing: building consumer connections to brands. J Consumer Psychol 2004:168-180. [doi: 10.1207/s15327663jcp1401&2_19]
59. Kim H, Bigman C, Leader A, Lerman C, Cappella J. Narrative health communication and behavior change: the influence of exemplars in the news on intention to quit smoking. J Commun 2012 Jun 01;62(3):473-492 [FREE Full text] [doi: 10.1111/j.1460-2466.2012.01644.x] [Medline: 22736808]
60. Kreuter M, Holmes K, Alcaraz K, Kalesan B, Rath S, Richert M, et al. Comparing narrative and informational videos to increase mammography in low-income African American women. Patient Educ Couns 2010 Dec;81 Suppl:S6-14 [FREE Full text] [doi: 10.1016/j.pec.2010.09.008] [Medline: 21071167]
61. Funk T. Social Media Playbook for Business. Santa Barbara, CA: Praeger; 2011.
62. Park H, Reber B, Chon M. Tweeting as health communication: health organizations' use of Twitter for health promotion and public engagement. J Health Commun 2016;21(2):188-198. [doi: 10.1080/10810730.2015.1058435] [Medline: 26716546]
63. Huo C, Zhang M, Ma F. Factors influencing people's health knowledge adoption in social media: the mediating effect of trust and the moderating effect of health threat. Library Hi Tech 2018 Mar 19;36(1):129-151. [doi: 10.1108/lht-04-2017-0074]
64. Kang M. Understanding public engagement: conceptualizing and measuring its influence on supportive behavioral intentions. J Public Relations Res 2014 Nov 11;26(5):399-416. [doi: 10.1080/1062726x.2014.956107]
65. Yuki T. What makes brands' social content shareable on Facebook? J Advertising Res 2015 Dec 01;55(4):458-470. [doi: 10.2501/jar-2015-026]
66. Gittelman S, Lange V, Gotway Crawford CA, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: Facebook likes. J Med Internet Res 2015 Apr 20;17(4):e98 [FREE Full text] [doi: 10.2196/jmir.3970] [Medline: 25895907]
67. Sun N, Rau P, Ma L. Understanding lurkers in online communities: a literature review. Comput Hum Behav 2014 Sep;38:110-117. [doi: 10.1016/j.chb.2014.05.022]
68. Ngai C, Einwiller S, Singh R. An exploratory study on content and style as driving factors facilitating dialogic communication between corporations and publics on social media in China. Public Relations Rev 2020 Mar;46(1):101813. [doi: 10.1016/j.pubrev.2019.101813]
69. Qin B, Strömberg D, Wu Y. Media bias in China. Am Econ Rev 2018 Sep 01;108(9):2442-2476. [doi: 10.1257/aer.20170947]
70. Ho J. "Sensible protesters began leaving the protests": a comparative study of opposing voices in the Hong Kong political movement. Lang Commun 2019 Jan;64:12-24. [doi: 10.1016/j.langcom.2018.09.001]
71. Introduction. People's Daily. URL: http://www.people.com.cn/GB/50142/104580/index.html [accessed 2020-04-21]
72. Xi orders resolute efforts to curb virus spread. The State Council Information Office of the People's Republic of China. URL: http://english.scio.gov.cn/topnews/2020-01/21/content_75635065.htm [accessed 2020-04-21]

73. China to curb spread of new coronavirus-related pneumonia. The State Council Information Office of the People's Republic of China. URL: http://english.scio.gov.cn/topnews/2020-01/21/content_75635465.htm [accessed 2020-04-21]
74. Aggarwal C. Outlier analysis. In: Data Mining. Cham, Switzerland: Springer, Cham; 2015:237-263.
75. Boettger RK, Palmer LA. Quantitative content analysis: its use in technical communication. IEEE Trans Profess Commun 2010 Dec;53(4):346-357. [doi: 10.1109/tpc.2010.2077450]
76. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
77. Blanca MJ, Alarcón R, Arnau J, Bono R, Bendayan R. Non-normal data: is ANOVA still a valid option? Psicothema 2017 Nov;29(4):552-557. [doi: 10.7334/psicothema2016.383] [Medline: 29048317]
78. Herbison P, Robertson MC, McKenzie JE. Do alternative methods for analysing count data produce similar estimates? Implications for meta-analyses. Syst Rev 2015 Nov 17;4:163 [FREE Full text] [doi: 10.1186/s13643-015-0144-x] [Medline: 26577545]
79. Chau A, Lo E, Wong M, Chu C. Interpreting Poisson regression models in dental caries studies. Caries Res 2018;52(4):339-345 [FREE Full text] [doi: 10.1159/000486970] [Medline: 29478049]
80. Sroka C, Nagaraja H. Odds ratios from logistic, geometric, Poisson, and negative binomial regression models. BMC Med Res Methodol 2018 Oct 20;18(1):112 [FREE Full text] [doi: 10.1186/s12874-018-0568-9] [Medline: 30342488]
81. Gurland J. Some applications of the negative binomial and other contagious distributions. Am J Public Health Nations Health 1959 Oct;49:1388-1399. [doi: 10.2105/ajph.49.10.1388] [Medline: 13855879]
82. Kim C, Yang S. Like, comment, and share on Facebook: how each behavior differs from the other. Public Relations Rev 2017 Jun;43(2):441-449. [doi: 10.1016/j.pubrev.2017.02.006]
83. Men L, Tsai W. How companies cultivate relationships with publics on social network sites: evidence from China and the United States. Public Relations Rev 2012 Dec;38(5):723-730. [doi: 10.1016/j.pubrev.2011.10.006]
84. Xu Z, Yang Q. Analyzing user retweet behavior on twitter. 2013 Feb Presented at: 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; August 2012; Istanbul, Turkey p. 46-50. [doi: 10.1109/asonam.2012.18]
85. Choi Y, Lin Y. Consumer responses to Mattel product recalls posted on online bulletin boards: exploring two types of emotion. J Public Relations Res 2009 Apr 14;21(2):198-207. [doi: 10.1080/10627260802557506]
86. Kaye B. It's a blog, blog, blog world: users and uses of weblogs. Atlantic J Comm 2005 Jun 1;13(2):73-95 [FREE Full text] [doi: 10.1207/s15456889ajc1302_2]
87. Lima A, De Domenico M, Pejovic V, Musolesi M. Disease containment strategies based on mobility and information dissemination. Sci Rep 2015 Jun 02;5:10650. [doi: 10.1038/srep10650] [Medline: 26035297]
88. Cho M, Schweickart T, Haase A. Public engagement with nonprofit organizations on Facebook. Public Relations Rev 2014 Sep;40(3):565-567. [doi: 10.1016/j.pubrev.2014.01.008]

Abbreviations

ANOVA: analysis of variance
CEO: chief executive officer
COVID-19: coronavirus disease 2019
NB2: negative binomial regression
RQ: research question
SEM: standard error of the mean
WHO: World Health Organization

Edited by G Eysenbach; submitted 12.06.20; peer-reviewed by R Petelin, Q Wang, W Tang; comments to author 15.07.20; revised version received 29.07.20; accepted 03.08.20; published 24.08.20.

Please cite as:
Ngai CSB, Singh RG, Lu W, Koon AC
Grappling With the COVID-19 Health Crisis: Content Analysis of Communication Strategies and Their Effects on Public Engagement on Social Media
J Med Internet Res 2020;22(8):e21360
URL: http://www.jmir.org/2020/8/e21360/
doi:10.2196/21360
PMID:32750013

Original Paper

Managing COVID-19 With a Clinical Decision Support Tool in a Community Health Network: Algorithm Development and Validation

Michael P McRae[1], PhD; Isaac P Dapkins[2], MD; Iman Sharif[3], MD, MPH; Judd Anderman[4], MSc; David Fenyo[5], PhD; Odai Sinokrot[6], MD; Stella K Kang[7,8], MD, MSc; Nicolaos J Christodoulides[1], PhD; Deniz Vurmaz[9], BSc; Glennon W Simmons[1], BSc; Timothy M Alcorn[10], PhD; Marco J Daoura[11], MS; Stu Gisburne[11], BA; David Zar[11], MS; John T McDevitt[1], PhD

[1]Department of Biomaterials, Bioengineering Institute, New York University College of Dentistry, New York, NY, United States

[2]Department of Population Health and Internal Medicine, Family Health Centers at NYU Langone, New York University School of Medicine, New York, NY, United States

[3]Departments of Pediatrics and Population Health, Family Health Centers at NYU Langone, New York University School of Medicine, New York, NY, United States

[4]Family Health Centers at NYU Langone, New York, NY, United States

[5]Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY, United States

[6]Department of Medicine, New York University School of Medicine, New York, NY, United States

[7]Department of Radiology, New York University School of Medicine, New York, NY, United States

[8]Department of Population Health, New York University School of Medicine, New York, NY, United States

[9]Department of Chemical and Biomolecular Engineering, NYU Tandon School of Engineering, New York University, New York, NY, United States

[10]Latham BioPharm Group, Cambridge, MA, United States

[11]OraLiva, Naples, FL, United States

Corresponding Author:
John T McDevitt, PhD
Department of Biomaterials, Bioengineering Institute
New York University College of Dentistry
433 First Avenue
Room 820
New York, NY, 10010
United States
Phone: 1 (212) 998-9204
Email: mcdevitt@nyu.edu

Abstract

Background: The coronavirus disease (COVID-19) pandemic has resulted in significant morbidity and mortality; large numbers of patients require intensive care, which is placing strain on health care systems worldwide. There is an urgent need for a COVID-19 disease severity assessment that can assist in patient triage and resource allocation for patients at risk for severe disease.

Objective: The goal of this study was to develop, validate, and scale a clinical decision support system and mobile app to assist in COVID-19 severity assessment, management, and care.

Methods: Model training data from 701 patients with COVID-19 were collected across practices within the Family Health Centers network at New York University Langone Health. A two-tiered model was developed. Tier 1 uses easily available, nonlaboratory data to help determine whether biomarker-based testing and/or hospitalization is necessary. Tier 2 predicts the probability of mortality using biomarker measurements (C-reactive protein, procalcitonin, D-dimer) and age. Both the Tier 1 and Tier 2 models were validated using two external datasets from hospitals in Wuhan, China, comprising 160 and 375 patients, respectively.

Results: All biomarkers were measured at significantly higher levels in patients who died vs those who were not hospitalized or discharged (P<.001). The Tier 1 and Tier 2 internal validations had areas under the curve (AUCs) of 0.79 (95% CI 0.74-0.84) and 0.95 (95% CI 0.92-0.98), respectively. The Tier 1 and Tier 2 external validations had AUCs of 0.79 (95% CI 0.74-0.84) and 0.97 (95% CI 0.95-0.99), respectively.

Conclusions: Our results demonstrate the validity of the clinical decision support system and mobile app, which are now ready to assist health care providers in making evidence-based decisions when managing COVID-19 patient care. The deployment of these new capabilities has potential for immediate impact in community clinics and sites, where application of these tools could lead to improvements in patient outcomes and cost containment.

(J Med Internet Res 2020;22(8):e22033) doi:10.2196/22033

KEYWORDS

COVID-19; coronavirus; clinical decision support system; point of care; mobile app; disease severity; biomarkers; artificial intelligence; app; family health center

Introduction

Coronavirus disease (COVID-19) was first reported in Wuhan, Hubei, China, in December 2019 [1], and it was declared a pandemic by the World Health Organization (WHO) [2] soon thereafter. As of June 15, 2020, about 8 million cases have been confirmed, with approximately 435,000 deaths from the disease worldwide [3]. The COVID-19 crisis has exposed critical gaps in diagnostic testing and population-level surveillance [4]. With hospitalization rates of 20% to 31% and intensive care unit (ICU) admission rates of 5% to 12% [5], surges of patients are requiring care, which has overwhelmed local health care systems and depleted reserves of medical resources.

Physicians are tasked with evaluating large amounts of rapidly changing patient data and making critical decisions in a short amount of time. Well-designed clinical decision support systems (CDSSs) deliver pertinent knowledge and individualized patient information to health care providers to enhance medical decisions [6]. These systems may rely on surveys of similar cases, while others may use a "black box" approach [7]. Traditional scores such as Sepsis-related Organ Failure Assessment (SOFA) [8-10] and Acute Physiology and Chronic Health Evaluation (APACHE) II [11,12] are commonly used in hospitals for determining disease severity and mortality, whereas clinical decision management systems, such as electronic ICU (eICU), enable systematic collection of comprehensive data [13]. However, CDSSs that use conventional variables, such as demographics, symptoms, and medical history, often do not reach their full diagnostic potential [14]. There is a compelling need for a COVID-19 disease severity assessment to help prioritize care for patients at elevated risk of mortality and manage low-risk patients in outpatient settings or at home through self-quarantine.

Several scoring systems for COVID-19 severity have been developed or adapted from existing tools, such as the Brescia-COVID Respiratory Severity Scale [15], African Federation for Emergency Medicine COVID-19 Severity Scoring Tool [16], Berlin Criteria for Acute Respiratory Distress Syndrome [17,18], and Epic Deterioration Index [19]. However, these tools have either not yet been externally validated in peer-reviewed publications or were not developed specifically for COVID-19 patient populations. Recently, we developed an integrated point-of-care COVID-19 Severity Score and CDSS that combines multiplex biomarker measurements and risk factors in a statistical learning algorithm to predict mortality with excellent diagnostic accuracy [20]. The COVID-19 Severity Score was trained and evaluated using data from 160 hospitalized COVID-19 patients from Wuhan, China. The COVID-19 Severity Score was significantly higher for patients who died than for patients who were discharged, with median scores of 59 (IQR 40-83) and 9 (IQR 6-17), respectively, and an area under the curve (AUC) of 0.94 (95% CI 0.89-0.99).

COVID-19 has caused and continues to cause significant morbidity and mortality globally. A validated tool to assess and quantify viral sepsis severity and patient mortality risk would address the urgent need for disease severity categorization. Toward the goal of improving prognostic judgement and outcomes, we assembled a multidisciplinary team representing stakeholders from technology, machine learning, engineering, primary care, and in vitro diagnostic testing to develop a COVID-19 disease severity test. The unfolding novel COVID-19 pandemic has greatly illuminated the important role of community health centers in providing safe and effective patient care. The Family Health Centers (FHC) at New York University (NYU) Langone is a large Federally Qualified Health Center; it provides comprehensive primary and preventive health care to a diverse population of patients across the New York City metropolitan area and is well-positioned to improve survival by fast-tracking hospitalization of patients at high risk of severe disease. This study describes a clinical decision support tool for COVID-19 disease severity developed using recent data from the FHC and externally validated using data from two recent studies from hospitals in Wuhan, China. We describe a practical and efficient tiered approach that involves a model with nonlaboratory inputs (Tier 1), a model with biomarkers commonly measured in ambulatory settings (Tier 2), and a mobile app to deliver and scale these tools. The deployment of these new capabilities has potential for immediate clinical impact in community clinics, where these tools could lead to improvements in patient outcomes and prognostic judgment.

Methods

Patient Data

Data from 701 patients with COVID-19 were collected across 9 clinics and hospitals within the FHC network at NYU Langone, one of the largest Federally Qualified Health Center networks in the United States. All patients had detectable severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection as evidenced by polymerase chain reaction testing. The following outcomes were recorded: not hospitalized, discharged, ventilated, and deceased. The data that support the Tier 1 Outpatient Model and Tier 2 Biomarker Model development are available from the authors upon reasonable request and with permission of FHC at NYU Langone.

Validation data for the Tier 1 Outpatient Model were derived from a study of 160 hospitalized patients with COVID-19 from Zhongnan Hospital of Wuhan University. The data that support validation of the Tier 1 Outpatient Model are available from the authors upon reasonable request and with permission of Zhongnan Hospital of Wuhan University. Validation data for the Tier 2 Biomarker Model were derived from a study of 375 hospitalized patients with COVID-19 from Tongji Hospital in Wuhan, China. The data that support the validation of the Tier 2 Biomarker Model are available as Supplementary Data in a publication by Yan et al [21].

Clinical Decision Support Tool

This study describes the development of a two-tiered CDSS for the assessment of COVID-19 disease severity using similar methods to those described previously [20,22]. The Tier 1 Outpatient Model uses nonlaboratory data that are readily available prior to laboratory measurements and is intended to help determine whether Tier 2 biomarker-based testing and/or hospitalization are necessary. Here, a lasso logistic regression model was trained to distinguish between patients who were not hospitalized or who were hospitalized and discharged home without need for ventilation vs patients who were ventilated or died. Patients who were still hospitalized when the data were compiled were excluded. The following predictors were considered in model training: age, gender, BMI, systolic blood pressure, temperature, symptoms (cough, fever, or shortness of breath), known cardiovascular comorbidities (patient problem list includes one or more of cerebrovascular disease, heart failure, ischemic heart disease, myocardial infarction, peripheral vascular disease, and hypertension), pulmonary comorbidities (asthma and chronic obstructive pulmonary disease), and diabetes.

The Tier 2 Biomarker Model predicts disease severity using biomarker measurements and patient characteristics. A lasso logistic regression model was trained to distinguish patients who died versus patients who were either never hospitalized or discharged home. Patients who were ventilated or still hospitalized when the data were compiled were excluded. The following predictors were considered in model training: age, gender, comorbidities, C-reactive protein (CRP), cardiac troponin I (cTnI), D-dimer, procalcitonin (PCT), and N-terminal fragment of the prohormone brain natriuretic peptide (NT-proBNP). Predictors that were not relevant to the model (ie, coefficients equal to zero) were removed. Laboratory measurements across all time points were log-transformed. Patients with no measurements for the aforementioned biomarkers were excluded. Biomarker values below the limits of detection were set to the minimum measured value divided by the square root of 2.

Model Development and Statistical Analysis

Both Tier 1 and Tier 2 models were developed using the same procedure. All continuous predictors were standardized with a mean of 0 and a variance of 1. Missing data were imputed using the multivariate imputation by the chained equations algorithm in the statistical software R (R Project) [23]. Predictive mean matching and logistic regression imputation models were used to generate 10 imputations for continuous and categorical predictors, respectively. Samples in the training and test sets were partitioned using stratified 5-fold cross-validation to preserve the relative proportions of outcomes in each fold. Model training and selection were performed on each of the 10 imputation datasets for 10 Monte Carlo repetitions and optimized for the penalty parameter corresponding to one standard error above the minimum deviance for additional shrinkage. After the initial training, only predictors with nonzero regression coefficients were retained, and the model was retrained with a reduced number of predictors. The training process was repeated until all predictors yielded nonzero coefficients. Model performance was documented in terms of the mean (95% CI) of the AUC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Median (IQR) cross-validated COVID-19 scores were compared across disease outcomes. The COVID-19 scores for both models and biomarker measurements were compared using the Wilcoxon rank sum test. Normally distributed predictors were compared using an independent *t* test. Proportions were compared using the chi-squared test [24,25]. Two-sided tests were considered statistically significant for $P<.05$.

External Validation

We externally validated the Tier 1 Outpatient Model using data from a study of 160 hospitalized patients with COVID-19 from Zhongnan Hospital of Wuhan University. Only patients with complete information (age, systolic blood pressure, gender, diabetes, and cardiovascular comorbidities) were included. The model performance was documented in terms of AUC, sensitivity, specificity, PPV, and NPV. Results were presented in a scatter/box plot of COVID-19 outpatient scores for patients who were discharged and those who died.

Similarly, we externally validated the Tier 2 Biomarker Model using data from a study of 375 hospitalized patients with COVID-19 from Tongji Hospital in Wuhan, China, collected between January 10 and February 18, 2020 [21]. While most patients had multiple lab measurements over time, the first available lab value for each biomarker was used to validate the model to maximize lead time. Patients with one or more missing predictor values were excluded. Model performance was documented in terms of AUC, sensitivity, specificity, PPV, and NPV. Results were presented in a scatter/box plot of COVID-19 Biomarker Scores for patients who were discharged and who died.

To demonstrate how the COVID-19 Biomarker Score could be used to track changes in disease severity over time, the model was evaluated based on time series biomarker data. Because the lab measurements were reported asynchronously, the model was reevaluated every time a new biomarker measurement became available. Time series plots of the COVID-19 Biomarker Score were generated for each patient.

Results

This study describes the development of a 2-tiered CDSS to assess COVID-19 disease severity using similar methods to those described previously [20,22]. The Tier 1 Outpatient Model uses nonlaboratory data that are readily available prior to

laboratory measurements and is intended to help determine whether Tier 2 biomarker-based testing and/or hospitalization are warranted. The Tier 2 Biomarker Model predicts disease severity using biomarker measurements and patient characteristics.

The CDSS and mobile app are designed to support decisions made in multiple settings, including home care, primary care or urgent care clinics, emergency departments, and hospital and intensive care (Figure 1). The process starts with symptomatic patients who are positive or presumably positive for COVID-19 and seeking care at a family health center or emergency room. In the family health center, decisions are made in two key stages, or tiers. The Tier 1 algorithm is intended for individuals in an outpatient setting where laboratory data are not yet readily available, and it uses only age, gender, blood pressure, and comorbidities. Patients with a low COVID-19 Outpatient Score may be managed in a home or telemedicine setting, while patients with a high COVID-19 Outpatient Score are referred for a blood draw and Tier 2 biomarker-based test. The Tier 2 algorithm, which is directly related to mortality risk, predicts disease severity using biomarker measurements and age. Patients with a low COVID-19 Biomarker Score are expected to be managed in a low-to-moderate risk group (eg, 5-day telehealth follow-up), while patients with a high COVID-19 Biomarker Score are expected to be hospitalized in most cases or managed in a high risk group (eg, 24- to 48-hour follow-up). Providers encountering clinically evident severe cases, as in urgent care or emergency departments, may choose to bypass the Tier 1 Outpatient Score and perform biomarker testing and Tier 2 triage on all patients with COVID-19. Last, in the hospital setting, patients are serially monitored for their COVID-19 Biomarker Scores. This personalized time series information directly related to mortality risk has strong potential to optimize therapy, improve patient care, and ultimately save lives. For both algorithms, we selected cutoffs that balanced sensitivity and specificity; however, these algorithms can be easily tuned for high sensitivity or high specificity by adjusting the weighting or relative importance of sensitivity and specificity in clinical practice.

Of the 701 patients with detectable COVID-19 infection cared for in one of the 9 clinics within the FHC network, 402 (57.3%) were not hospitalized, 185 (26.4%) were hospitalized and discharged, 19 (2.7%) were ventilated, and 95 (13.6%) died (Table 1). Ventilated and deceased patients were older than those who were not hospitalized or discharged (P=.03 and P<.001, respectively). Of patients who were ventilated and deceased, 14/19 (73.7%) and 60/95 (63.2%) were male, respectively, vs 271/587 (46.1%) for patients with less severe disease (ie, not hospitalized or discharged) (P=.02 and P=.002, respectively). Diabetes was also a statistically significant factor, with 9/19 (47.4%) and 52/95 (54.7%) in the ventilated and deceased groups vs 149/587 (25.3%) in the nonhospitalized and discharged groups (P=.03 and P<.001, respectively). Likewise, 10/19 (52.6%) of ventilated patients (P=.04) and 65/95 (68.4%) of deceased patients (P<.001) had one or more cardiovascular comorbidities, vs 181/587 (30.8%) for the less severe disease categories, with hypertension being the most common comorbidity. Interestingly, systolic blood pressure was significantly higher for patients who were not hospitalized vs those who were discharged (P=.004), and patients who died had abnormally low blood pressure relative to patients with less severe disease (P<.001). All biomarkers (cTnI, CRP, PCT, D-dimer, and NT-proBNP) were measured at significantly higher levels in patients who died vs those who were not hospitalized or discharged (P<.001).

Figure 1. Clinical decision support system and mobile app for managing COVID-19 care. COVID-19: coronavirus disease; CRP: C-reactive protein; PCT: procalcitonin.

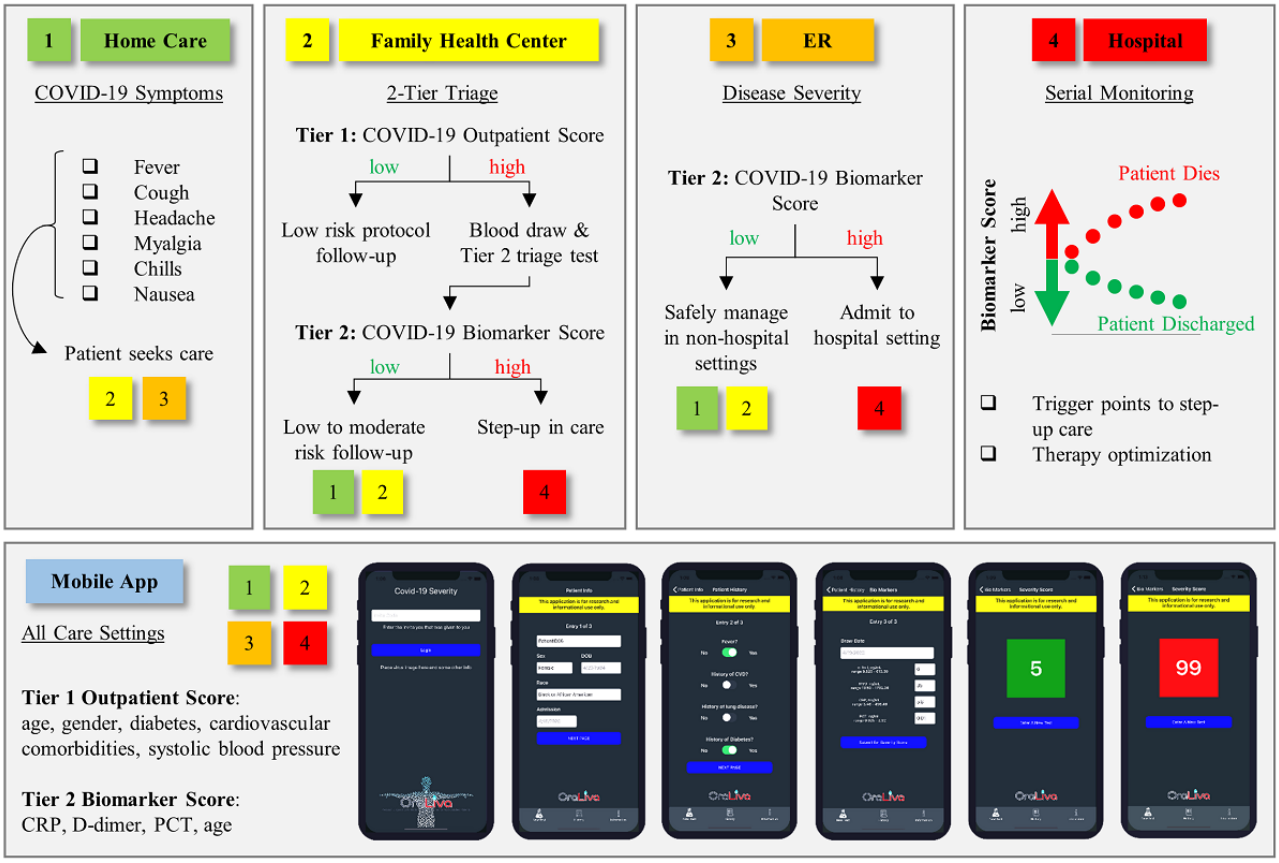


Table 1. Characteristics of the patients included in model training. Data are represented as n (%), mean ± standard deviation, or median (IQR).

Characteristic	Not hospitalized (n=402)	Discharged (n=185)		Ventilated (n=19)		Deceased (n=95)	
	Value	Value	*P* value[a]	Value	*P* value[b]	Value	*P* value[b]
Age (years), mean (SD)	48 (17)	50 (17)	.32	58 (20)	.03	67 (14)	<.001
Male sex, n (%)	182 (45.3)	89 (48.1)	.52	14 (73.7)	.02	60 (63.2)	.002
BMI, kg/m^2, mean (SD)	25 (4)	28 (6)	.16	29 (5)	.46	25 (6)	.06
Systolic BP[c] (mm Hg), mean (SD)	132 (14)	123 (19)	.004	126 (20)	.78	94 (40)	<.001
Diastolic BP (mm Hg), mean (SD)	82 (8)	71 (11)	<.001	70 (12)	.29	54 (26)	<.001
Temperature (ºF), mean (SD)	99 (1)	98 (5)	.54	99 (1)	.66	100 (2)	.12
Pulse (beats per minute), mean (SD)	90 (18)	84 (14)	.06	93 (14)	.03	74 (54)	.02
Asthma, n (%)	44 (10.9)	12 (6.5)	.09	3 (15.8)	.37	6 (6.3)	.31
COPD[d], n (%)	60 (14.9)	17 (9.2)	.06	3 (15.8)	.74	15 (15.8)	.48
Cancer, n (%)	13 (3.2)	5 (2.7)	.73	2 (10.5)	.07	14 (14.7)	<.001
Cardiovascular comorbidities[e], n (%)	120 (29.9)	61 (33.0)	.45	10 (52.6)	.04	65 (68.4)	<.001
Diabetes, n (%)	96 (23.9)	53 (28.6)	.22	9 (47.4)	.03	52 (54.7)	<.001
HIV/AIDS, n (%)	3 (0.7)	2 (1.1)	.68	0 (0.0)	.69	3 (3.2)	.053
Liver disease, n (%)	11 (2.7)	10 (5.4)	.11	2 (10.5)	.12	4 (4.2)	.76
Renal disease, n (%)	20 (4.9)	17 (9.2)	.051	3 (15.8)	.10	21 (22.1)	<.001
cTnI[f] (pg/mL), median (IQR)	7.07 (7.07-7.07)	7.07 (7.07-7.07)	.30	20.00 (7.07-63.75)	<.001	73.50 (7.07-712.00)	<.001
CRP[g] (mg/L), median (IQR)	51.40 (16.55-101.35)	67.90 (17.95-121.50)	.28	37.30 (27.30-139.72)	.44	176.00 (115.00-287.00)	<.001
PCT[h] (ng/mL), median (IQR)	0.12 (0.06-0.36)	0.10 (0.05-0.31)	.31	0.69 (0.07-1.91)	.008	1.61 (0.35-8.31)	<.001
D-Dimer (µg/mL[i]), median (IQR)	0.39 (0.20-0.71)	0.27 (0.18-0.56)	.047	0.86 (0.50-3.02)	<.001	1.58 (0.72-5.35)	<.001
NT-proBNP[j] (pg/mL), median (IQR)	93.00 (36.50-375.25)	88.00 (28.50-298.00)	.60	217.00 (78.00-394.25)	.13	937.00 (160.25-5728.50)	<.001

[a]Compared to patients who were not hospitalized.

[b]Compared to patients who were not hospitalized or discharged.

[c]BP: blood pressure.

[d]COPD: chronic obstructive pulmonary disease.

[e]Cardiovascular comorbidities: one or more of cerebrovascular disease, heart failure, ischemic heart disease, myocardial infarction, peripheral vascular disease, and hypertension.

[f]cTnI: cardiac troponin I.

[g]CRP: C-reactive protein.

[h]PCT: procalcitonin.

[i]µg/mL: micrograms per milliliter.

[j]NT-proBNP: N-terminal fragment of the prohormone brain natriuretic peptide.

Tier 1 Outpatient Model

The Tier 1 Outpatient Model for COVID-19 disease severity was developed and internally validated using data from the FHCs at NYU Langone (Figure 2). The model retained the following predictors: age, gender, systolic blood pressure, cardiovascular comorbidities (one or more of cerebrovascular disease, heart failure, ischemic heart disease, myocardial

infarction, peripheral vascular disease, and hypertension), and diabetes. The median COVID-19 Outpatient Scores were 11, 13, 20, and 27 for not hospitalized, discharged, ventilated, and deceased patients, respectively. The AUC of the model was 0.79 (95% CI 0.74-0.84) at the optimal cutoff COVID-19 Outpatient Score of 18 (Table 2). The median scores (Figure 2) had statistically significant differences for comparisons between all patient groups, except for not hospitalized vs discharged (*P*=.18).

Figure 2. Validation of the Tier 1 Outpatient Model. A. Lasso logistic regression coefficients revealing the relative importance of predictors in generating the score. B. Box/scatter plot from the internal validation showing the Tier 1 Outpatient Scores for the four outcomes. A cutoff score of 18 (red dotted line) balances sensitivity and specificity for "Noncase" vs "Case" patients (gray line). COVID-19: coronavirus disease; CV comorbidities: cardiovascular comorbid conditions; No Hosp.: patients who were not hospitalized; Vent.: patients who were ventilated.

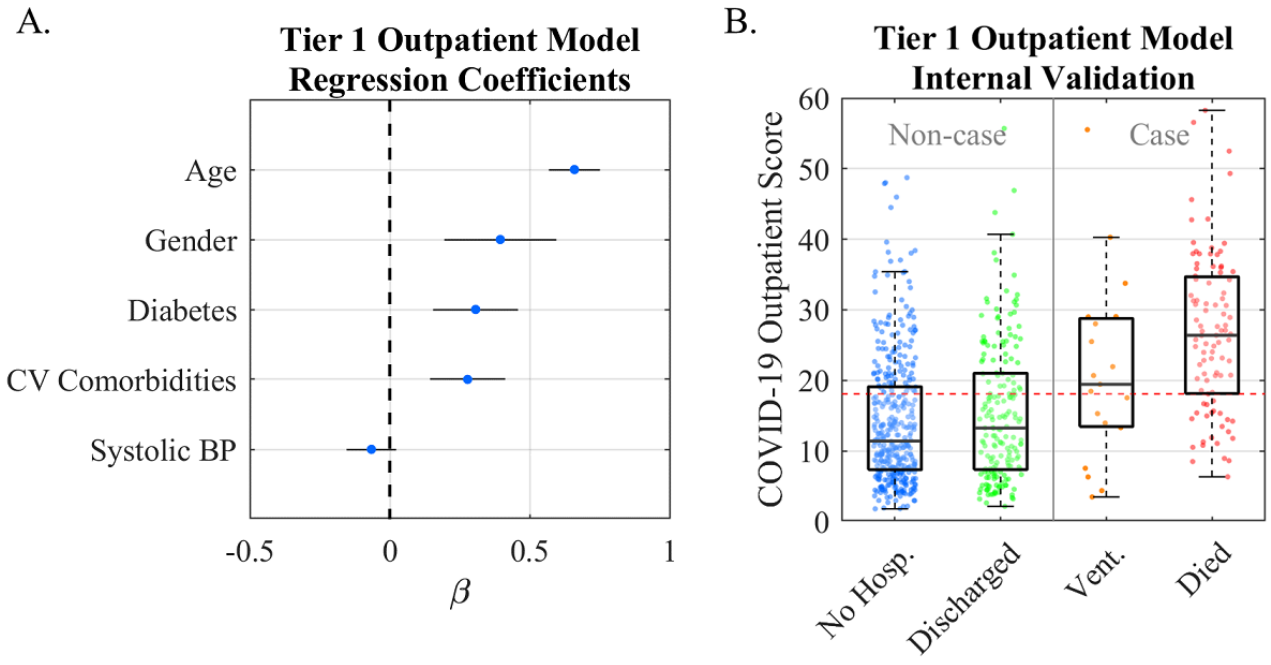


Table 2. Internal validation performance in terms of AUC, sensitivity, specificity, PPV, and NPV (95% CI) from 5-fold cross-validation. The Tier 1 and 2 models were trained and tested using data from Family Health Centers at New York University.

	Tier 1 Outpatient Model	Tier 2 Biomarker Model
AUC[a]	0.79 (0.74-0.84)	0.95 (0.92-0.98)
Sensitivity	0.73 (0.69-0.76)	0.89 (0.86-0.92)
Specificity	0.73 (0.69-0.76)	0.89 (0.86-0.92)
PPV[b]	0.34 (0.30-0.38)	0.70 (0.65-0.74)
NPV[c]	0.93 (0.91-0.95)	0.97 (0.94-0.98)

[a]AUC: area under the curve.

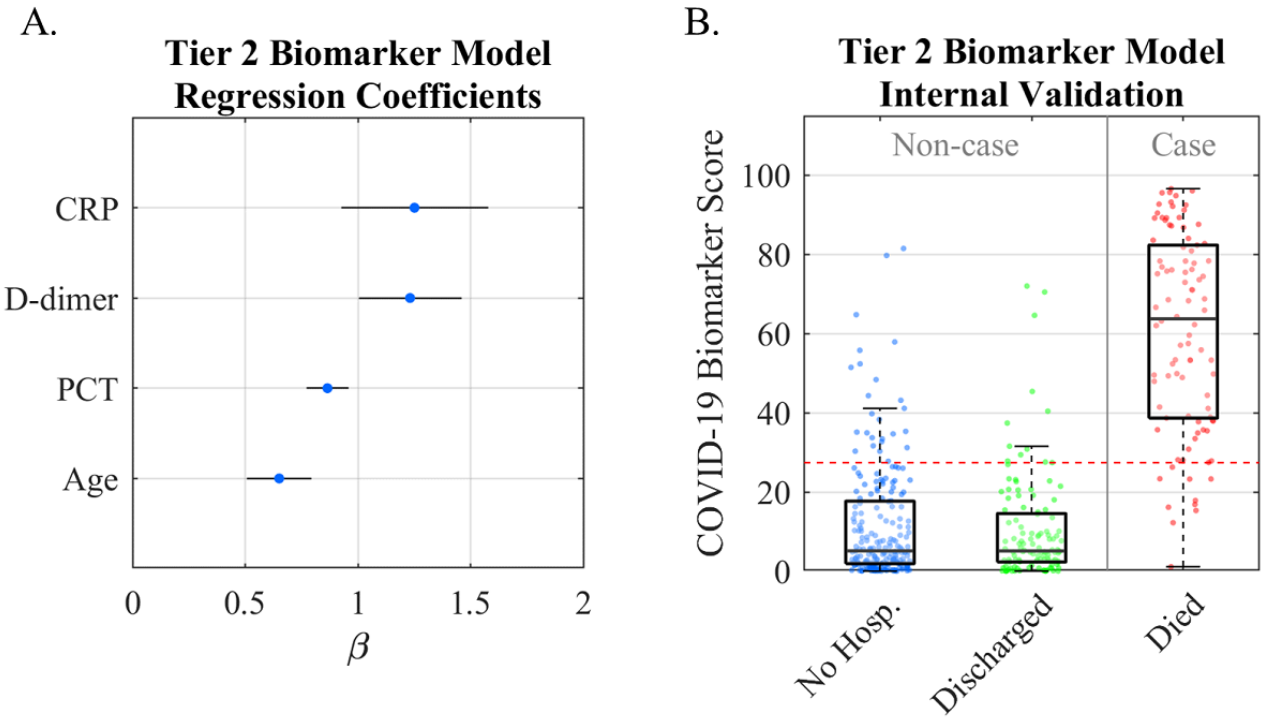
[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

Tier 2 Biomarker Model

The Tier 2 Biomarker Model for COVID-19 disease severity was developed and internally validated using data from the FHCs at NYU Langone (Figure 3). Patients who were ventilated (n=19) and still hospitalized (n=19) were excluded. Patients with fewer than one biomarker measurement were excluded (n=190 not hospitalized, n=64 discharged, n=1 deceased). The remaining 427 patients with one or more biomarker measurements were included in the analysis (n=212 not hospitalized, n=121 discharged, n=94 deceased). The model retained the following predictors after shrinkage and selection: age, D-dimer, PCT, and CRP. The median COVID-19 Outpatient Scores were 5, 5, and 64 for not hospitalized, discharged, and deceased patients, respectively. The AUC of the model was 0.95 (95% CI 0.92-0.98) at the optimal cutoff COVID-19 Outpatient Score of 27 (Table 2). The median COVID-19 Outpatient Scores (Figure 3) had statistically significant differences for comparisons between patients who were not hospitalized and patients who died (*P*<.001) and between patients who were discharged and patients who died (*P*<.001).

Figure 3. Validation of the Tier 2 Biomarker Model. A. Lasso logistic regression coefficients revealing the relative importance of predictors in generating the score. B. The box/scatter plot from internal validation shows Tier 2 Biomarker Scores for the three patient outcomes. A cutoff score of 27 (horizontal red dotted line) balances sensitivity and specificity for "Noncase" vs "Case" patients (vertical gray line) COVID-19: coronavirus disease; No Hosp.: patients who were not hospitalized.



External Validation

We externally validated the Tier 1 Outpatient Model using data from a study of 160 hospitalized patients with COVID-19 who had hypertension from Zhongnan Hospital of Wuhan University, Wuhan, China [26]. Of the 160 patients in the study, 4 (2.5%) were missing one or more predictors and were excluded from the analysis. The COVID-19 Biomarker Scores were evaluated for 115 patients who were discharged and 41 patients who died (Figure 4A). The median COVID-19 Biomarker Scores were 27.9 (IQR 22.0-36.4) for patients who were discharged and 39.7 (34.2-47.4) for patients who died. The external validation diagnostic performance was determined using a cutoff score of 34 (Table 3).

We externally validated the Tier 2 Biomarker Model using data from a study of 375 hospitalized COVID-19 patients from Tongji Hospital in Wuhan, China, collected between January 10 and February 18, 2020 [21]. To maximize potential lead time, the first available laboratory measurements during hospitalization were used to generate cross-sectional COVID-19 Biomarker Scores, representing the first in a series of measurements collected for hospital stays lasting a median of 12.5 (IQR 8-17.5) days prior to the outcomes (discharged or deceased). Out of the 375 patients in the study, 133 were missing one or more lab values and excluded from the analysis. The COVID-19 Biomarker Scores were evaluated for 112 patients who were discharged and 130 patients who died (Figure 4B). The median COVID-19 Biomarker Scores were 1.6 (IQR 0.5-6.2) for patients who were discharged and 59.1 (IQR 36.6-78.9) for patients who died. The external validation diagnostic performance was determined using a cutoff score of 19 (Table 3).

Figure 4. External validation results. A. The Tier 1 Outpatient Model was evaluated using data from patients with COVID-19 at Zhongnan Hospital of Wuhan University [26]. B. The Tier 2 Biomarker Model was evaluated using data from patients with COVID-19 at Tongji Hospital [21]. COVID-19: coronavirus disease.

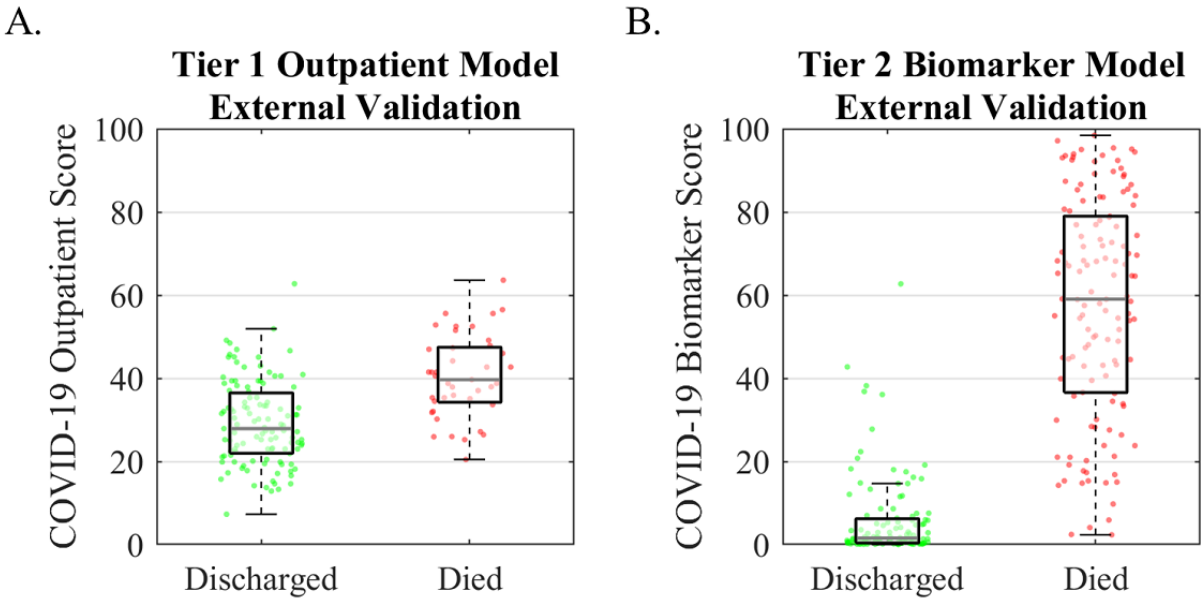


Table 3. External validation performance in terms of AUC, sensitivity, specificity, PPV, and NPV (95% CI). The Tier 1 Outpatient Model was evaluated on the Zhongnan Hospital dataset [26]. The Tier 2 model was evaluated on the Tongji Hospital dataset [21].

	Tier 1 Outpatient Model	Tier 2 Biomarker Model
AUC[a]	0.79 (0.70-0.88)	0.97 (0.95-0.99)
Sensitivity	0.76 (0.68-0.82)	0.89 (0.84-0.93)
Specificity	0.73 (0.65-0.80)	0.93 (0.89-0.96)
PPV[b]	0.50 (0.42-0.58)	0.94 (0.90-0.96)
NPV[c]	0.89 (0.83-0.94)	0.88 (0.83-0.92)

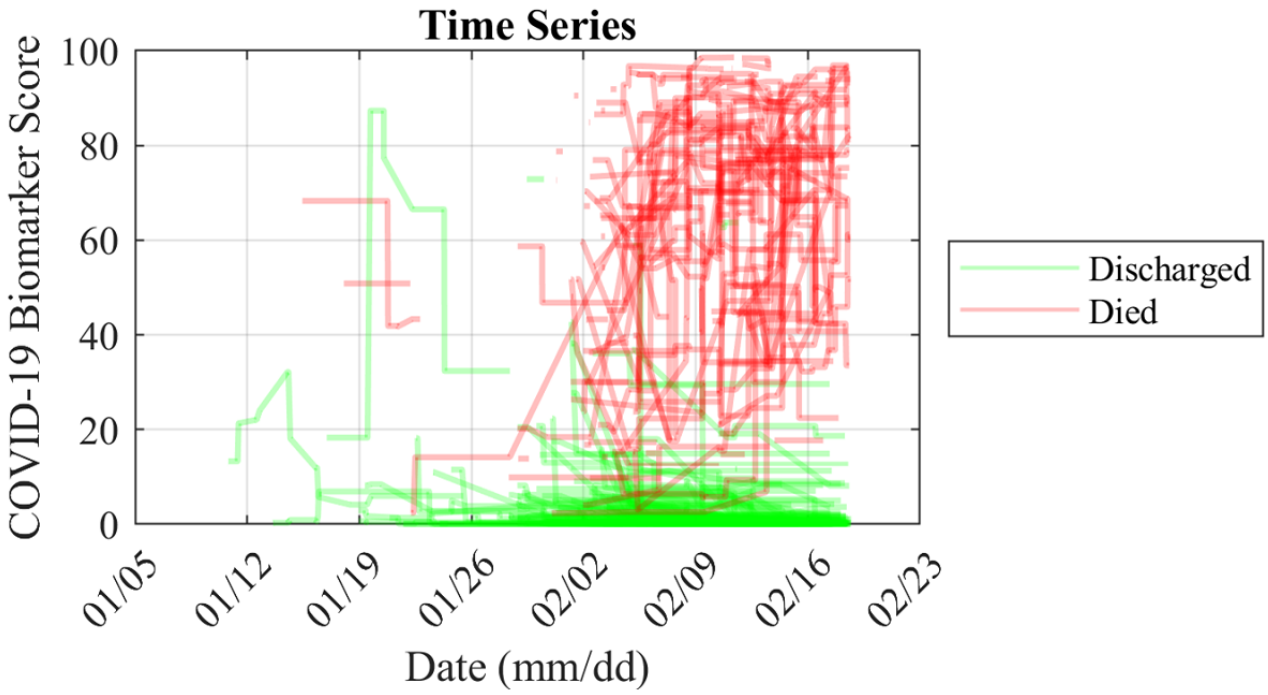
[a]AUC: area under the curve.

[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

We also evaluated the COVID-19 Biomarker Scores for patients over time using longitudinal biomarker measurement data from individual patients in the external validation set (Figure 5). These data represent individual patients' scores over a median of 12.5 days (IQR 8-17.5) between admission and outcomes of discharge or death. The first scores available after admission were significantly higher in patients who died vs patients who were discharged (AUC 0.97, cutoff score of 19); over time, patients who were discharged had an average decrease in score (–4.7), while patients who died had an average increase in score (+11.2).

Figure 5. Spaghetti plot of longitudinal COVID-19 Biomarker Scores for patients in the external validation set from Tongji Hospital [21] between January 10 and February 18, 2020. These data represent individual patients' scores over a median (IQR) of 12.5 (8–17.5) days between admission and outcomes of discharged or deceased. The first scores available after admission were significantly higher in those that died vs those that were discharged (AUC 0.97, cutoff score of 19), and over time patients who were discharged had an average decrease in score (-4.7) while those that died had an average increase in score (+11.2).



Discussion

As the COVID-19 pandemic continues to create surges and resurgences without an effective vaccine, the goal of this multidisciplinary team was to develop a triage and prognostication tool that strengthens community-level testing and disease severity monitoring. A CDSS and mobile app for COVID-19 severity have been designed, developed, and validated using data from 1236 patients with COVID-19 across numerous clinics and hospitals in the coronavirus disease epicenters of Wuhan, China, and New York, United States. These clinically validated tools have potential to assist health care providers in making evidence-based decisions in managing the care of patients with COVID-19. The significance of this work is realized by the algorithms developed and validated here, which are accurate, interpretable, and generalizable.

Accurately identifying patients with elevated risk for developing severe COVID-19 complications can empower health care providers to save lives by prioritizing critical care, medical resources, and therapies. With respect to accuracy, both Tier 1 and Tier 2 models were effective in discriminating disease outcomes, with statistically significant differences between the most relevant patient groups (AUCs of 0.79 and 0.97 for Tier 1 and Tier 2 external validation, respectively). As expected, the diagnostic accuracy of the Tier 1 Outpatient Model in terms of AUC was lower than that of the Tier 2 Biomarker Model, which demonstrates the importance of biomarker data in determining disease severity. The accuracy with which the Tier 2 Biomarker Score identified patients who eventually died reflects the unfortunate and morbid reality of the COVID-19 pandemic to date. However, as medical knowledge and experience with COVID-19 progresses, it is possible that future treatments and interventions could improve patient survival. In this context, the Tier 2 Biomarker Score could be used to monitor patients' treatment progression or regression over time and modify therapies accordingly.

Another strength of this approach is the interpretability of the models. While many predictive tools rely on "black box" methods in which algorithmic decisions and the logic supporting those decisions are uninterpretable, the lasso logistic regression method is transparent through its coefficients (ie, log odds) and probabilistic output. The Tier 1 Outpatient Score is the probability of severe disease (ventilation or death) based on the predictors (age, gender, diabetes, cardiovascular comorbidities, and systolic blood pressure). Likewise, the Tier 2 Biomarker Score is the probability of mortality based on CRP, D-dimer, PCT, and age. Predictive models such as these are more likely to be adopted for clinical applications in which transparency and interpretability are valued.

One of the most clinically relevant features of this new CDSS is the capacity to monitor individual patients over time. The use of this *precision diagnostic* approach allows for the amplification of early signs of disease, which can be achieved by focusing on time-course changes of biomarker signatures that are referenced not to population metrics, but rather back to the individual patient. As an example, the use of time course changes in individual biomarker fingerprints has been explored previously in the study of early detection in ovarian cancer [27]. Studies demonstrated that cancer antigen 125 by itself for a single time point was a poor diagnostic marker due to overlapping reference range problems across the population. However, when each patient was treated as their own point of

reference and biomarker slopes for individual patients were considered, the diagnostic accuracy for this same biomarker increased significantly. Similarly, the COVID-19 Biomarker Score time series (Figure 5) reveals a strong capacity to separate patients who die of COVID-19 complications from those who are discharged from the hospital. Note that the app includes capabilities to use proximal biomarker measurements, allowing for biomarker measurements to be collected over time without the rigid restriction of requiring completion of all biomarker measurements at the same time for all time points. This flexibility is anticipated to afford more convenience for longitudinal monitoring of patients.

Lastly, the models developed here demonstrated generalizability through external model validation. External validation is essential before implementing prediction models in clinical practice [28]. We found that the AUCs for both the Tier 1 and Tier 2 models were similar for internal vs external validation, demonstrating that the models are generalizable to making predictions for these disease indications in different care settings and for different patient demographics. Usually, prediction models perform better on the training data than on new data; however, in this study, we found that the external validation results were approximately the same or better (Tier 1: AUC of 0.79 vs 0.79; Tier 2: 0.95 and 0.97 for internal and external validation, respectively), suggesting that patients in the external validation sets may have suffered from more severe disease.

Despite the potential for CDSSs to transform health care, major challenges remain for translating and scaling these tools. Future data and, thus, future model performance may have large heterogeneity, which may be exacerbated by missing data (potentially not missing at random), nonstandard definitions of outcomes, and incomplete laboratory measurements and follow-up times [29]. The mobile app developed here is intended to reduce heterogeneity by encouraging the harmonization of data collection across multiple care settings. Further, models may be tuned through optimization of cutoffs for certain patient subpopulations. Another challenge in deploying a CDSS that relies on biomarker measurements is accounting for differences in laboratory testing across hospitals and clinics. The variability of these measurements across institutions may have a large impact on the distribution of COVID-19 Biomarker Scores. This challenge creates a unique opportunity for standardized, well-calibrated, and highly scalable point-of-care tests for COVID-19 disease severity [20,30,31]. Finally, the COVID-19 pandemic is a fluid and rapidly evolving crisis. Not only will our epidemiological and physiological understanding of the disease evolve over time, but viral mutations could also alter disease severity in future outbreaks. The two-tiered algorithms developed here are highly amenable to future adaptations in which new data are included in the training through periodic or continuous learning.

A commercial app has been developed in collaboration with OraLiva, Inc for deployment of these tools to frontline health care workers managing COVID-19 patients. Plans are now in place to assess the usability, user satisfaction, and confidence in results of this CDSS and mobile app in the FHCs at NYU. Future efforts will focus on point-of-care testing capabilities to more rapidly assess the Tier 2 biomarkers described in this study using a previously developed and published platform [20,30,31]. The deployment of these new capabilities has potential for immediate clinical impact in community clinics, where the application of these tools could significantly improve the quality of care.

Acknowledgments

A portion of this work was funded by Renaissance Health Service Corporation and Delta Dental of Michigan. The authors thank Zhibing Lu and colleagues of Zhongnan Hospital of Wuhan University for providing the data used for external validation in this study. The authors also thank Ye Yuan, Hui Xu, Shusheng Li, and colleagues of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology for providing open access to the data used for external validation in this study.

Conflicts of Interest

MPM has served as a paid consultant for SensoDx, LLC, and has a provisional patent pending. NJC and IPD have a provisional patent pending. SKK has received royalties from Wolters Kluwer for work performed outside of this study. MJD has served as a paid consultant for Hubwerx, LLC. MJD, DZ, and SG have served as paid consultants for OraLiva, Inc. JTM has a provisional patent pending. In addition, he has an ownership position and an equity interest in both SensoDx II, LLC, and OraLiva, Inc, and serves on their advisory boards.

References

1. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
2. World Health Organization. URL: https://www.who.int/ [accessed 2020-08-13]
3. Coronavirus disease (COVID-19) Situation Report – 133. World Health Organization. 2020 Jun 01. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200601-covid-19-sitrep-133.pdf?sfvrsn=9a56f2ac_4 [accessed 2020-08-13]
4. Sharfstein JM, Becker SJ, Mello MM. Diagnostic Testing for the Novel Coronavirus. JAMA 2020 Mar 09;323(15):1437-1438. [doi: 10.1001/jama.2020.3864] [Medline: 32150622]

5. CDC COVID-19 Response Team. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) - United States, February 12-March 16, 2020. MMWR Morb Mortal Wkly Rep 2020 Mar 27;69(12):343-346 [FREE Full text] [doi: 10.15585/mmwr.mm6912e2] [Medline: 32214079]
6. Clinical Decision Support. The Office of the National Coordinator for Health Information Technology. URL: https://www.healthit.gov/topic/safety/clinical-decision-support [accessed 2020-08-13]
7. Wasylewicz A, Scheepers-Hoeks A. Clinical decision support systems. In: Kubben P, Dumontier M, Dekker A, editors. editors. Fundamentals of Clinical Data Science. Cham (CH): Springer; 2019:153-169.
8. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3]
9. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016 Feb 23;315(8):762-774 [FREE Full text] [doi: 10.1001/jama.2016.0288] [Medline: 26903335]
10. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Med 1996 Jul;22(7):707-710. [doi: 10.1007/bf01709751]
11. Zou X, Li S, Fang M, Hu M, Bian Y, Ling J. Acute Physiology and Chronic Health Evaluation II Score as a predictor of hospital mortality in patients of coronavirus disease 2019. Crit Care Med 2020 May;48(8):e657-e665. [doi: 10.1097/ccm.0000000000004411]
12. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985 Oct;13(10):818-829. [Medline: 3928249]
13. Lilly CM, McLaughlin JM, Zhao H, Baker SP, Cody S, Irwin RS, UMass Memorial Critical Care Operations Group. A multicenter study of ICU telemedicine reengineering of adult critical care. Chest 2014 Mar 01;145(3):500-507. [doi: 10.1378/chest.13-1973] [Medline: 24306581]
14. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data 2018 Sep 11;5:180178 [FREE Full text] [doi: 10.1038/sdata.2018.178] [Medline: 30204154]
15. Duca A, Piva S, Focà E, Latronico N, Rizzi M. Calculated Decisions: Brescia-COVID Respiratory Severity Scale (BCRSS)/Algorithm. Emerg Med Pract 2020 Apr 16;22(5 Suppl):CD1-CD2. [Medline: 32297727]
16. Wallis LA. African Federation for Emergency Medicine resources for managing COVID-19 in low resourced settings. Afr J Emerg Med 2020 Jun;10(2):49 [FREE Full text] [doi: 10.1016/j.afjem.2020.06.001] [Medline: 32566467]
17. ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, et al. Acute respiratory distress syndrome: the Berlin Definition. JAMA 2012 Jun 20;307(23):2526-2533. [doi: 10.1001/jama.2012.5669] [Medline: 22797452]
18. Fan E, Brodie D, Slutsky AS. Acute Respiratory Distress Syndrome: Advances in Diagnosis and Treatment. JAMA 2018 Feb 20;319(7):698-710. [doi: 10.1001/jama.2017.21907] [Medline: 29466596]
19. Singh K, Valley T, Tang S, Li B, Kamran F, Sjoding M, et al. Validating a Widely Implemented Deterioration Index Model Among Hospitalized COVID-19 Patients. medRxiv 2020 Apr 29:1-22 [FREE Full text] [doi: 10.1101/2020.04.24.20079012] [Medline: 32511650]
20. McRae MP, Simmons GW, Christodoulides NJ, Lu Z, Kang SK, Fenyo D, et al. Clinical decision support tool and rapid point-of-care platform for determining disease severity in patients with COVID-19. Lab Chip 2020 Jun 21;20(12):2075-2085. [doi: 10.1039/d0lc00373e] [Medline: 32490853]
21. Yan L, Zhang H, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell 2020 May 14;2(5):283-288. [doi: 10.1038/s42256-020-0180-7]
22. McRae MP, Bozkurt B, Ballantyne CM, Sanchez X, Christodoulides N, Simmons G, et al. Cardiac ScoreCard: A Diagnostic Multivariate Index Assay System for Predicting a Spectrum of Cardiovascular Disease. Expert Syst Appl 2016 Jul 15;54:136-147 [FREE Full text] [doi: 10.1016/j.eswa.2016.01.029] [Medline: 31467464]
23. Buuren SV, Groothuis-Oudshoorn K. MICE: Multivariate Imputation by Chained Equations in R. J Stat Soft 2011;45(3):online. [doi: 10.18637/jss.v045.i03]
24. Campbell I. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Stat Med 2007 Aug 30;26(19):3661-3675. [doi: 10.1002/sim.2832] [Medline: 17315184]
25. Richardson JTE. The analysis of 2 × 2 contingency tables--yet again. Stat Med 2011 Apr 15;30(8):890. [doi: 10.1002/sim.4116] [Medline: 21432882]
26. Guo T, Fan Y, Chen M, Wu X, Zhang L, He T, et al. Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19). JAMA Cardiol 2020 Mar 27;5(7):811-818 [FREE Full text] [doi: 10.1001/jamacardio.2020.1017] [Medline: 32219356]
27. Skates SJ, Pauler DK, Jacobs IJ. Screening Based on the Risk of Cancer Calculation From Bayesian Hierarchical Changepoint and Mixture Models of Longitudinal Markers. J Am Stat Assoc 2001 Jun;96(454):429-439. [doi: 10.1198/016214501753168145]

28. Bleeker S, Moll H, Steyerberg E, Donders A, Derksen-Lubsen G, Grobbee D, et al. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol 2003 Sep;56(9):826-832. [doi: 10.1016/s0895-4356(03)00207-5] [Medline: 14505766]
29. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 2016 Jun 22;353:i3140 [FREE Full text] [doi: 10.1136/bmj.i3140] [Medline: 27334381]
30. McRae MP, Simmons G, Wong J, McDevitt JT. Programmable Bio-nanochip Platform: A Point-of-Care Biosensor System with the Capacity To Learn. Acc Chem Res 2016 Jul 19;49(7):1359-1368 [FREE Full text] [doi: 10.1021/acs.accounts.6b00112] [Medline: 27380817]
31. McRae MP, Simmons GW, Wong J, Shadfan B, Gopalkrishnan S, Christodoulides N, et al. Programmable bio-nano-chip system: a flexible point-of-care platform for bioscience and clinical measurements. Lab Chip 2015 Oct 21;15(20):4020-4031 [FREE Full text] [doi: 10.1039/c5lc00636h] [Medline: 26308851]

Abbreviations

APACHE: Acute Physiology and Chronic Health Evaluation
AUC: area under the curve
CDSS: clinical decision support system
COVID-19: coronavirus disease
CRP: C-reactive protein
cTnI: cardiac troponin I
eICU: electronic intensive care unit
FHC: Family Health Centers
ICU: intensive care unit
NPV: negative predictive value
NT-proBNP: N-terminal fragment of the prohormone brain natriuretic peptide
NYU: New York University
PCT: procalcitonin
PPV: positive predictive value
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
SOFA: Sepsis-related Organ Failure Assessment

Edited by G Eysenbach; submitted 01.07.20; peer-reviewed by KL Ong, J Ropero; comments to author 22.07.20; revised version received 22.07.20; accepted 23.07.20; published 24.08.20.

Please cite as:
McRae MP, Dapkins IP, Sharif I, Anderman J, Fenyo D, Sinokrot O, Kang SK, Christodoulides NJ, Vurmaz D, Simmons GW, Alcorn TM, Daoura MJ, Gisburne S, Zar D, McDevitt JT
Managing COVID-19 With a Clinical Decision Support Tool in a Community Health Network: Algorithm Development and Validation
J Med Internet Res 2020;22(8):e22033
URL: http://www.jmir.org/2020/8/e22033/
doi:10.2196/22033
PMID:32750010

Original Paper

Prognostic Modeling of COVID-19 Using Artificial Intelligence in the United Kingdom: Model Development and Validation

Ahmed Abdulaal[1], MBBS, BSc; Aatish Patel[1], MBBS, BSc; Esmita Charani[2], PhD; Sarah Denny[1], MBBS, BSc; Nabeela Mughal[1], MBBS, MSc; Luke Moore[1,2], MBBS, PhD

[1]Chelsea and Westminster NHS Foundation Trust, London, United Kingdom

[2]NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, Imperial College London, London, United Kingdom

Corresponding Author:
Luke Moore, MBBS, PhD
NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance
Imperial College London
Exhibition Rd, South Kensington
London, SW7 2AZ
United Kingdom
Phone: 44 783 436 6302
Email: l.moore@imperial.ac.uk

Abstract

Background: The current severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) outbreak is a public health emergency and the case fatality rate in the United Kingdom is significant. Although there appear to be several early predictors of outcome, there are no currently validated prognostic models or scoring systems applicable specifically to patients with confirmed SARS-CoV-2.

Objective: We aim to create a point-of-admission mortality risk scoring system using an artificial neural network (ANN).

Methods: We present an ANN that can provide a patient-specific, point-of-admission mortality risk prediction to inform clinical management decisions at the earliest opportunity. The ANN analyzes a set of patient features including demographics, comorbidities, smoking history, and presenting symptoms and predicts patient-specific mortality risk during the current hospital admission. The model was trained and validated on data extracted from 398 patients admitted to hospital with a positive real-time reverse transcription polymerase chain reaction (RT-PCR) test for SARS-CoV-2.

Results: Patient-specific mortality was predicted with 86.25% accuracy, with a sensitivity of 87.50% (95% CI 61.65%-98.45%) and specificity of 85.94% (95% CI 74.98%-93.36%). The positive predictive value was 60.87% (95% CI 45.23%-74.56%), and the negative predictive value was 96.49% (95% CI 88.23%-99.02%). The area under the receiver operating characteristic curve was 90.12%.

Conclusions: This analysis demonstrates an adaptive ANN trained on data at a single site, which demonstrates the early utility of deep learning approaches in a rapidly evolving pandemic with no established or validated prognostic scoring systems.

(J Med Internet Res 2020;22(8):e20259) doi:10.2196/20259

KEYWORDS

COVID-19; coronavirus; machine learning; deep learning; modeling; artificial intelligence; neural network; prediction

Introduction

Since the outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Wuhan, China in December 2019, there have been over 229,705 confirmed cases in the United Kingdom, with a case fatality rate of 14.4% as of May 13, 2020 [1,2]. In the United Kingdom, the highest number of coronavirus disease (COVID-19) deaths (the disease caused by the SARS-CoV-2 virus) has been reported in London [3], with many health care providers having experienced a rapid, difficult-to-predict increase in intensive therapy unit (ITU) bed requirements.

Although there appear to be several early predictors of outcome such as age, high sequential organ failure assessment score and elevated D-dimer levels [4], being male [5], poor glycemic control in patients with diabetes [6], being immunocompromised

[7], and obesity [8], currently there are no validated prognostic models or scoring systems applicable specifically to patients with SARS-CoV-2, despite attempts to delineate general predictors of mortality [9]. Emerging clinical risk scores have been limited by small sample sizes [10], predicting outcomes in suspected as well as confirmed cases [11-14], or using regression analysis to produce static models that have been applied to specific population subgroups, limiting generalizability [15].

Predicting patient-specific adverse events including ITU admissions and mortality with sufficient lead time is crucial during a pandemic, as it allows clinicians, managers, and service providers to admit patients based on risk of deterioration, forecast ITU bed demand, and determine appropriate ceilings of care. From a public health perspective, this would play a significant role in allowing policy makers to respond efficiently to surges of COVID-19, which would otherwise risk overwhelming critical care capacity [16].

There have been significant recent advances in modeling clinical data on electronic health records (EHRs) [17,18] and specifically in the capability of machine learning techniques to predict mortality [19-21]. In view of this, our study proposes an artificial neural network (ANN) that analyzes a set of patient features including demographics, comorbidities, lifestyle factors, and presenting symptoms and predicts patient-specific mortality risk during the current hospital admission. Crucially, this data could be collected during the initial encounter of a patient with a physician, and therefore allows for a prediction of outcome at the earliest opportunity along the patient pathway.

Classically, deep learning approaches created models that were difficult to interpret. This had led clinicians to retreat from complex but accurate techniques to simpler (eg, linear) models [22]. However, significant recent advances have been made in deep learning interpretability research [23] and specifically in creating predictive machine learning models for health care [24-26]. We use an algorithm capable of revealing which features were important for making predictions while maintaining accuracy and consistency [27].

Methods

The ANN was trained on retrospective data extracted from EHRs in a digital format. Demographic, comorbidity, lifestyle, and symptom data were encoded from admission notes of patients admitted to an accident and emergency department at a West London teaching hospital.

Study Population and Data Description

The clinical data used in this study was collected from all hospital admissions for SARS-CoV-2 from February 2, 2020, to April 22, 2020, at a West London teaching hospital. All patients were included in the analyses. Data was anonymized at point of extraction from EHR software (Millennium, Cerner Corporation) and included admission notes, current active medical conditions, and discharge summaries or electronic certifications of death.

The inclusion criteria included all patients with real-time reverse transcription polymerase chain reaction (RT-PCR) test–confirmed SARS-CoV-2 (proprietary Public Health England Assay until March 10, 2020, and an AusDiagnostics assay thereafter). A SARS-CoV-2 infection must have been the principal diagnosis and the reason for that admission episode. Outcome data including the presence of either a discharge summary or electronic certification of death were collected for each patient. SARS-CoV-2 mortality was defined as an in-patient death, which occurred during the current admission episode, in patients with confirmed SARS-CoV-2.

Data Preprocessing

Individual patients were represented as an array of possible prognostic factors. Demographic factors included age and gender. Comorbidities included the presence or absence of chronic obstructive pulmonary disease, asthma, or a chronic respiratory disease; hypertension, diabetes, ischemic heart disease, congestive cardiac failure, hepatic cirrhosis, chronic kidney disease, or a cerebrovascular event history. Smoking history was also collected. Symptom data included the number of days of symptoms prior to hospital admission and the presence or absence of fever, cough, dyspnea, myalgia, abdominal pain, diarrhea or vomiting, altered mentation, collapse, and olfactory change or ageusia. These data points represented a feature set that was then used as the input for the ANN. Information regarding patient ethnicity was not felt to be robust due to 23.4% missing data. Ethnic subgroups are reported using descriptive statistics; however, ethnicity is not included in the ANN as a variable.

All data was encoded (by AA, AP, and EC) after reading the admission episode notes for each patient. Comorbidity, lifestyle, and symptom data were encoded as binary presence features. Age was recorded as a discrete quantitative feature. Gender was recorded as a categorical variable, which was later encoded as a numerical binary feature. All numerical features were standardized (centered and scaled) by subtracting the mean and dividing by the standard deviation of the training samples. The target value was defined as an in-patient death in a patient clinically suspected of having SARS-CoV-2 on admission and who had a positive RT-PCR test. Deaths were counted if they occurred during the admission episode for which outcome predictions were made. Mortality was encoded as a binary target value.

Model for Predicting SARS-CoV-2 Outcomes

Overview

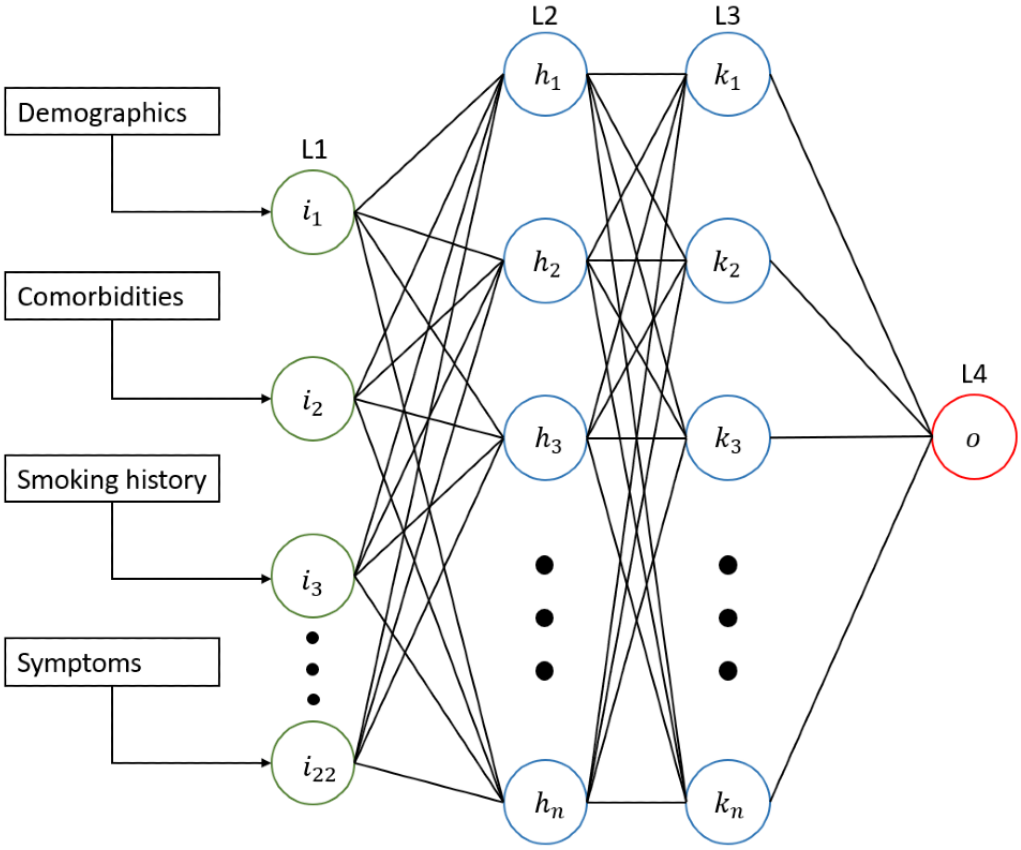
The ANN used clinical data accrued from the admission notes to predict mortality for that admission. Input features were provided to the system, and the output is the probability of death for a patient during their current admission. Patients were randomized to training (80%) and testing (20%) sets. Data for each patient were randomized to only one of these sets. To reduce model overfitting, k-fold cross-validation was used. Ten folds were chosen during training, which represented 10% of the training sets being used as validation sets. If the probability of mortality was above a threshold of 50%, the prediction was

considered positive in that the model predicted the patient was likely to experience a poor outcome.

Artificial Neural Network Input and Core

Figure 1 demonstrates the ANN. The input layer had an input dimension equal to the number of patient features and used rectifier activation (n=22). Information was then fed into two densely connected further layers (known as hidden layers), which also used rectifier activation. Hidden layers are input-output transformation of incoming data. Each layer attempts to create increasingly meaningful representations of the input data (patient variables) before attempting to make an outcome prediction.

Figure 1. Architecture of an artificial neural network to prognosticate severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in a UK population. L1 represents the input layer, which accepts patient features as input. L2 and L3 represent two densely connected hidden layers, with a variable number of units. L4 represents the output layer which predicts mortality for a given patient.



Prediction Targets and Training Objectives

The output layer consisted of a single node which used sigmoid activation to provide a probability for the target value. In other words, the output layer collects the data representations from the prior hidden layers and predicts whether a patient is likely to die during their admission. Each resulting probability output was compared to the ground-truth label using the cross-entropy loss function, which measures the performance of a classification model whose output is a probability value between 0 and 1 [28]. This means that the ANN checks whether the prediction it made was correct against known patient outcomes.

Training and Hyperparameters

Model architecture was chosen based on validation set performance. A grid search technique was used to exhaustively search over parameters in an iterative process to establish optimal model parameters (known as hyperparameters). All variables were initialized with normalized uniform (Xavier normal) initialization [29] and trained using the Adam adaptive learning rate optimization algorithm [30]. The input layer had an input dimension to match the number of patient variables (n=22). Optimal validation results were achieved with an ANN with 22 units in the first hidden layer with a dropout rate of 20%, and 6 units in the second and third hidden layers with a dropout rate of 40% to prevent overfitting. To further prevent overfitting, L2 regularization is used on all layers except for the output layer.

Relative Importance of Clinical Attributes

We used a high-speed approximation algorithm for Shapley additive explanations (SHAP) values, which in effect reveal the contribution of each patient variable (clinical attribute) to their mortality prediction against a mean prediction [30,31].

Evaluation of Model

A validation set represented by 10% of the training set was used during cross-validation and a cross-validated grid search. The validation set was used to improve model architecture and select for optimal hyperparameters. The metrics selected for model performance evaluation were accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and the area under the receiver operating characteristic curve (AUROC).

K-fold cross-validation allowed a mean model accuracy with 95% CI to be calculated.

Code Availability

The open source machine learning framework Tensorflow 2.1.0 [32] was used to develop the neural network. The architecture was written in the Python programming language (Python 3.7.7). Scikit-learn 0.22 and its dependencies were used to create the data preprocessing pipeline and to create the graphs in this analysis.

Ethical Considerations

Data was collected as part of routine care by the responsible clinical team. No patient-identifiable data was used in this analysis. The need for written informed consent was waived by the Research Governance Office of Chelsea & Westminster NHS Foundation Trust. The study protocol was approved by the antimicrobial stewardship group at Chelsea & Westminster NHS Foundation Trust. The study was conducted in accordance with the Helsinki declaration.

Results

Patient Demographics, Comorbidities, and Symptoms

There were a total of 11,144 data points for 398 patients, encompassing 22 input features. The training and testing populations consisted of n=318 and n=80 patients, respectively. Out of the 398 patients included in the analysis, 389 (97.8%) had completed outcomes and 9 (2.2%) were still hospitalized at the end of the study period. There were 223 (56%) males. Of patients with completed outcomes, 275 (69%) were discharged alive and 93 (23%) died. There were 53 admissions to the ITU (13%), of which 17 (32%) died.

The median age of all patients was 65 years (IQR 51-80). The median age of patients who were admitted to the ITU was 56 years (IQR 51-65) and the median age of patients who died was 79 years (IQR 72-86). Regarding ethnicity, 157 (39.4%) patients were White, 37 (9.3%) were Black, 42 (10.6%) were Asian, 66 (16.6%) were from other ethnicities, and 3 (0.7%) were of mixed ethnicity. In total, 93 (23.4%) did not have a recorded ethnicity. The median time from symptom onset to hospital admission was 5 days (IQR 2-10). Mean length of stay for all patients with completed outcomes was 9.8 days (SD 9.6). Mean length of stay for patients who died was 8.5 days (SD 7.9).

The most common comorbidities were hypertension (n=147; 37%), diabetes (n=104; 26%), and chronic respiratory disease (n=84; 21%). The most common presenting symptoms were cough (n=247; 62%), dyspnea (n=223; 56%), and fever (n=216; 54%). Tables 1 and 2 demonstrate comorbidities and presenting symptoms in order of prevalence, respectively.

Table 1. Prevalence of comorbidities in a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) population in West London.

Comorbidity	Patients, n (%)
Hypertension	147 (36.9)
Diabetes	104 (26.1)
Chronic obstructive pulmonary disease, asthma, or chronic respiratory pathology	84 (21.1)
Ischemic heart disease	47 (11.8)
Chronic kidney disease	33 (8.3)
Cerebrovascular event history	29 (7.3)
Cardiac failure	22 (5.5)
Obesity	15 (3.7)
Hepatic cirrhosis	6 (1.5)

Table 2. Prevalence of symptoms in a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) population in West London.

Symptom	Patients, n (%)
Cough	247 (62.1)
Dyspnea	223 (56.0)
Fever	216 (54.2)
Diarrhea or vomiting	105 (26.4)
Myalgia	68 (17.1)
Altered mentation	59 (14.8)
Abdominal pain	40 (10.1)
Collapse	37 (9.3)
Anosmia or ageusia	36 (9.0)

ANN Performance

With this ANN, cross-validated accuracy (accuracy on the training and validation set) was 89% (95% CI 81%-97%). Patient-specific mortality was predicted with 86.25% accuracy on the test set (Figure 2), with a sensitivity of 87.50% (95% CI 61.65%-98.45%) and specificity of 85.94% (95% CI 74.98%-93.36%). The positive predictive value was 60.87% (95% CI 45.23%-74.56%), and the negative predictive value was 96.49% (95% CI 88.23%-99.02%). Binary cross-entropy loss is demonstrated in Figure 3. The AUROC was 90.12% (Figure 4).

Figure 2. Model accuracy of an artificial neural network prognostic model for coronavirus disease (COVID-19) in a UK population. Model accuracy is based on training and test set results per epoch. Accuracy is defined as (TP+TN)/(TP+TN+FP+FN), where TP=true positive, TN=true negative, FP=false positive, and FN=false negative. One epoch represents one full cycle through the training set data. As the model trains for a greater number of epochs, the accuracy (ie, its ability to predict true positives and true negatives relative to all outcomes) increases.

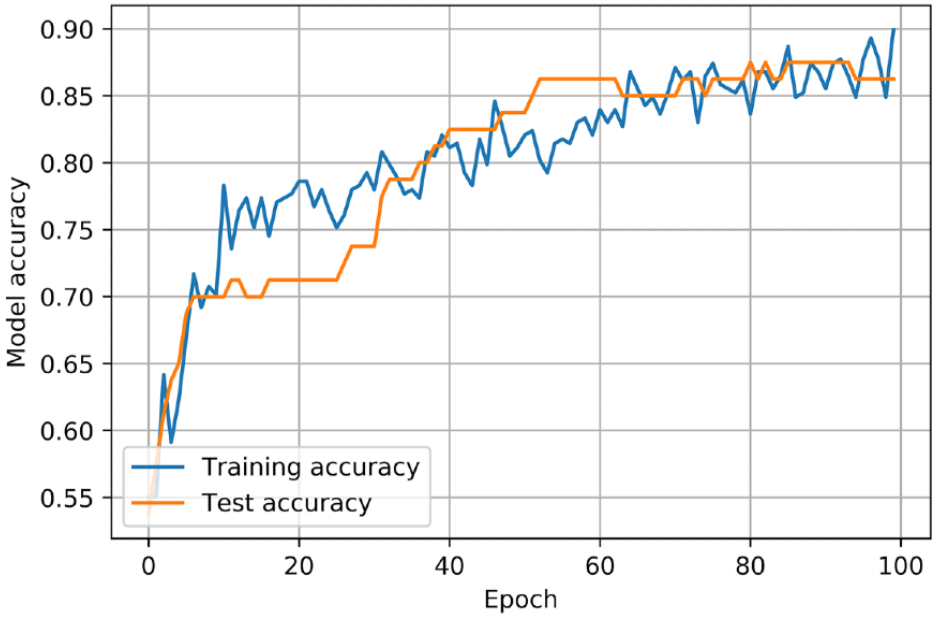


Figure 3. Performance metric of an artificial neural network prognostic model for coronavirus disease (COVID-19) in a UK population. Binary cross-entropy loss for training and tests per epoch. Cross-entropy loss measures the performance of a model that outputs a prediction between 0 and 1. It is a measure of how far the predictions made by the model are from the truth. As loss decreases, the probabilities estimated by the model match the actual target value (in this case, correct mortality predictions) more closely.

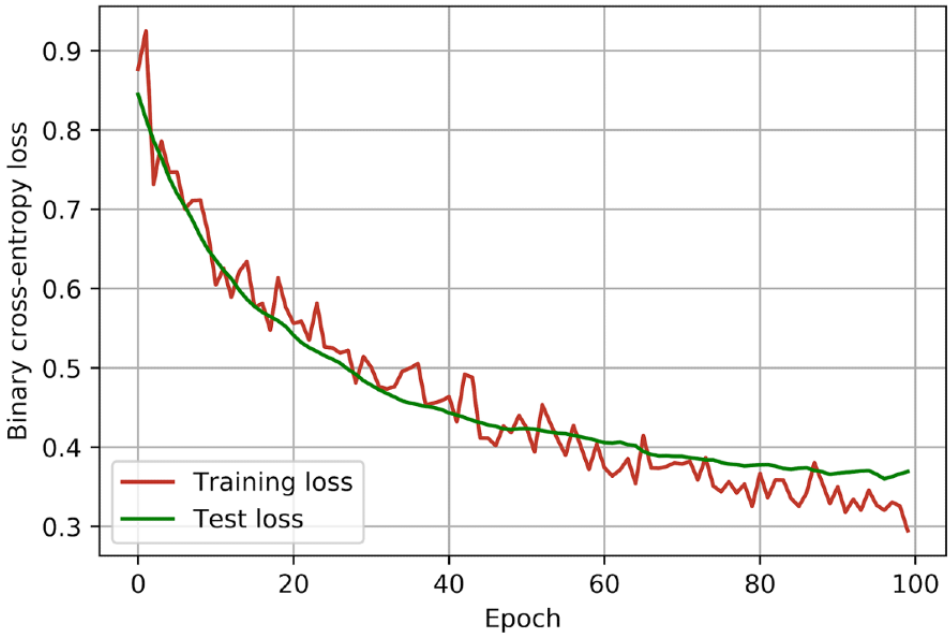
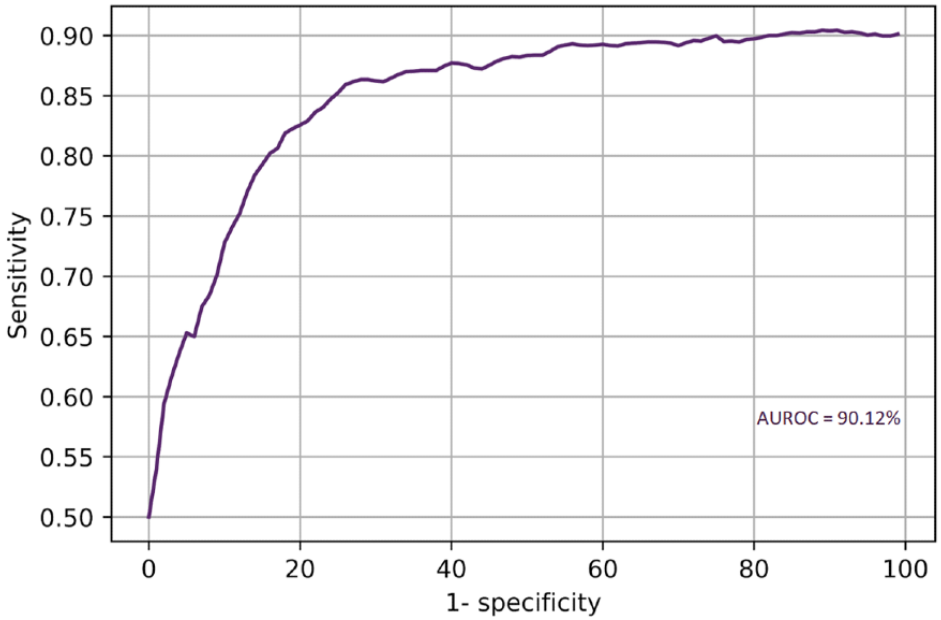


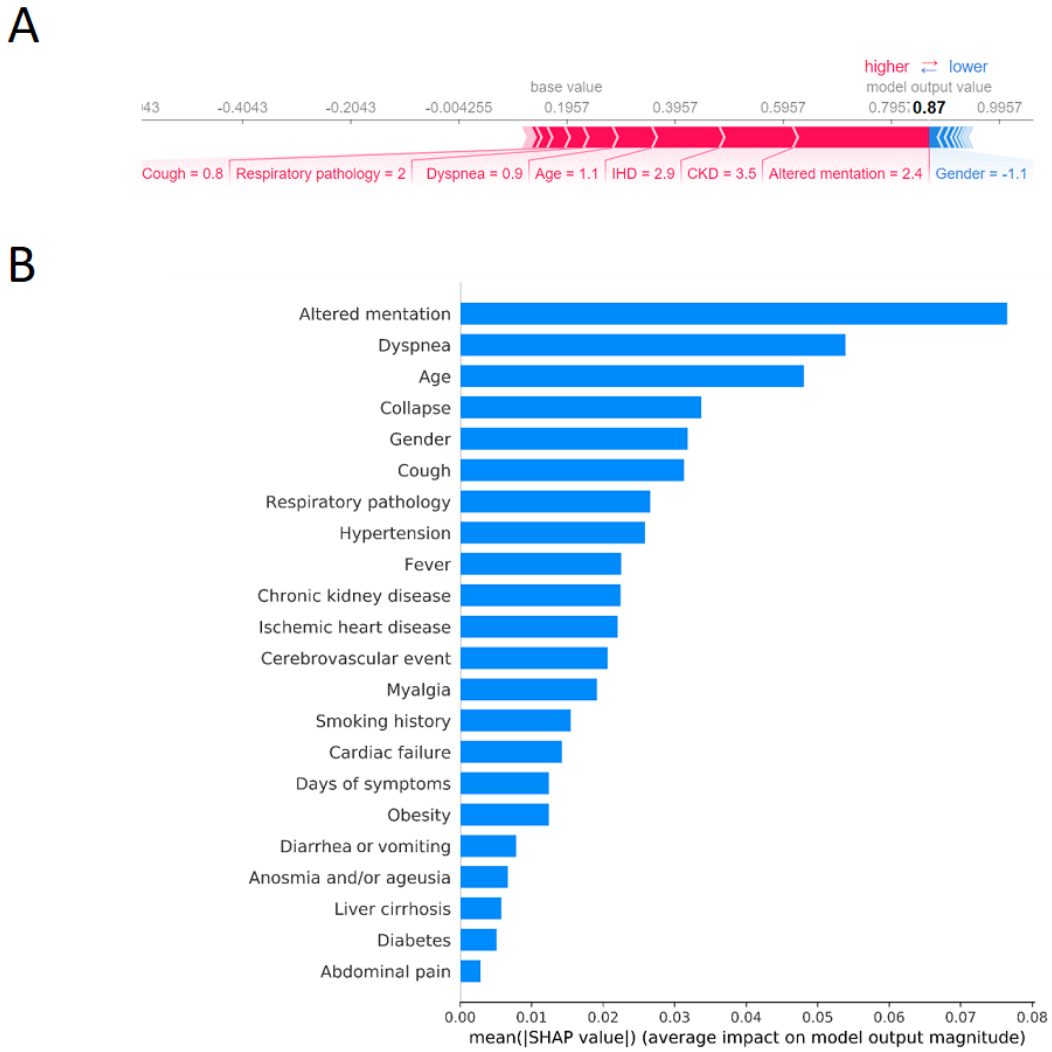
Figure 4. Receiver operating characteristic curve for an artificial neural network prognostic model for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in a UK population. AUROC: area under the receiver operating characteristic curve.



Relative Importance of Clinical Attributes

The approximation algorithm for SHAP values could reveal feature importance on a patient-specific basis (Figure 5A). The model dynamically adjusts mortality risk prediction for each patient, and illustrates the predictors it used (and their relative importance) to form the prediction. Overall feature importance for the model was also calculated (Figure 5B). Altered mentation, new dyspnea, and increasing age were the most significant predictors of mortality. Moderate predictors of mortality were collapse, male gender, new cough, and known respiratory pathology.

Figure 5. Relative importance of clinical attributes in an artificial neural network prognostic model for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in a UK population. A) An example of an accurate mortality prediction. Feature importance is proportional to feature bar width. Features which increased risk of death are shown in red, and those which decreased risk are shown in blue. Altered mentation, ischemic heart disease (IHD), and chronic kidney disease (CKD) were the most salient predictors of mortality for this patient. Female gender was protective. B) Overall feature importance as considered by the model. SHAP: Shapley additive explanations.



Discussion

Principal Findings

In this analysis, we provide details of an ANN capable of predicting patient-specific mortality with high sensitivity and specificity. Furthermore, the model provides information on which features are most salient when predicting risk, delivering explainable predictions on a patient-specific basis to clinicians and potentially allowing more informed discussions with patients and relatives.

Our aim was to provide a patient-specific, point-of-admission mortality risk prediction to help inform clinical management decisions at the earliest opportunity. The contribution of this analysis is in the proof-of-concept ANN trained on data from a single site, which demonstrates the early utility of deep learning approaches in a rapidly evolving pandemic with no established or validated prognostic scoring systems. Intensivists and respiratory physicians can be alerted of a patient with a higher relative risk of deterioration at an earlier stage, and ITU departments can better anticipate bed needs and adjust staffing and capacity appropriately.

Altered mentation, dyspnea, and increasing age were found to be the most salient overall features in predicting mortality. Moderate predictors of mortality included collapse, male gender, new cough, and previous respiratory pathology. These features are broadly in line with the current literature [31].

Of note is smoking history, which appears less important to the model than might be intuitively assumed. Smoking history was encoded as a presence feature, meaning that current smokers were grouped with ex-smokers, and this may provide an explanation as to why smoking history was considered as a more minor feature by the ANN. Indeed, the largest study to investigate factors associated with SARS-CoV-2 deaths to date (n=5683 SARS-CoV-2–linked deaths) demonstrated a lower risk of death in current smokers (hazard ratio 0.88; 0.79-0.99) but a higher risk in ex-smokers (hazard ratio 1.25; 1.18-1.33) [33]. Although this may suggest smokers are underrepresented

in groups with severe disease, a protective mechanism related to nicotine function has been suggested [34].

An advantage of our current model is that all demographic, comorbidity, lifestyle, and symptom data can be collected on first encounter with a physician, and therefore an early outcome prediction can be produced following clerking. Unlike previous work [11,12], the intended use of this prognostic model as a point-of-admission mortality risk predictor is clearly described. Furthermore, the ANN models outcomes in the context of current SARS-CoV-2 practice guidelines and is therefore directly applicable to the current cohort of hospitalized patients with confirmed SARS-CoV-2.

There were several limitations to consider in our analysis. Although the ANN is representative of hospitalized patients with confirmed SARS-CoV-2 and their outcomes within the geographic remit of the study site, these results should be generalized with caution to other populations. Validating the predictive ability of the model would require prospective studies with a larger number of patients across multiple sites in the United Kingdom. In addition, we were unable to account for patients who were admitted for clinically suspected SARS-CoV-2 and subsequently tested positive for the virus, but who may have died due to an unrelated morbidity. However, such patients likely represent a small minority of our cohort. There were 9 patients (2.2% of our data set) who were alive at the end of the study period but had not yet been discharged. They were considered as alive in the analysis. However, these censored patients may be more likely to have a favorable outcome, as we found that the mean length of stay for patients who died was shorter relative to the whole cohort. Mean length of stay for all patients was 9.8 days, and therefore the ANN would be unsuitable for use to predict outcomes for significantly longer durations than this. Lastly, the current model does not include other potentially important predictors of outcomes, including hematological, biochemical, radiological, microbiological, and histological results where appropriate. The ANN architecture is such that adding further input variables (including clinical investigation results) is easily achievable. We plan to extend the ANN in the future with these parameters to maximize its predictive capability.

More complex deep learning models such as recurrent neural networks (RNNs) allow for time-series forecasting and have been successfully used to predict outcomes in real time [17]. Use of RNNs in the future would allow for real-time outcomes predictions throughout an individual admission and could account for factors such as ITU admissions as they occur.

Conclusions

Increasingly, hospitals document and store patient data in EHRs, and machine learning techniques are becoming more ubiquitous in health care. In the context of an evolving pandemic with no established prognostic scoring system, deep learning approaches can be used to rapidly develop empirical prognostic models. These models have the inherent advantage of becoming progressively more accurate and representative as data sets increase in size. With larger, more representative data sets and more accurate artificial intelligence models, it may be possible for patient-specific outcome predictions to help guide physicians to tailor management and establish appropriate ceilings of care more generally.

Acknowledgments

LM acknowledges academic support from the National Institute of Health Research (NIHR) Imperial Biomedical Research Centre (BRC) and the National Institute for Health Research Health Protection Research Unit (HPRU) in Healthcare Associated Infection and Antimicrobial Resistance at Imperial College London in partnership with Public Health England. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the UK Department of Health.

Authors' Contributions

AA and LM designed the study methodology. AA, AP, and EC collated the data. All authors reviewed the themes during data analysis and contributed comments. AA drafted the initial manuscript with all authors contributing significantly to revising the document for submission. All authors agreed on the final version for submission to the journal.

Conflicts of Interest

LM has consulted for bioMerieux (2013-2020), DNAelectronics (2015), Dairy Crest (2017-2018), Pfizer (2018-2020), and Umovis Lab (2020), received speaker fees from Profile Pharma (2018), received research grants from the National Institute for Health Research (2013-2019), Leo Pharma (2016), and CW+ Charity (2018-2019), and received educational support from Eumedica (2016-2017). All other authors have no conflicts of interest to declare.

References

1. Mortality Analyses - Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/data/mortality [accessed 2020-05-11]
2. Coronavirus cases in the UK: daily updated statistics. URL: https://www.gov.uk/guidance/coronavirus-covid-19-information-for-the-public [accessed 2020-05-11]

3. Office for National Statistics. Deaths registered weekly in England and Wales, provisional. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredweeklyinenglandandwalesprovisional/weekending10april2020 [accessed 2020-05-11]
4. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet 2020 Mar;395(10229):1054-1062. [doi: 10.1016/s0140-6736(20)30566-3]
5. Zhang J, Wang X, Jia X, Li J, Hu K, Chen G, et al. Risk factors for disease severity, unimprovement, and mortality in COVID-19 patients in Wuhan, China. Clin Microbiol Infect 2020 Jun;26(6):767-772 [FREE Full text] [doi: 10.1016/j.cmi.2020.04.012] [Medline: 32304745]
6. Zhu L, She Z, Cheng X, Qin JJ, Zhang XJ, Cai J, et al. Association of Blood Glucose Control and Outcomes in Patients with COVID-19 and Pre-existing Type 2 Diabetes. Cell Metab 2020 Jun 02;31(6):1068-1077.e3 [FREE Full text] [doi: 10.1016/j.cmet.2020.04.021] [Medline: 32369736]
7. Montagud-Marrahi E, Cofan F, Torregrosa J, Cucchiari D, Ventura-Aguiar P, Revuelta I, et al. Preliminary data on outcomes of SARS-CoV-2 infection in a Spanish single center cohort of kidney recipients. Am J Transplant 2020 May 05:1-2 [FREE Full text] [doi: 10.1111/ajt.15970] [Medline: 32368838]
8. Simonnet A, Chetboun M, Poissy J, Raverdy V, Noulette J, Duhamel A, LICORN and the Lille COVID-19 and Obesity study group. High Prevalence of Obesity in Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) Requiring Invasive Mechanical Ventilation. Obesity (Silver Spring) 2020 Jul 10;28(7):1195-1199 [FREE Full text] [doi: 10.1002/oby.22831] [Medline: 32271993]
9. Yang H, Zhang Y, Yang M, Huang X. Re: Predictors of mortality for patients with COVID-19 pneumonia caused by SARSCoV-2: a prospective cohort study. Eur Respir J 2020 Aug 03:2002439 [FREE Full text] [doi: 10.1183/13993003.02439-2020] [Medline: 32747392]
10. Bai X, Fang C, Zhou Y, Bai S, Liu Z, Xia L, et al. Predicting COVID-19 Malignant Progression with AI Techniques. medRxiv 2020 Mar 27:1. [doi: 10.1101/2020.03.20.20037325]
11. Caramelo F, Ferreira N, Oliveiros B. Estimation of risk factors for COVID-19 mortality - preliminary results. medRxiv 2020:1 [FREE Full text] [doi: 10.1101/2020.02.24.20027268]
12. Lu J, Hu S, Fan R, Liu Z, Yin X, Wang Q, et al. ACP Risk Grade: A Simple Mortality Index for Patients with Confirmed or Suspected Severe Acute Respiratory Syndrome Coronavirus 2 Disease (COVID-19) During the Early Stage of Outbreak in Wuhan, China. MedRxiv 2020:1. [doi: 10.1101/2020.02.20.20025510]
13. Yan L, Zhang H, Xiao Y, Wang M, Sun C, Liang J, et al. Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. MedRxiv 2020:1. [doi: 10.1101/2020.02.27.20028027]
14. Yuan M, Yin W, Tao Z, Tan W, Hu Y. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. PLoS One 2020 Mar 19;15(3):e0230548 [FREE Full text] [doi: 10.1371/journal.pone.0230548] [Medline: 32191764]
15. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, China Medical Treatment Expert Group for COVID-19. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA Intern Med 2020 May 12;180(8):1081 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2033] [Medline: 32396163]
16. Phua J, Weng L, Ling L, Egi M, Lim CM, Divatia JV, Asian Critical Care Clinical Trials Group. Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. Lancet Respir Med 2020 May;8(5):506-517 [FREE Full text] [doi: 10.1016/S2213-2600(20)30161-2] [Medline: 32272080]
17. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 2019 Aug 31;572(7767):116-119 [FREE Full text] [doi: 10.1038/s41586-019-1390-1] [Medline: 31367026]
18. Li M, Leslie H, Qi B, Nan S, Feng H, Cai H, et al. Development of an openEHR Template for COVID-19 Based on Clinical Guidelines. J Med Internet Res 2020 Jun 10;22(6):e20239 [FREE Full text] [doi: 10.2196/20239] [Medline: 32496207]
19. Lim B, van der Schaar M. Disease-Atlas: Navigating Disease Trajectories with Deep Learning. ArXiv 2018 Mar 27:1 [FREE Full text]
20. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018 May 8;1(1):18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
21. Henao R, Lu J, Lucas J, Ferranti J, Carin L, Henao C. Electronic Health Record Analysis via Deep Poisson Factor Models. Phenotype KnowledgeBase. 2015. URL: https://phekb.org/ [accessed 2020-05-11]
22. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018 Oct 10;2(10):749-760 [FREE Full text] [doi: 10.1038/s41551-018-0304-0] [Medline: 31001455]
23. Xie N, Ras G, van Gerven M, Doran D. Explainable Deep Learning: A Field Guide for the Uninitiated. ArXiv 2020 Apr 30:A [FREE Full text]
24. ElMoaqet H, Tilbury DM, Ramachandran SK. Multi-Step Ahead Predictions for Critical Levels in Physiological Time Series. IEEE Trans Cybern 2016 Jul;46(7):1704-1714. [doi: 10.1109/tcyb.2016.2561974]

25. Lipton Z, Kale D, Wetzel R. Phenotyping of Clinical Time Series with LSTM Recurrent Neural Networks. ArXiv 2015:1-5 [FREE Full text]
26. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med 2015 Aug 05;7(299):299ra122. [doi: 10.1126/scitranslmed.aab3719] [Medline: 26246167]
27. Lundberg S, Allen P. A Unified Approach to Interpreting Model Predictions. GitHub. 2017. URL: https://github.com/slundberg/shap [accessed 2020-05-11]
28. Loss Functions - ML Glossary documentation. ML Cheatsheet. URL: https://ml-cheatsheet.readthedocs.io/en/latest/loss_functions.html [accessed 2020-05-13]
29. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. URL: http://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf [accessed 2020-05-13]
30. Kingma D, Ba J. Adam: A method for stochastic optimization. ArXiv 2014:1-15 [FREE Full text]
31. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 16,749 hospitalised UK patients with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol. MedRxiv 2020 Apr 28:1-21 [FREE Full text] [doi: 10.1101/2020.04.23.20076042]
32. Tensorflow. GitHub. URL: https://github.com/tensorflow/tensorflow [accessed 2020-08-11]
33. The OpenSAFELY Collaborative, Williamson E, Walker A, Bhaskaran KJ, Bacon S, Bates C, et al. OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. MedRxiv 2020 May 07:1-22 [FREE Full text] [doi: 10.1101/2020.05.06.20092999]
34. Simons D, Shahab L, Brown J, Perski O. The association of smoking status with SARS-CoV-2 infection, hospitalisation and mortality from COVID-19: A living rapid evidence review (version 6). Qeios 2020 Aug 02:A. [doi: 10.32388/ujr2aw.7]

Abbreviations

ANN: artificial neural network
AUROC: area under the receiver operating characteristic curve
COVID-19: coronavirus disease
EHR: electronic health record
RNN: recurrent neural network
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
SHAP: Shapley additive explanations

Edited by G Eysenbach; submitted 14.05.20; peer-reviewed by W Sun, L Zhang, H Demir; comments to author 20.06.20; revised version received 25.06.20; accepted 24.07.20; published 25.08.20.

Please cite as:
Abdulaal A, Patel A, Charani E, Denny S, Mughal N, Moore L
Prognostic Modeling of COVID-19 Using Artificial Intelligence in the United Kingdom: Model Development and Validation
J Med Internet Res 2020;22(8):e20259
URL: http://www.jmir.org/2020/8/e20259/
doi:10.2196/20259
PMID:32735549

Original Paper

Impact of the COVID-19 Epidemic on Lifestyle Behaviors and Their Association With Subjective Well-Being Among the General Population in Mainland China: Cross-Sectional Study

Zhao Hu[1], MD; Xuhui Lin[1], MD; Atipatsa Chiwanda Kaminga[2,3], MD; Huilan Xu[1], PhD

[1]Department of Social Medicine and Health Management, Xiangya School of Public Health, Central South University, Changsha, China

[2]Department of Mathematics and Statistics, Mzuzu University, Luwinga, Mzuzu, Malawi

[3]Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China

Corresponding Author:
Huilan Xu, PhD
Department of Social Medicine and Health Management
Xiangya School of Public Health
Central South University
238 Shangmayuanling
Changsha, 410078
China
Phone: 86 073184135459
Email: xhl6363@sina.com

Abstract

Background: The world is experiencing an unprecedented challenge due to the coronavirus disease (COVID-19) pandemic. However, it is unclear whether people's lifestyles will change as a result.

Objective: The aim of this study is to explore perceived lifestyle changes after the outbreak of COVID-19 and their association with subjective well-being (SWB) among the general population in Mainland China.

Methods: An online survey was conducted in May 2020. Lifestyle behaviors including leisure-time physical exercise, leisure-time screen time, and dietary intake were self-reported. SWB was measured using the General Wellbeing Schedule (GWS). Other covariates including sociodemographic factors, self-rated physical health, perceived social support, and loneliness were also assessed by a structured questionnaire. A multivariate ordinal regression method was used to analyze the association between SWB and lifestyle behaviors as well as perceived lifestyle changes.

Results: A total of 1033 participants aged between 18 and 60 years were included in this study. The mean GWS score was 71.7 points. About 70% of the respondents reported spending more time looking at screens, whereas about 30% reported an increased frequency of vegetable and fruit intake after the outbreak of COVID-19. Inactive physical exercise (odds ratio [OR] 1.16, 95% CI 1.02-1.48), infrequent vegetable intake (OR 1.45, 95% CI 1.10-1.90), infrequent fruit intake (OR 1.31, 95% CI 1.01-1.70), and often skipping breakfast (OR 1.43, 95% CI 1.08-1.91) were associated with lower SWB after adjusting for sociodemographic factors, self-rated physical health, perceived social support, and loneliness. Moreover, participants who perceived a decrease in the frequency of vegetable, fruit, and breakfast intake were more likely to report lower SWB.

Conclusions: The COVID-19 pandemic may have positive and negative impacts on different aspects of lifestyle behaviors. Both unhealthy lifestyle behaviors and negative lifestyle changes were associated with lower SWB. These findings provide scientific evidence that can inform lifestyle guidelines and public mental health interventions during the COVID-19 outbreak.

(J Med Internet Res 2020;22(8):e21176) doi:10.2196/21176

KEYWORDS

COVID-19; coronavirus disease; subjective well-being; lifestyle behaviors

Introduction

The ongoing coronavirus disease (COVID-19) outbreak has led to an unprecedented public health crisis worldwide [1]. COVID-19, similar to severe acute respiratory syndrome (SARS), is a beta-coronavirus that can be spread to humans through intermediate hosts such as bats [2]. The COVID-19 outbreak was first revealed in late December 2019 in Wuhan, Hubei Province, China. On January 30, 2020, the World Health Organization (WHO) declared the COVID-19 outbreak a public health emergency of international concern [3]. In the following months, COVID-19 spread rapidly in China and worldwide.

At the same time, the global population is experiencing life-altering challenges due to the COVID-19 pandemic [4]. Given the lack of effective treatment for COVID-19, nonpharmacological interventions (NPIs) are required to decrease its transmission [5]. In essence, NPIs are intended to modify disease-related lifestyle behaviors such that these should no longer contribute to the spread of the disease. However, lifestyle choices or behavioral change may lead to unforeseen detrimental or protective consequences for mental health outcomes [6-10]. For example, some NPIs, such as personal restrictions, mass confinement, and compulsory home isolation, may be associated with worse psychological conditions [8], whereas some NPIs, such as good ventilation in the workplace and wearing a face mask, were found to be protective of mental health during the COVID-19 pandemic [9,10]. Furthermore, evidence from recent studies suggests that current pandemic-related coping strategies may have an adverse impact on mental health, such as decreased well-being and increased posttraumatic stress disorders (PTSD), depression and anxiety symptoms, insomnia, and anger [11-13]. In addition, fear of the disease and social isolation may lead to stress reactions that could develop into other psychological disorders [14].

Unhealthy lifestyle habits such as poor diet, lack of physical activity, smoking, and alcohol use are not only major contributors to the global burden of disease [15], but are also positively associated with worse mental health outcomes [16]. Recently, lifestyle guidelines have emphasized maintaining a healthy nutritional status and engaging in physical exercise at home during the COVID-19 outbreak [17,18]. Many other studies have suggested that the focus should be on addressing the mental health aspect when implementing public disease control and prevention interventions [19,20]. However, observational studies on the characteristics of behavior patterns such as dietary intake, physical exercise, and screen time among the general population after the outbreak of COVID-19 are lacking [21]. Similarly, data about perceived lifestyle changes among the general population after the outbreak of COVID-19 are also lacking. Moreover, the associations between mental health outcomes and lifestyle behaviors as well as lifestyle changes during the COVID-19 pandemic represent a research gap.

Therefore, this study aimed to explore the perceived lifestyle changes after the outbreak of COVID-19, and their association with subjective well-being among the general population in Mainland China. It is hoped that, from the public health and preventative care perspectives, this study can provide valuable information to inform public health policies or interventions aimed at maintaining good mental health and a healthy lifestyle during the COVID-19 pandemic.

Methods

Overview

This study was conducted from May 10 to May 15, 2020, through an online survey using an internet platform [22]. This platform has more than 2.6 million members distributed across more than 30 provinces in China, of which 52.00% are female, 29.34% are aged 26 to 30 years, and 39.20% are general staff. All participants signed an online informed consent form before filling out this questionnaire. The target population comprised all Chinese people in Mainland China aged between 18 and 60 years. A computer-assisted simple random sampling method was used to select eligible participants from this internet platform. Adolescents were not interviewed because it would be difficult to obtain parental consent over the internet, and older adults (≥60 years old) were not investigated as recall bias may be strong in this group. Participants were also excluded if they refused to participate in this survey, did not reply, or could not complete the online survey independently.

Data Collection

Sociodemographic Information

A self-administrated structured questionnaire was used to collect the participants' sociodemographic information including age, gender (male or female), residential location (urban or rural area), marital status (married or unmarried), education (high school and below, college or university degree, master's degree and above) and personal monthly income. The unmarried group included those who were divorced, single, and widowed.

Lifestyle Behaviors Assessment

Lifestyle behaviors in the 4 months immediately before and the 4 months immediately after the COVID-19 outbreak were assessed. In addition, the perceived lifestyle changes after the COVID-19 outbreak were assessed in relation to lifestyle behaviors in the 4 months before the COVID-19 outbreak. The specific lifestyle behaviors that were assessed included the following: leisure-time physical exercise, leisure-time screen time, and dietary intake. To measure leisure-time physical exercise, four items from the long version of the International Physical Activity Questionnaire (IPAQ), a valid and reliable questionnaire assessing physical activity [23], were used. Of the four items, two assess the frequency of moderate or vigorous physical exercise (such as running and dancing) per week and the other two items assess the duration of physical exercise every time. Participants were defined as active if they reported a total moderate-vigorous physical exercise time of more than 150 minutes/week [24]. The Cronbach α of these items in this study was 0.685. Leisure-time screen time was viewed as a major indicator of sedentary behavior, and was calculated by summing up the screen time related to watching TV/videos, internet use (such as news and Douban) through a smartphone or the internet, playing computer or smartphone games, studying online, and social platform use (such as QQ and Wechat).

Leisure-time screen time was further classified as short (<2 hours/day) and long (≥2 hours/day). Three items from the simplified food frequency questionnaire (FFQ-25) [25] were used to assess the weekly frequency of consuming vegetables (including light- and deep-colored vegetables) and fruits (such as apples and pears). Thus, participants were dichotomized as frequently (≥5 times/week) and infrequently (<5 times/week) consuming fruits and vegetables. The Cronbach α of these items in this study was 0.725. The question "How many days do you skip breakfast weekly after the outbreak of COVID-19?" was used to assess the frequency of skipping breakfast and this dichotomized participants as seldom (<3 days/week) and often (≥3 days/week). Perceived lifestyle changes compared to lifestyle behaviors before the outbreak of COVID-19 were classified into three categories as follows: increased or much increased, same as before, and decreased or much decreased.

Covariates

Other covariates assessed were self-rated physical health, perceived social support (PSS), and loneliness during the COVID-19 epidemic. Self-rated physical health was assessed using a 5-item Likert scale. For example, one of the questions asked, "How do you feel about your physical condition?" Participants rated their physical condition as good, fair, or poor. PSS was measured using the Chinese version of the Perceived Social Support Scale (PSSS). The PSSS was developed by Zimet in 1988 [26] and was used to measure PSS from three sources: family, friends, and significant others. This scale contain 12 items, scored on a 7-point rating scale from 1 (very strongly disagree) to 7 (very strongly agree). The total score ranges from 12 to 84, and PSS is divided into low (a score from 12 to 36), intermediate (a score from 37 to 60), and high (a score from 61 to 84). The Cronbach α of this scale in this study was 0.881. Loneliness was measured using the short from of the University of California, Los Angeles (UCLA) Loneliness Scale (ULS-8) [27], which was adapted by Hays and DiMatteo in 1987 on the basis of UCLA-20 [28]. This scale contains 8 items and the total score ranges from 8 to 32, with higher scores indicating higher levels of loneliness. The loneliness of participants was classified into three levels (low, intermediate, and high) according to the tertile of total score of ULS-8. The Cronbach α of this scale in this study was 0.818.

Subjective Well-Being Assessment

Subjective well-being (SWB), a main indicator of mental health, is a broad category of phenomena that includes people's emotional response, domain satisfaction, and global judgments of life satisfaction [29]. SWB was measured using the Chinese adapted version of the General Wellbeing Schedule (GWS) [30], which was first developed by the US National Center for Health Statistics in 1977. It contains the following 6 dimensions: satisfaction and interest in life; health concerns; energy; depression or pleasant mood; control of emotions and behavior; and relaxation and tension. The total scores for this instrument are yielded by summing up the items' scores, with higher scores indicating higher levels of SWB. The SWB of participants was classified into three levels (high, intermediate, and low) according to the tertile of total scores on the GWS. The GWS has high reliability and validity in the Chinese population, with a Cronbach of 0.91 and 0.95 for men and women, respectively [30]. The Cronbach α of this scale in this study was 0.812.

Statistical Analysis

Categorical variables were summarized as counts and percentages, while continuous variables were summarized as mean and standard deviation. The statistical difference in the distribution of the GWS scores across different sociodemographic characteristics and lifestyle behaviors was assessed using the Student *t* test or the one-way analysis of variance (ANOVA). The statistical difference in the distribution of perceived lifestyle changes across different groups was assessed using the chi-square test. Multivariate ordinal regression models were used to explore the association between SWB (1=low; 2=intermediate; 3=high) and lifestyle behaviors as well as perceived lifestyle changes. The strength of association was reported as odds ratio (OR) and 95% CI. An unadjusted relationship between SWB and each lifestyle behavior as well as its change was examined in Model 1 (crude model), whereas an adjusted relationship between SWB and each lifestyle behavior as well as its change was examined in Model 2, adjusted for the variables age, gender, education, marital status, residential location, and personal monthly income. Finally, a relationship between SWB and each lifestyle behavior as well as its change was examined in Model 3, which adjusted for covariates including self-rated physical health, perceived social support, loneliness, and all the covariates entered in Model 2. The statistical analyses were conducted in SPSS (Version 22.0; IBM Corp). All statistical tests were two-tailed and a $P<.05$ was considered statistically significant.

Results

A total of 1600 registrants were invited to take part in this online survey and a total of 1081 subjects consented to take part in the study, providing a response rate of 67.6%. The main reasons for nonresponse were declining to participate and no reply. After excluding 48 participants due to an incomplete questionnaire or missing data, a total of 1033 participants were included in this study. Among them, 637 participants were aged between 18 and 30 years (61.7%) and 498 were female (48.2%). Further, a very small proportion of the participants (n=95, 9.2%) had attended high school and lower levels of education. In addition, 730 subjects (70.7%) were living in urban areas, 846 participants (81.9%) considered their physical health status as good, and 675 participants (65.3%) felt that they had been receiving high social support. Regarding lifestyle, 58.8% (n=607) of the participants had leisure-time physical exercise of more than 150 min/week, about 86.8% (n=897) had screen time of 2 hours/day or more, and 67.9% (n=701) ate vegetables 5 times/week or more, whereas 41.8% (n=432) of the subjects ate fruits 5 times/week or more. Characteristics of participants are summarized in Table 1.

The mean GWS score in this sample was 71.7 points (SD 12.5). In addition, the distribution of GWS scores was not significantly different across the categories of gender, education, and leisure-time screen time ($P>.05$). Table 1 shows the details.

Table 1. Distribution of sample characteristics and General Wellbeing Schedule scores.

Characteristics	Participants, n (%)	General Wellbeing Schedule score, mean (SD)	*P* value[a]
Age (years)			
18-30	637 (61.7)	70.9 (12.1)	.03
31-40	281 (27.2)	72.3 (12.6)	N/A[b]
≥41	115 (11.1)	74.1 (14.3)	N/A
Gender			
Male	535 (51.8)	71.8 (12.2)	.68
Female	498 (48.2)	71.5 (12.9)	N/A
Residential location			
Urban	730 (70.7)	72.6 (12.5)	<.001
Rural	303 (29.3)	69.5 (12.4)	N/A
Education			
High school and below	95 (9.2)	69.5 (12.2)	.06
College or university degree	858 (83.1)	71.7 (12.5)	N/A
Master's degree and above	80 (7.7)	74.1 (12.5)	N/A
Marital status			
Married	485 (47.0)	73.2 (12.4)	<.001
Unmarried	548 (53.0)	69.9 (12.5)	N/A
Personal monthly income (¥)			
<3000	283 (27.4)	69.2 (12.4)	<.001
3000-7999	471 (45.6)	71.2 (12.4)	N/A
≥8000	279 (27.0)	74.9 (12.3)	N/A
Self-rated physical health			
Good	846 (81.9)	73.6 (11.8)	<.001
Fair	177 (17.1)	63.4 (12.0)	N/A
Poor	10 (1.0)	54.7 (11.6)	N/A
Perceived social support			
High	675 (65.3)	75.7 (12.5)	<.001
Intermediate	340 (32.9)	64.8 (11.3)	N/A
Low	18 (1.8)	52.3 (16.3)	N/A
Loneliness			
High	340 (32.9)	81.0 (9.3)	<.001
Intermediate	387 (37.5)	71.1 (10.1)	N/A
Low	306 (29.6)	62.0 (10.7)	N/A
Leisure-time physical exercise			
Active	607 (58.8)	72.9 (12.4)	<.001
Inactive	426 (41.2)	69.9 (12.6)	N/A
Leisure-time screen time			
Short	136 (13.2)	72.1 (14.0)	.64
Long	897 (86.8)	71.6 (12.3)	N/A
Frequency of vegetable intake			
Infrequently	332 (32.1)	69.2 (12.1)	<.001

Characteristics	Participants, n (%)	General Wellbeing Schedule score, mean (SD)	*P* value[a]
Frequently	701 (67.9)	72.8 (12.6)	N/A
Frequency of fruit intake			
Infrequently	601 (58.2)	69.8 (12.0)	<.001
Frequently	432 (41.8)	74.2 (12.8)	N/A
Frequency of skipping breakfast			
Seldom	721 (69.8)	73.4 (12.5)	<.001
Often	312 (30.2)	67.7 (11.7)	N/A

[a]*P* values were obtained according to the Student *t* test or one-way analysis of variance.

[b]N/A: not applicable.

Furthermore, about 17.0% of the respondents stated that they spent more time doing physical exercise, and about 2/3 of the respondents reported that they spent more time looking at screens. Additionally, a small proportion of participants decreased the frequency of their intake of vegetables, fruits, and breakfast. In addition, changes in the frequency of fruit and breakfast intake were associated with participants' residential locations. There were no statistical differences in perceived lifestyle changes associated with gender. These results are shown in Table 2.

Table 2. Perceived lifestyle changes by gender and place of residence.

Lifestyle habits	Overall, %	Gender			Place of residence		
		Male, %	Female, %	*P* value[a]	Urban, %	Rural, %	*P* value[a]
Time spent exercising							
Increased or much increased	17.8	18.5	17.1	.12	18.6	19.5	.36
Same as before	63.3	65.1	61.4	N/A[b]	62.5	65.3	N/A
Decreased or much decreased	18.9	16.4	21.5	N/A	18.9	15.2	N/A
Time spent looking at screens							
Increased or much increased	68.3	65.2	71.5	.09	70.0	64.0	.12
Same as before	23.9	25.8	21.9	N/A	23.0	26.1	N/A
Decreased or much decreased	7.8	9.0	6.6	N/A	7.0	9.9	N/A
Frequency of vegetable intake							
Increased or much increased	28.8	28.4	29.1	.32	29.2	27.7	.89
Same as before	55.8	54.6	57.2	N/A	55.5	56.8	N/A
Decreased or much decreased	15.4	17.0	13.7	N/A	15.3	15.5	N/A
Frequency of fruit intake							
Increased or much increased	35.3	36.7	33.9	.21	37.1	31.0	.004
Same as before	46.0	46.7	45.2	N/A	46.7	44.2	N/A
Decreased or much decreased	18.7	16.6	20.9	N/A	16.2	24.8	N/A
Frequency of skipping breakfast							
Increased or much increased	23.1	20.2	26.1	.06	23.4	22.1	.02
Same as before	56.9	58.1	55.6	N/A	58.8	52.5	N/A
Decreased or much decreased	20.0	21.7	18.3	N/A	17.8	25.4	N/A

[a]*P* value was obtained by chi-square test.

[b]N/A: not applicable.

The association between lifestyle behaviors during the COVID-19 pandemic and subjective well-being among the general population in Mainland China is summarized in Table 3. The multivariate ordinal regression model showed that participants with inadequate leisure-time physical exercise (OR 1.16, 95% CI 1.02-1.48), infrequent vegetable intake (OR 1.45, 95% CI 1.10-1.90), infrequent fruit intake (OR 1.31, 95% CI 1.01-1.70), as well as those who often skipped breakfast (OR 1.43, 95% CI 1.08-1.91) were associated with a higher risk of lower subjective well-being after adjusting for age, gender,

education, marital status, residential location, personal monthly income, self-rated physical health, perceived social support, and loneliness.

The association between perceived lifestyle changes, before and after the outbreak of COVID-19, and subjective well-being is summarized in Table 4. An ordinal regression model indicated that participants with decreased frequency of vegetable and fruit intake, and increased frequency of skipping breakfast were more likely to report lower subjective well-being after adjusting for sociodemographic factors and other covariates.

Table 3. The association between lifestyle behaviors and subjective well-being.

Lifestyle behaviors	Model 1, OR[a] (95% CI)	Model 2[b], OR (95% CI)	Model 3[c], OR (95% CI)
Leisure-time physical exercise			
Active	1.00	1.00	1.00
Inactive	1.51 (1.20-1.89)	1.42 (1.12-1.79)	1.16 (1.02-1.48)
Leisure-time screen time			
Short	1.00	1.00	1.00
Long	1.12 (0.81-1.57)	1.09 (0.78-1.53)	1.37 (0.94-1.99)
Frequency of vegetable intake			
Frequently	1.00	1.00	1.00
Infrequently	1.82 (1.41-2.30)	1.81 (1.41-2.31)	1.45 (1.10-1.90)
Frequency of fruit intake			
Frequently	1.00	1.00	1.00
Infrequently	1.88 (1.49-2.37)	1.70 (1.34-2.14)	1.31 (1.01-1.70)
Frequency of skipping breakfast			
Seldom	1.00	1.00	1.00
Often	2.28 (1.78-2.92)	2.12 (1.64-2.73)	1.43 (1.08-1.91)

[a]OR: odds ratio.

[b]Adjusted for age, gender, marital status, residential location, education, and personal monthly income.

[c]Adjusted for covariates in Model 2 and plus self-rated physical health, perceived social support, and loneliness.

Table 4. The association between perceived lifestyle changes and subjective well-being.

Perceived lifestyle changes	Model 1, OR[a] (95% CI)	Model 2[b], OR (95% CI)	Model 3[c], OR (95% CI)
Decreased time spent exercising	1.15 (0.85-1.54)	1.14 (0.86-1.56)	1.11 (0.80-1.54)
Increased time spent looking at screens	1.33 (1.05-1.69)	1.32 (0.98-1.60)	1.03 (0.80-1.34)
Decreased frequency of vegetable intake	1.84 (1.34-2.53)	1.78 (1.30-2.44)	1.73 (1.21-2.46)
Decreased frequency of fruit intake	1.56 (1.16-2.08)	1.48 (1.10-1.98)	1.41 (1.02-1.96)
Increased frequency of skipping breakfast	2.05 (1.45-2.88)	1.83 (1.29-2.59)	1.49 (1.01-2.18)

[a]OR: odds ratio.

[b]Adjusted for age, gender, marital status, residential location, education, and personal monthly income.

[c]Adjusted for covariates in Model 2 and plus self-rated physical health, perceived social support, and loneliness.

Discussion

Previous studies have predominantly focused on the psychological impact of the COVID-19 epidemic, rather than lifestyle issues. For the first time, some perceived lifestyle changes after the outbreak of COVID-19 have been assessed, and the impact of such changes on mental health was also explored among the general population in Mainland China. Noticeably, NPIs have modified some lifestyle behaviors positively and others negatively. Both unhealthy lifestyle behaviors and negative lifestyle changes were associated with lower SWB. Although about half of the participants reported no lifestyle changes, the percentages of reported favorable lifestyle changes were larger than the percentages of reported unfavorable lifestyle changes, especially in relation to the frequency of vegetable and fruit intake. However, the situation was the opposite when considering leisure-time physical exercise and screen time. Thus, it is possible that the sudden occurrence of COVID-19 made people reconsider their healthy lifestyle habits. In addition, the social or home isolation policy made people avoid public places and increase their indoor time, which may have increased their use of electronic media at home.

Cognitive behavior therapy enables activity scheduling during home isolation and improves mental health [31].

Accordingly, unhealthy lifestyle behaviors were recorded among the Chinese population after the outbreak of COVID-19. For example, about 40% of the participants had inactive leisure-time physical exercise and about 90% had longer screen time. In addition, vegetable and fruit intake was less than 5 times/week for about 30% and 60% of the participants, respectively. However, it is not known whether such lifestyle patterns would persist during the COVID-19 pandemic or after the COVID-19 pandemic. Therefore, further studies to assess the lasting effects of the COVID-19 pandemic on lifestyle behaviors are warranted.

The findings of this study have added to the existing evidence that physical exercise is associated with mental well-being [32,33]. For example, a larger cross-sectional study indicated that sports and vigorous recreational activities were positively associated with emotional well-being even after adjusting for sex, social class, and health status [34]. In addition, active physical exercises were associated with reduced risk of mental health conditions such as depression and anxiety [35]. Similarly, during the SARS epidemic, increased exercise time was associated with decreased perceived stress and incidence of PTSD in the general population of Hong Kong [36]. However, increased exercise time did not significantly predict subjective well-being. A possible explanation could be that SWB may be associated with the intensity of physical exercise. Many studies demonstrated that vigorous physical exercises were positively associated with SWB, while moderate physical exercises were either not associated or negatively associated with SWB [37,38]. Another reason is that the environment for physical exercise may have a great impact on mental well-being [39]. Therefore, further studies about intensity of physical exercises under different environments are needed to fully elucidate the effects on wellbeing.

The finding that a higher frequency of vegetable and fruit intake was positively associated with subjective well-being is consistent with previous studies [40-42]. There is an urgent need to disseminate this health information to the public during the COVID-19 pandemic via the internet and health collaborators [43,44]. For example, Stranges et al [45] reported that the odds for low mental well-being were increased in those with decreasing fruit and vegetable intake. Furthermore, a systematic review of 61 studies indicated that higher total intake of fruits and vegetables may promote higher levels of optimism and reduce psychological distress, thereby having a positive influence on mental health [42]. Even though the total quantities of vegetable and fruit intake were not obtained in this study, we assumed that a higher frequency of intake may lead to higher total volume intake; thus, the former and the latter may have the same influence. Nevertheless, further studies are warranted to understand the association between mental health and quantities of vegetables and fruits consumed. The finding about the relationship between eating habits, like skipping breakfast, and well-being added to the existing evidence that skipping breakfast is associated with poorer physical and mental health outcomes [41,46,47]. For instance, breakfast skippers had significantly worse health-related quality of life both physically and mentally in a Taiwanese national representative sample [48]. In addition, a large cohort study indicated that an eating pattern characterized by skipping or delaying breakfast was associated with mood disorders among Australian adults [49]. However, the relationship between mental health and the manner in which breakfast is eaten needs further investigation.

In summary, this original preliminary study examined some positive and negative lifestyle changes due to the influence of the COVID-19 pandemic. It has revealed the pressing need to provide individuals, communities, and health agencies with information to help maintain healthy lifestyles to some degree while in isolation. Moreover, this study has contributed to the scant literature on the association between lifestyle behaviors and mental health during the COVID-19 pandemic. However, this study has several limitations. First, this study had a cross-sectional design and was conducted during the COVID-19 pandemic. Thus, causality relationships could not be inferred, and it is not clear whether the association between subjective well-being and lifestyle behaviors as well as their changes will last medium- and long-term. Second, precise information about lifestyle behaviors was difficult to collect before the start of the COVID-19 pandemic. Therefore, only perceived lifestyle changes were measured. Moreover, lifestyle behaviors and perceived changes were self-reported, thus these measurements may be susceptible to recall bias. Third, this study population had a high proportion of subjects aged 18 to 40 years (approximately 90%) and a high proportion of participants had a higher education level. Whether the observed changes and associations are present in a more representative study population requires further exploration. Therefore, longitudinal studies with representative samples should be conducted during the COVID-19 pandemic to better understand the lasting effects of this pandemic on lifestyle behaviors and their changes.

Acknowledgments

We thank all the participants for their cooperation throughout the study. There was no funding for this study.

Authors' Contributions

ZH and XL contributed to analyzing data and writing the first original draft. ACK contributed to the editing of all versions of the manuscript. HX contributed to the editing of the first draft and provided supervision. All authors have read and agreed to the final draft of the manuscript.

Conflicts of Interest

None declared.

References

1. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Feb 26;76:71-76. [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
2. Paules CI, Marston HD, Fauci AS. Coronavirus Infections-More Than Just the Common Cold. JAMA 2020 Jan 23:1. [doi: 10.1001/jama.2020.0757] [Medline: 31971553]
3. Mahase E. China coronavirus: WHO declares international emergency as death toll exceeds 200. BMJ 2020 Jan 31;368:m408. [doi: 10.1136/bmj.m408] [Medline: 32005727]
4. Coronavirus disease (COVID-19) Pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-08-19]
5. Balanzá-Martínez V, Atienza-Carbonell B, Kapczinski F, De Boni RB. Lifestyle behaviours during the COVID-19 - time to connect. Acta Psychiatr Scand 2020 May;141(5):399-400 [FREE Full text] [doi: 10.1111/acps.13177] [Medline: 32324252]
6. Velten J, Bieda A, Scholten S, Wannemüller A, Margraf J. Lifestyle choices and mental health: a longitudinal survey with German and Chinese students. BMC Public Health 2018 May 16;18(1):632 [FREE Full text] [doi: 10.1186/s12889-018-5526-2] [Medline: 29769115]
7. Zaman R, Hankir A, Jemni M. Lifestyle Factors and Mental Health. Psychiatr Danub 2019 Sep;31(Suppl 3):217-220 [FREE Full text] [Medline: 31488729]
8. Shah K, Kamrai D, Mekala H, Mann B, Desai K, Patel RS. Focus on Mental Health During the Coronavirus (COVID-19) Pandemic: Applying Learnings from the Past Outbreaks. Cureus 2020 Mar 25;12(3):e7405 [FREE Full text] [doi: 10.7759/cureus.7405] [Medline: 32337131]
9. Wang C, Pan R, Wan X, Tan Y, Xu L, McIntyre RS, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. Brain Behav Immun 2020 Jul;87:40-48 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.028] [Medline: 32298802]
10. Tan W, Hao F, McIntyre RS, Jiang L, Jiang X, Zhang L, et al. Is returning to work during the COVID-19 pandemic stressful? A study on immediate mental health status and psychoneuroimmunity prevention measures of Chinese workforce. Brain Behav Immun 2020 Jul;87:84-92 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.055] [Medline: 32335200]
11. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
12. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. Asian J Psychiatr 2020 Apr 10;52:102066 [FREE Full text] [doi: 10.1016/j.ajp.2020.102066] [Medline: 32302935]
13. Torales J, O'Higgins M, Castaldelli-Maia JM, Ventriglio A. The outbreak of COVID-19 coronavirus and its impact on global mental health. Int J Soc Psychiatry 2020 Jun;66(4):317-320. [doi: 10.1177/0020764020915212] [Medline: 32233719]
14. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920. [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
15. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1923-1994 [FREE Full text] [doi: 10.1016/S0140-6736(18)32225-6] [Medline: 30496105]
16. Rao S, Shah N, Jawed N, Inam S, Shafique K. Nutritional and lifestyle risk behaviors and their association with mental health and violence among Pakistani adolescents: results from the National Survey of 4583 individuals. BMC Public Health 2015 Apr 28;15:431 [FREE Full text] [doi: 10.1186/s12889-015-1762-x] [Medline: 25927931]
17. Chen P, Mao L, Nassis GP, Harmer P, Ainsworth BE, Li F. Coronavirus disease (COVID-19): The need to maintain regular physical activity while taking precautions. J Sport Health Sci 2020 Mar;9(2):103-104 [FREE Full text] [doi: 10.1016/j.jshs.2020.02.001] [Medline: 32099716]
18. Zhang L, Liu Y. Potential interventions for novel coronavirus in China: A systematic review. J Med Virol 2020 May;92(5):479-490 [FREE Full text] [doi: 10.1002/jmv.25707] [Medline: 32052466]
19. Dong L, Bouey J. Public Mental Health Crisis during COVID-19 Pandemic, China. Emerg Infect Dis 2020 Jul;26(7):1616-1618 [FREE Full text] [doi: 10.3201/eid2607.200407] [Medline: 32202993]
20. Galea S, Merchant RM, Lurie N. The Mental Health Consequences of COVID-19 and Physical Distancing: The Need for Prevention and Early Intervention. JAMA Intern Med 2020 Apr 10:2020. [doi: 10.1001/jamainternmed.2020.1562] [Medline: 32275292]

21. Tran BX, Ha GH, Nguyen LH, Vu GT, Hoang MT, Le HT, et al. Studies of Novel Coronavirus Disease 19 (COVID-19) Pandemic: A Global Analysis of Literature. Int J Environ Res Public Health 2020 Jun 08;17(11):1-16 [FREE Full text] [doi: 10.3390/ijerph17114095] [Medline: 32521776]
22. Wjx. URL: https://www.wjx.cn [accessed 2020-08-19]
23. Ren YJ, Su M, Liu QM, Tan YY, Li LM, Lyu J. Validation of the Simplified Chinese-character Version of the International Physical Activity Questionnaire-Long Form in Urban Community-dwelling Adults: a Cross-sectional Study in Hangzhou, China. Biomed Environ Sci 2017 Apr;30(4):255-263 [FREE Full text] [doi: 10.3967/bes2017.035] [Medline: 28494835]
24. Haskell WL, Lee I, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007 Aug;39(8):1423-1434. [doi: 10.1249/mss.0b013e3180616b27] [Medline: 17762377]
25. Zhao W, Hasegawa K, Chen J. The use of food-frequency questionnaires for various purposes in China. Public Health Nutr 2002 Dec;5(6A):829-833. [doi: 10.1079/phn2002374] [Medline: 12638592]
26. Zimet GD, Powell SS, Farley GK, Werkman S, Berkoff KA. Psychometric characteristics of the Multidimensional Scale of Perceived Social Support. J Pers Assess 1990;55(3-4):610-617. [doi: 10.1080/00223891.1990.9674095] [Medline: 2280326]
27. Xu S, Qiu D, Hahne J, Zhao M, Hu M. Psychometric properties of the short-form UCLA Loneliness Scale (ULS-8) among Chinese adolescents. Medicine (Baltimore) 2018 Sep;97(38):e12373 [FREE Full text] [doi: 10.1097/MD.0000000000012373] [Medline: 30235699]
28. Hays RD, DiMatteo MR. A short-form measure of loneliness. J Pers Assess 1987;51(1):69-81. [doi: 10.1207/s15327752jpa5101_6] [Medline: 3572711]
29. Diener E, Suh EM, Lucas RE, Smith HL. Subjective well-being: Three decades of progress. Psychological Bulletin 1999;125(2):276-302. [doi: 10.1037/0033-2909.125.2.276]
30. Duan J. The application of General Well-Being Schedule in university students [Chinese]. Chinese Journal of Clinical Psychology 1996;4(1):56-57.
31. Ho CS, Chee CY, Ho RC. Mental Health Strategies to Combat the Psychological Impact of COVID-19 Beyond Paranoia and Panic. Ann Acad Med Singapore 2020 Mar 16;49(3):155-160 [FREE Full text] [Medline: 32200399]
32. McMahon EM, Corcoran P, O'Regan G, Keeley H, Cannon M, Carli V, et al. Physical activity in European adolescents and associations with anxiety, depression and well-being. Eur Child Adolesc Psychiatry 2017 Jan;26(1):111-122. [doi: 10.1007/s00787-016-0875-9] [Medline: 27277894]
33. Haapasalo V, de Vries H, Vandelanotte C, Rosenkranz RR, Duncan MJ. Cross-sectional associations between multiple lifestyle behaviours and excellent well-being in Australian adults. Prev Med 2018 Nov;116:119-125. [doi: 10.1016/j.ypmed.2018.09.003] [Medline: 30218725]
34. Steptoe A, Butler N. Sports participation and emotional wellbeing in adolescents. Lancet 1996 Jun 29;347(9018):1789-1792. [doi: 10.1016/s0140-6736(96)91616-5] [Medline: 8667922]
35. Ströhle A. Physical activity, exercise, depression and anxiety disorders. J Neural Transm (Vienna) 2009 Jun;116(6):777-784. [doi: 10.1007/s00702-008-0092-x] [Medline: 18726137]
36. Lau JTF, Yang X, Tsui HY, Pang E, Wing YK. Positive mental health-related impacts of the SARS epidemic on the general public in Hong Kong and their associations with other negative impacts. J Infect 2006 Aug;53(2):114-124 [FREE Full text] [doi: 10.1016/j.jinf.2005.10.019] [Medline: 16343636]
37. Wicker P, Frick B. Intensity of physical activity and subjective well-being: an empirical analysis of the WHO recommendations. J Public Health (Oxf) 2017 Jun 01;39(2):e19-e26. [doi: 10.1093/pubmed/fdw062] [Medline: 27412174]
38. Costigan SA, Lubans DR, Lonsdale C, Sanders T, Del Pozo Cruz B. Associations between physical activity intensity and well-being in adolescents. Prev Med 2019 Aug;125:55-61. [doi: 10.1016/j.ypmed.2019.05.009] [Medline: 31125627]
39. Thompson Coon J, Boddy K, Stein K, Whear R, Barton J, Depledge MH. Does participating in physical activity in outdoor natural environments have a greater effect on physical and mental wellbeing than physical activity indoors? A systematic review. Environ Sci Technol 2011 Mar 01;45(5):1761-1772. [doi: 10.1021/es102947t] [Medline: 21291246]
40. Ocean N, Howley P, Ensor J. Lettuce be happy: A longitudinal UK study on the relationship between fruit and vegetable consumption and well-being. Soc Sci Med 2019 Feb;222:335-345 [FREE Full text] [doi: 10.1016/j.socscimed.2018.12.017] [Medline: 30626498]
41. Lesani A, Mohammadpoorasl A, Javadi M, Esfeh JM, Fakhari A. Eating breakfast, fruit and vegetable intake and their relation with happiness in college students. Eat Weight Disord 2016 Dec;21(4):645-651. [doi: 10.1007/s40519-016-0261-0] [Medline: 26928281]
42. Głąbska D, Guzek D, Groele B, Gutkowska K. Fruit and Vegetable Intake and Mental Health in Adults: A Systematic Review. Nutrients 2020 Jan 01;12(1):1-34 [FREE Full text] [doi: 10.3390/nu12010115] [Medline: 31906271]
43. Tran BX, Dang AK, Thai PK, Le HT, Le XTT, Do TTT, et al. Coverage of Health Information by Different Sources in Communities: Implication for COVID-19 Epidemic Response. Int J Environ Res Public Health 2020 May 20;17(10):3577 [FREE Full text] [doi: 10.3390/ijerph17103577] [Medline: 32443712]

44. Tran BX, Phan HT, Nguyen TPT, Hoang MT, Vu GT, Thi Lei H, et al. Reaching further by Village Health Collaborators: The informal health taskforce of Vietnam for COVID-19 responses. J Glob Health 2020 Jun;10(1):010354 [FREE Full text] [doi: 10.7189/jogh.10.010354] [Medline: 32509285]
45. Stranges S, Samaraweera PC, Taggart F, Kandala N, Stewart-Brown S. Major health-related behaviours and mental well-being in the general population: the Health Survey for England. BMJ Open 2014 Sep 19;4(9):e005878 [FREE Full text] [doi: 10.1136/bmjopen-2014-005878] [Medline: 25239293]
46. Rong S, Snetselaar LG, Xu G, Sun Y, Liu B, Wallace RB, et al. Association of Skipping Breakfast With Cardiovascular and All-Cause Mortality. J Am Coll Cardiol 2019 Apr 30;73(16):2025-2032 [FREE Full text] [doi: 10.1016/j.jacc.2019.01.065] [Medline: 31023424]
47. Park S, Rim SJ, Lee JH. Associations between dietary behaviours and perceived physical and mental health status among Korean adolescents. Nutr Diet 2018 Nov;75(5):488-493. [doi: 10.1111/1747-0080.12444] [Medline: 29978549]
48. Huang C, Hu H, Fan Y, Liao Y, Tsai P. Associations of breakfast skipping with obesity and health-related quality of life: evidence from a national survey in Taiwan. Int J Obes (Lond) 2010 Apr;34(4):720-725. [doi: 10.1038/ijo.2009.285] [Medline: 20065977]
49. Wilson JE, Blizzard L, Gall SL, Magnussen CG, Oddy WH, Dwyer T, et al. An eating pattern characterised by skipped or delayed breakfast is associated with mood disorders among an Australian adult cohort. Psychol Med 2019 Oct 16:1-11. [doi: 10.1017/S0033291719002800] [Medline: 31615586]

Abbreviations

COVID-19: coronavirus disease
FFQ-25: Food Frequency Questionnaire
GWS: General Wellbeing Schedule
NPI: nonpharmacological intervention
OR: odds ratio
PSS: perceived social support
PSSS: Perceived Social Support Scale
PTSD: posttraumatic stress disorder
SARS: severe acute respiratory syndrome
SWB: subjective well-being
ULS-8: University of California, Los Angeles Loneliness Scale
WHO: World Health Organization

Edited by G Eysenbach; submitted 07.06.20; peer-reviewed by CF Yen, R Ho; comments to author 13.07.20; revised version received 26.07.20; accepted 02.08.20; published 25.08.20.

Please cite as:
Hu Z, Lin X, Chiwanda Kaminga A, Xu H
Impact of the COVID-19 Epidemic on Lifestyle Behaviors and Their Association With Subjective Well-Being Among the General Population in Mainland China: Cross-Sectional Study
J Med Internet Res 2020;22(8):e21176
URL: http://www.jmir.org/2020/8/e21176/
doi:10.2196/21176
PMID:32759103

Original Paper

Features and Functionalities of Smartphone Apps Related to COVID-19: Systematic Search in App Stores and Content Analysis

Roberto Collado-Borrell[1], PharmD, PhD; Vicente Escudero-Vilaplana[1], PharmD, PhD; Cristina Villanueva-Bueno[1], PharmD; Ana Herranz-Alonso[1], PharmD, PhD; Maria Sanjurjo-Saez[1], PharmD, PhD

Hospital General Universitario Gregorio Marañón, Madrid, Spain

Corresponding Author:
Vicente Escudero-Vilaplana, PharmD, PhD
Hospital General Universitario Gregorio Marañón
C/del Dr. Esquerdo, 46
Madrid, 28007
Spain
Phone: 34 678239017
Email: vicente.escudero@salud.madrid.org

Abstract

Background: Knowledge of the quantity and quality of apps related to coronavirus disease (COVID-19) is lacking. In addition, no directory has been established listing all the apps developed to address the COVID-19 pandemic.

Objective: The aim of this study was to identify smartphone apps designed to address the COVID-19 pandemic and to analyze their characteristics.

Methods: We performed an observational, cross-sectional, descriptive study of all smartphone apps associated with COVID-19. Between April 27 and May 2, 2020, we searched the App Store (iOS) and Google Play Store (Android) for COVID-19 apps. The search terms used were *coronavirus*, *COVID-19*, and *SARS-COV-2*. The apps were downloaded and evaluated. The variables analyzed were name, platform, country, language, category, cost, update date, size, version, number of downloads, developer, and purpose. Purpose was further classified into the following categories: news, general information, self-diagnosis, contact tracing, notices to contacts, notification of close cases, awareness, helplines, monitoring of clinical parameters, recording of symptoms and treatment, and messaging with health care professionals.

Results: We identified 114 apps on the investigated platforms. Of these, 62/114 (54.4%) were on Android and 52/114 (45.6%) were on iOS. Of the 114 apps, 37 (32.5%) were developed in Europe, 32 (28.1%) in Asia, and 30 (26.3%) in North America. The most frequent languages were English (65/114, 57.0%), Spanish (34/114, 29.8%), and Chinese (14/114, 12.3%). The most common categories were health and well-being/fitness apps (41/114, 41.2%) and medicine apps (43/114, 37.7%). Of the 114 apps, 113 (99.1%) were free. The mean time between the date of the analysis and the date of the last update was 11.1 days (SD 11.0). Overall, 95 of the 114 apps (83.3%) were intended for the general population, 99 apps (7.9%) were intended for health professionals, and 3 apps (2.6%) were intended for both. Regarding the type of developer, 64/114 apps (56.1%) were developed by governments; 42/114 (64.1%) were developed by national governments, and 23/114 (35.9%) were developed by regional governments. The apps with the highest number of downloads (100,000+) were developed by governments (P=.13), except for the World Health Organization app (500,000+). The purposes of the apps available in Western languages (107/114, 93.9%) were determined; the most common purposes were general information about COVID-19 (66, 64.0%), COVID-19 news (53, 51.0%), recording of symptoms (53, 51.0%), and contact tracing (51, 47.7%). More than one purpose was identified for 99/107 apps (92.5%).

Conclusions: This paper offers a comprehensive and unique review of all available COVID-19 apps. Governments have adopted these tools during the pandemic, and more than half of the apps were developed by government agencies. The most common purposes of the apps are providing information on the numbers of infected, recovered, and deceased patients, recording of symptoms, and contact tracing.

(J Med Internet Res 2020;22(8):e20334) doi:10.2196/20334

KEYWORDS

COVID-19; mobile apps; contact tracing; monitoring; telemedicine; smartphone

Introduction

The novel coronavirus SARS-CoV-2, which causes coronavirus disease (COVID-19), has created a health care crisis that is unprecedented in our recent history. Since the first case was detected in late 2019 [1], more than 4 million infections and more than 300,000 deaths have been confirmed worldwide [2]. These data, together with the high capacity for progression of COVID-19 and the lack of knowledge about its behavior patterns, highlight the urgent need to find a solution to this health emergency [2].

Governments, companies, and several public and private institutions worldwide are combining their efforts to find an effective solution to reduce the risk of spreading COVID-19 [3]. In this scenario, digital technologies constitute essential tools for improving the health of populations as well as basic services provided to them [4,5]. The World Health Organization (WHO) recently published 10 recommendations for improving population health and essential services through digital technologies [6].

It is currently estimated that more than 5 billion people use mobile phones [7]; moreover, according to the "State of Mobile in 2019" report published in 2018 [8], 194 billion apps have been downloaded worldwide. Apps are therefore easily accessed and implemented by the vast majority of the world's population [9]. Apps have various beneficial functions related to the COVID-19 pandemic, from remote follow-up by health care professionals to monitoring the number of infections. One method of monitoring the number of infections is contact tracing [10-13]. According to Yasaka et al [14], smartphone-based contact tracing presents a novel solution that preserves privacy and has the potential to suppress an epidemic or pandemic. Another scenario in which apps may be useful is self-diagnosis, in which an app can process a patient's information and provide an approximate diagnosis [4]. Taking these data into account, apps can be useful and effective tools for managing the COVID-19 pandemic.

The success of an app depends on the context in which it is used and the appropriateness of its design. However, the current lack of knowledge about the quantity and quality of apps related to COVID-19 is compounded by the lack of directories listing the apps that have been developed to address the pandemic.

The objective of this study was to identify smartphone apps designed to address the COVID-19 pandemic and to analyze their characteristics.

Methods

We performed an observational, cross-sectional, descriptive study of all smartphone apps associated with COVID-19 available on the iOS and Android platforms.

The methodology used for the selection of the apps was based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) system. Between April 27 and May 2, 2020, we searched the App Store (iOS) and the Google Play Store (Android) for all COVID-19–related apps. The search terms were *coronavirus*, *COVID-19*, and *SARS-COV-2*, and no language restrictions were applied. After identification, the information available on each platform was analyzed, and all apps were downloaded and evaluated. The apps included in the App Store were downloaded onto an iPhone 8 (iOS version 12.3.2), and the apps found in the Google Play Store were downloaded onto a Xiaomi Mi A1 (Android version 9.0).

The variables analyzed were the app name, platform (Android or iOS), country of origin, language, category, cost, date of last update, size, version, number of downloads, developer, and purpose. Purpose was further classified into the following categories: news, general information, self-diagnosis, contact tracing, alerts to contacts, notification of closed cases, awareness, helplines, monitoring of clinical parameters (temperature, blood pressure, heart rate, respiratory frequency, oxygen saturation), symptom log, treatment log, and messaging with health care professionals. "Contact tracing" was defined as the ability of the app to identify the locations of COVID-19 cases and thus to control the spread of the virus.

The app was evaluated by two independent researchers with experience in app analysis, design, and development (RC and VE). Continuous variables were compared using the *t* test when the distribution was normal or the Mann-Whitney test when it was not. Categorical variables were compared using an uncorrected chi-square test or Fisher exact test. The Cohen κ test was performed to guarantee the reliability of the data analyzed by the two independent researchers. Data were analyzed using Stata version IC-14 (descriptive statistics). A P value <.05 was considered statistically significant.

Results

App Characteristics

We identified 114 apps: 62/114 (54.4%) were on Android and 52/114 (45.6%) were on iOS. 50/114 apps (43.9%) were available on both platforms. Supplementary Table 1 (Multimedia Appendix 1) shows the names of the apps classified according to the country in which they were developed. Of the 114 apps, 37 (32.5%) were developed in Europe, 32 (28.1%) in Asia, 30 (26.3%) in North America, 12 (10.5%) in Central America–South America, and 3 (2.6%) in Africa. The most common languages were English (65/114 apps, 57.0%), Spanish (34 apps, 29.8%), and Chinese (14 apps, 12.3%). The average number of languages per app was 1.9 (range 1-12). Seven of the 114 analyzed apps (6.1%) were only available in Asian languages.

The most frequent categories of the 114 apps were health and well-being/fitness (47 apps, 41.2%), medicine (43 apps, 37.7%), utilities (5 apps, 4.4%), education (3 apps, 2.6%), and lifestyle (3 apps, 2.6%). A full list of the categories is shown in Table 1. Regarding cost, 113/144 apps (99.1%) were free. At least one version update was found for 96/114 apps (86.8%). Among the 96 apps that updated, 90 (93.8%) updated in April and 6 (6.3%) updated in March. The average time between the date of analysis and the date of the last update was 11.1 days (SD 11.0). The average size of the apps was 27.3 MB (SD 28.9). Regarding content classification, 85 of the 114 apps (74.6%)

were intended for all age groups, 12 (10.5%) for people older than 17 years, and 9 (7.9%) for people older than 12 years; 8 (7.0%) did not specify an age group. The target populations of the 114 apps were the general population (95 apps, 88.8%), health professionals (9 apps, 78.4%), and both (3 apps, 2.8%).

Regarding the type of developer, 64/114 apps (56.1%) were developed by governments. Of these, 42/114 (64.1%) were national governments, and 22 (35.9%) were regional. A full list of developers is shown in Table 2.

Table 1. Categories of the apps (N=114).

Category	n (%)
Health and well-being/fitness	47 (41.2)
Medicine	43 (37.7)
Utilities	5 (4.4)
Education	3 (2.6)
Lifestyle	3 (2.6)
Communication	2 (1.8)
Reference	2 (1.8)
News	2 (1.8)
Productivity	2 (1.8)
Parental control	1 (0.9)
Electronic health administration and medicine	1 (0.9)
Business	1 (0.9)
Economy and business	1 (0.9)
Travel and guides	1 (0.9)

Table 2. Developers of the apps (N=114).

Developer	n (%)
Government	64 (56.1)
Health-related technology company	19 (16.7)
Nonhealth-related technology company	11 (9.6)
Hospital	7 (6.1)
University	4 (3.5)
Independent health professionals	3 (2.6)
Not specified	2 (1.8)
World Health Organization	1 (0.9)
University or hospital	1 (0.9)
Voluntary humanitarian institution or public interest	1 (0.9)
Foundation	1 (0.9)

The numbers of downloads of the apps were determined where available (this information is provided on the Google Play store, but not the App Store). The median number of downloads of the apps was 5000 (range 1 to 50,000,000+). The apps with the most downloads (100,000+) were developed by governments (P=.13), except for the WHO app (500,000+). The app with the highest number of downloads (50,000,000+) was developed by the Indian government, followed by apps developed by the Polish and Colombian governments (1,000,000+). Supplementary Table 2 (Multimedia Appendix 1) shows the apps developed by various governments and the number of times they were downloaded.

Regarding purpose, we did not find a statistically significant relationship between the number of purposes and the number of downloads (P=.06). Analysis of the purpose of the 107 apps available in Western languages revealed that 99 apps (92.5%) had more than 1 purpose. The most frequent purposes were general information about the coronavirus pandemic (66/107 apps, 64%), news (53/107 apps, 51.0%), recording of symptoms (53/107 apps, 51.0%), and contact tracing (51/107 apps, 47.7%). The remaining purposes are shown in Table 3.

Table 3. Purposes of the apps (n=107).

Purpose	n (%)
General information about COVID-19[a]	66 (64.0)
News	53 (51.0)
Recording of symptoms	53 (51.0)
Contact tracing	51 (47.7)
Awareness	46 (43.0)
Helplines	46 (43.0)
Self-diagnosis	37 (34.6)
Warning of nearby cases	28 (26.2)
Contact with health professionals	22 (20.6)
Monitoring of clinical parameters	13 (12.1)
Recording treatments	6 (5.6)
Access control	4 (3.7)
Mobile contact warning	5 (4.7)

[a]COVID-19: coronavirus disease.

Apps developed for the purposes of contact tracing (19 apps, 18.4%), notifications of close cases (10 apps, 9.7%), access control (3 apps, 2.9%), and notice to contacts of the mobile (2 apps, 1.9%) were more common in Asia (*P*=.06). The purposes of general information (23 apps, 22.3%), news (17 apps, 16.5%), recording of symptoms (16 apps, 15.5%), and contact with health care professionals (8 apps, 7.8%) were more common among the apps developed in Europe. The most frequently observed functionalities of the apps developed in North America were general information (15 apps, 14.6%), and recording of symptoms (14 apps, 13.6%). In Central America–South America, the purposes of the apps were general information (10 apps, 9.3%) and news (10 apps, 9.3%) (*P*=.06). The Cohen κ coefficient was .96, indicating a degree of agreement of 96%.

Discussion

Principal Findings and Comparison With Prior Work

We provide a comprehensive and unique review of all available COVID-19–related apps. This is especially important because due to recently heightened security measures of iOS and Google with respect to COVID-19–related apps that do not come from official sources, there is currently no site where all available apps can be consulted [15]. In the Google Play Store, it is very difficult to find apps related to COVID-19 from countries other than that in which the user resides.

The present study, in which a total of 114 apps are analyzed, underlines the high number of apps designed by governments; this contrasts with other app reviews related to health, in which the leading developer is usually a scientific foundation or hospital [16,17]. Mobile health (mHealth) has evolved in recent years to position itself as a critical tool in patient communication and monitoring. However, the current pandemic has proven to be a turning point for mHealth technologies. Due to the rapid expansion of COVID-19 and its transmission by asymptomatic patients, detecting and isolating positive cases of COVID-19 has become a priority [3]. The ease of use of mobile phones and their accessibility by the majority of the world's population, added to the need for information from users, avoidance of travel and contact, and identification of cases, has positioned apps as critical tools in the management of the coronavirus pandemic. In this sense, the governments of the most affected countries have opted to implement app-based initiatives that collect user data, helping to increase knowledge about the prevention, detection, and monitoring of COVID-19 [3,5,18]. These apps have been designed in record time, taking into account that the average time to develop an app is estimated to be one year [19].

The purposes of the leading apps developed by governments include general information, news about the pandemic, and monitoring of the areas affected by the virus. The latter type of app can be used to collect specific sociodemographic information and trace contacts. Many apps are based on a GPS system through which users can determine whether they have traveled through places where cases of COVID-19 were previously detected, as well as the date on which the contagion was confirmed [13,14,18]. Contact tracing has been used successfully for other infections, such as tuberculosis and Ebola virus [4]. According to Yasaka *et al* [14], contact tracing is a viable solution to limit the transmission of disease. Because half of infections occur before the patient develops symptoms, a geolocation system can help users avoid critical areas and possible infection. These apps developed by governments were the most frequently downloaded. We did not find any relationship between the number of apps developed in a country or continent and the severity of the disease in that location. For example, apps have been developed by countries with early COVID-19 outbreaks (eg, Italy or Spain), later outbreaks (eg, Colombia or Argentina), or even countries where infection rates are relatively low (eg, Poland).

Data protection issues are a key problem with this type of app. While apps developed in Western countries collect most data

anonymously and in aggregate, some apps in Far Eastern countries collect greater amounts of personal data, enabling authorities to take more active measures to control the spread of the disease [11,12,18,20]. iOS and Google are currently working on a joint project to use people's mobility data to better monitor and predict COVID-19 infections while ensuring that users' privacy is respected [20].

Unlike Asia, in Europe and North America, the most common purposes of the apps are providing information on the numbers of persons infected, recovered, and deceased, enabling the recording of symptoms, and providing contact with health professionals. The number of apps related to symptom recording and contact with health care providers contrasts with data previously published by our group [16], which showed that only 6% of apps had this function. Recording of symptoms and communication between health care professionals and patients have been shown to improve access to health systems, reduce the use of health care resources, and improve monitoring and control of symptoms [16,19,21]. We also observed a high number of apps for self-diagnosis; the primary objective of these apps is to decongest health centers, thus enabling rapid and safe diagnosis of a larger number of people [14,22]. Digital/virtual triage apps enable the user to be evaluated for possible COVID-19 symptoms [4]. These systems are based on a series of questions related to the most frequent symptoms, such as cough, fever, dyspnea, and muscle pain. They also take into account whether an individual has been in contact with other people affected by COVID-19. Once the information has been processed, the app issues an approximate diagnosis. In the case of a positive result, the app itself shows a series of recommendations.

Medical follow-up apps have been designed in parallel for people who are self-isolating at home. According to Alwashmi et al [4], these apps have the potential to reduce transmission by minimizing physical contact between physician and patient. Through the recording and monitoring of the clinical parameters of COVID-19 (eg, temperature, oxygen saturation, and heart rate), together with virtual contact with health care professionals, the apps will enable people to adapt to the needs created by this situation, thus ensuring that patients receive medical attention.

Our Cohen κ analysis revealed an agreement of 96% between the two independent researchers, highlighting the robustness of our results.

Limitations

Our study is limited by its cross-sectional nature, which means that any information reported here could become outdated as the COVID-19 situation evolves. However, because there is currently no repository of all available COVID-19–related apps, this review ranks as the most current source.

Conclusions

Ease of access by the population and the use of artificial intelligence position apps as tools capable of detecting new foci of COVID-19 transmission, analyzing the rate of spread, monitoring possible symptoms, and approximately diagnosing positive cases at a distance. During the current pandemic, many governments have decided to implement apps to help curb the rapid expansion of COVID-19.

Acknowledgments

The authors would like to thank Tomas O'Boyle for their help preparing the manuscript. This project was supported by iPharma (Pharmacy Innovation Center at the Hospital General Universitario Gregorio Marañón).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 31 KB - jmir_v22i8e20334_app1.docx]

References

1. Danis K, Epaulard O, Bénet T, Gaymard A, Campoy S, Bothelo-Nevers E, Investigation Team. Cluster of coronavirus disease 2019 (Covid-19) in the French Alps, 2020. Clin Infect Dis 2020 Apr 11:online ahead of print [FREE Full text] [doi: 10.1093/cid/ciaa424] [Medline: 32277759]
2. Coronavirus disease (COVID-19) Situation Report – 109. World Health Organization. 2020 May 08. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200508covid-19-sitrep-109.pdf?sfvrsn=68f2c632_6 [accessed 2020-05-11]
3. Zamberg I, Manzano S, Posfay-Barbe K, Windisch O, Agoritsas T, Schiffer E. A Mobile Health Platform to Disseminate Validated Institutional Measurements During the COVID-19 Outbreak: Utilization-Focused Evaluation Study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18668 [FREE Full text] [doi: 10.2196/18668] [Medline: 32250958]
4. Alwashmi MF. The Use of Digital Health in the Detection and Management of COVID-19. Int J Environ Res Public Health 2020 Apr 23;17(8):2906 [FREE Full text] [doi: 10.3390/ijerph17082906] [Medline: 32340107]

5. Chan AT, Drew DA, Nguyen LH, Joshi AD, Ma W, Guo C, COPE Consortium. The COronavirus Pandemic Epidemiology (COPE) Consortium: A Call to Action. Cancer Epidemiol Biomarkers Prev 2020 Jul;29(7):1283-1289. [doi: 10.1158/1055-9965.EPI-20-0606] [Medline: 32371551]
6. mHealth: use of appropriate digital technologies for public health: report by the Director-General. World Health Organization. 2017. URL: https://apps.who.int/iris/handle/10665/274134 [accessed 2020-07-03]
7. The Mobile Economy 2020. GSM Association. 2020. URL: https://www.gsma.com/mobileeconomy/wp-content/uploads/2020/03/GSMA_MobileEconomy2020_Global [accessed 2020-05-11]
8. Sydow L. The State of Mobile in 2019 – The Most Important Trends to Know. App Annie. URL: https://www.appannie.com/en/insights/market-data/the-state-of-mobile-2019/ [accessed 2020-05-11]
9. Collado-Borrell R, Escudero-Vilaplana V, Calles A, Garcia-Martin E, Marzal-Alfaro B, Gonzalez-Haba E, et al. Oncology Patient Interest in the Use of New Technologies to Manage Their Disease: Cross-Sectional Survey. J Med Internet Res 2018 Oct 23;20(10):e11006 [FREE Full text] [doi: 10.2196/11006] [Medline: 30355554]
10. Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science 2020 May 05:1362-1367 [FREE Full text] [doi: 10.1126/science.abc0473] [Medline: 32371477]
11. Abeler J, Bäcker M, Buermeyer U, Zillessen H. COVID-19 Contact Tracing and Data Protection Can Go Together. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e19359 [FREE Full text] [doi: 10.2196/19359] [Medline: 32294052]
12. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020 Jul;46(7):427-431 [FREE Full text] [doi: 10.1136/medethics-2020-106314] [Medline: 32366705]
13. Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr 2020 Mar 11;19(1):8 [FREE Full text] [doi: 10.1186/s12942-020-00202-8] [Medline: 32160889]
14. Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e18936 [FREE Full text] [doi: 10.2196/18936] [Medline: 32240973]
15. Ensuring the Credibility of Health & Safety Information. Apple Developer. 2020 Mar 14. URL: https://developer.apple.com/news/?id=03142020a [accessed 2020-05-12]
16. Collado-Borrell R, Escudero-Vilaplana V, Ribed-Sánchez A, Ibáñez-García S, Herranz-Alonso A, Sanjurjo-Sáez M. Smartphone applications for cancer patients; what we know about them? Farm Hosp 2016 Jan 01;40(1):25-35 [FREE Full text] [doi: 10.7399/fh.2016.40.1.8993] [Medline: 26882831]
17. Cantudo-Cuenca MR, Robustillo-Cortés MA, Cantudo-Cuenca MD, Morillo-Verdugo R. A better regulation is required in viral hepatitis smartphone applications. Farm Hosp 2014 Apr 01;38(2):112-117 [FREE Full text] [Medline: 24669895]
18. Mayor S. Covid-19: Researchers launch app to track spread of symptoms in the UK. BMJ 2020 Mar 27;368:m1263. [doi: 10.1136/bmj.m1263] [Medline: 32220898]
19. Collado-Borrell R, Escudero-Vilaplana V, Ribed A, Anglada-Martínez H, Martín-Conde M, Herranz-Alonso A, et al. Design of the functional requirements of a smartphone app for patients receiving oral antineoplastic agents: The e-OncoSalud® app. J Oncol Pharm Pract 2020 Jan;26(1):105-115. [doi: 10.1177/1078155219840419] [Medline: 30955468]
20. Deb S. Privacy prescriptions for technology interventions on Covid-19 in India: IFF Working Paper No. 3/2020. Internet Freedom Foundation. 2020. URL: https://docs.google.com/document/d/1nDoPzygQyTetEguOlzula5O9y5f3f5YJDsA2Pd99O6U/edit# [accessed 2020-05-12]
21. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
22. Duong TA, Velter C, Rybojad M, Comte C, Bagot M, Sulimovic L, et al. Did Whatsapp reveal a new cutaneous COVID-19 manifestation? J Eur Acad Dermatol Venereol 2020 Apr 24:online ahead on print [FREE Full text] [doi: 10.1111/jdv.16534] [Medline: 32330322]

Abbreviations

COVID-19: coronavirus disease
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach, G Fagherazzi; submitted 17.05.20; peer-reviewed by M Perez, A Calles, O Windisch; comments to author 18.06.20; revised version received 01.07.20; accepted 01.07.20; published 25.08.20.

Please cite as:
Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M
Features and Functionalities of Smartphone Apps Related to COVID-19: Systematic Search in App Stores and Content Analysis
J Med Internet Res 2020;22(8):e20334
URL: https://www.jmir.org/2020/8/e20334
doi:10.2196/20334
PMID:32614777

Original Paper

Work-Related and Personal Factors Associated With Mental Well-Being During the COVID-19 Response: Survey of Health Care and Other Workers

Bradley A Evanoff[1,2], MD, MPH; Jaime R Strickland[1,2], MA; Ann Marie Dale[1,2], PhD, OTR/L; Lisa Hayibor[1], MPH; Emily Page[3], BFA; Jennifer G Duncan[1], MD; Thomas Kannampallil[1], PhD; Diana L Gray[1], MD

[1]Washington University School of Medicine, St. Louis, MO, United States

[2]Healthier Workforce Center of the Midwest, University of Iowa, Iowa City, IA, United States

[3]Human Resources, Washington University in St. Louis, St. Louis, MO, United States

Corresponding Author:
Bradley A Evanoff, MD, MPH
Washington University School of Medicine
4523 Clayton Avenue
Box 8005
St. Louis, MO, 63110
United States
Phone: 1 3144548340
Email: bevanoff@wustl.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/4/e29069

Abstract

Background: The response to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has created an unprecedented disruption in work conditions. This study describes the mental health and well-being of workers both with and without clinical exposure to patients with coronavirus disease (COVID-19).

Objective: The aim of this study is to measure the prevalence of stress, anxiety, depression, work exhaustion, burnout, and decreased well-being among faculty and staff at a university and academic medical center during the SARS-CoV-2 pandemic and describe work-related and personal factors associated with their mental health and well-being.

Methods: All faculty, staff, and postdoctoral fellows of a university, including its medical school, were invited in April 2020 to complete an online questionnaire measuring stress, anxiety, depression, work exhaustion, burnout, and decreased well-being. We examined associations between these outcomes and factors including work in high-risk clinical settings and family/home stressors.

Results: There were 5550 respondents (overall response rate of 34.3%). Overall, 34% of faculty and 14% of staff (n=915) were providing clinical care, while 61% of faculty and 77% of staff were working from home. Among all workers, anxiety (prevalence ratio 1.37, 95% CI 1.09-1.73), depression (prevalence ratio 1.28, 95% CI 1.03-1.59), and high work exhaustion (prevalence ratio 1.24, 95% CI 1.13-1.36) were independently associated with community or clinical exposure to COVID-19. Poor family-supportive behaviors by supervisors were also associated with these outcomes (prevalence ratio 1.40, 95% CI 1.21-1.62; prevalence ratio 1.69, 95% CI 1.48-1.92; and prevalence ratio 1.54, 95% CI 1.44-1.64, respectively). Age <40 years and a greater number of family/home stressors were also associated with these poorer outcomes. Among the subset of clinicians, caring for patients with COVID-19 and working in high-risk clinical settings were additional risk factors.

Conclusions: Our findings suggest that the pandemic has had negative effects on the mental health and well-being of both clinical and nonclinical employees. Mitigating exposure to COVID-19 and increasing supervisor support are modifiable risk factors that may protect mental health and well-being for all workers.

(J Med Internet Res 2020;22(8):e21366) doi:10.2196/21366

KEYWORDS

COVID-19; coronavirus; pandemic; mental health; health care workers; remote work; worker well-being; occupational health

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has created unprecedented disruption in social interactions and working conditions. Recent studies have described the effects of the pandemic on the mental health and well-being of frontline health care workers (HCWs) [1,2], and potential interventions to protect them [3-5]. Although concern over health and well-being has primarily focused on frontline HCWs, the pandemic has also affected working conditions in most other industries. Social and employment changes have led to concerns about an impending "second pandemic" of short- and long-term mental health issues [6], and predictions of a preventable surge of avoidable deaths from alcohol, drug use, and suicide [7]. Few studies describe the effects of the pandemic on the mental health and well-being of workers outside of health care. Such evidence is important for developing appropriate responses to the pandemic to preserve health and plan for economic and social recovery.

We describe results from the EMPOWER study (EMPlOyee Well-Being during Epidemic Response), which measured mental health and well-being among a large and diverse academic workforce, including those with and without clinical exposure to patients with coronavirus disease (COVID-19). The goals of the study were to measure the prevalence of stress, anxiety, depression, work exhaustion, burnout, and decreased mental well-being among faculty and staff at a university and its academic medical center during the SARS-CoV-2 pandemic; to compare mental health and well-being between clinical workers who were or were not caring for patients with COVID-19; and to identify other modifiable workplace and personal risk factors associated with mental health and well-being.

Methods

Study Design and Participants

We conducted a web-based survey of all benefits-eligible university employees (faculty, staff, and postdoctoral scholars) at Washington University in St. Louis, a private university with a large academic medical center where attending physicians and clinical staff are university employees. A separate survey was sent to physician trainees (residents and clinical fellows) and is not included in this report [8]. An email invitation was sent to all benefits-eligible employees on April 17, 2020, with a clickable link to a voluntary, anonymous online survey. A single reminder email was sent 10 days later. The survey period was approximately 4 to 5 weeks after the university had enacted work-from-home plans. The study was approved by the institutional review board of Washington University in St. Louis.

Survey Instrument

The survey was designed to take less than 10 minutes to complete. Definitions and sources of personal factors, work factors and well-being variables used in the survey are shown in Multimedia Appendix 1. Demographic questions included age, race, household income; children, dependents, and other adults living at home; and work status of partner. Questions about work included current work status (on-site work involving clinical care, on-site work not involving clinical care, working from home, or not working). Those doing on-site work in clinical care were asked about the clinical setting, and if they had cared for patients with COVID-19. All participants were asked if they or a member of their household had received a medical diagnosis or positive test result for COVID-19 or if they had been exposed to someone with COVID-19.

The questionnaire also included three questions from the FSSB-SF [9], which measures supervisor behaviors supportive of family roles (eg, "Your supervisor makes you feel comfortable talking to him/her about your conflicts between work and nonwork"; "Your supervisor demonstrates effective behaviors in how to juggle work and nonwork issues"; "Your supervisor works effectively with employees to creatively solve conflicts between work and nonwork.") We used the mean value of these three responses as the supervisor support variable. We also asked about 8 potential family/home stressors related to the pandemic (childcare, home schooling, caring for elderly relatives, having access to food and other necessities, being infected, friends and family being infected, keeping your job, and personal finances). These questions were asked in the format "Currently how stressed are you about…?" in a 5-point scale from "not at all" to "extremely" stressed. The number of stressors reported by each individual as "somewhat" to "extremely" were totaled to create a composite stress score (range 0-8).

Outcome Measures

Study outcomes included stress, anxiety, and depression as measured by the Depression, Anxiety and Stress Scale - 21 Items (DASS-21) [10], burnout and work exhaustion as measured by the Professional Fulfillment Index (PFI) [11], and changes in well-being [12]. The DASS-21 is a validated instrument with scales that correlate well with other measures of depression, anxiety, and stress. Due to the PFI questionnaire structure, burnout was only assessed among HCWs. Self-reported changes in well-being comparing current to prepandemic status were assessed in five domains (overall, financial, physical, mental, and social) by the question "To what extent have COVID-19–related work/life changes impacted your well-being?" using a 4-point scale from "much worse" to "much better/somewhat better."

Statistical Analyses

We contrasted the proportions or means of outcomes between faculty and staff and those in different clinical settings. We then conducted univariable and multivariable Poisson regression with robust sandwich estimators to examine personal and work factors associated with six mental health and well-being outcomes described above: stress, anxiety, depression, burnout, work exhaustion, and changes in well-being.

In conducting these analyses, we selected 10 a priori potential personal and work factors as independent variables for multivariable analysis (supervisor support, clinical work, staff [versus faculty or postdoctoral fellow], exposure to people [or patients for clinicians] with a diagnosis of COVID-19, age, sex, race, annual household income, children aged under 18 years living at home, and composite stressor count). Results were expressed as prevalence ratios (PRs) with 95% CIs. Independent variables were dichotomized at the median scores or at relevant cut-points for ordinal variables. We categorized race and ethnicity as "underrepresented groups" (those identifying as Black/African American, Native American, Hawaiian/Pacific Islander, or Hispanic) and "Other." The significance level was set at .05 and hypothesis tests were two-sided. All analyses were performed with R statistical software (Version 4.0.0; R Foundation for Statistical Computing) [13] and R studio (Version 1.2.504) [14].

Patient and Public Involvement

The survey was developed in collaboration with the university human resources department and the employee wellness director to ensure sensitivity to current issues and to address emerging concerns about employee wellness during the pandemic response. Initial survey results have been shared with university leaders to highlight the mental health needs of employees. Study results are driving plans to communicate broadly with faculty, staff, and trainees to highlight mental health challenges faced by our workforce and to better publicize and encourage employees to utilize available mental health resources.

Results

Email invitations were sent to all benefits-eligible university faculty, staff, and postdoctoral scholars (N=16,238). In total, 5706 responses were received (Figure 1); there were 5569 unique responses after the exclusion of 137 responses with a duplicate self-generated identifier that allows anonymous longitudinal follow-up. Of the remaining surveys, 19 were excluded for missing status as faculty, staff, or postdoctoral scholar, leaving 5550 surveys for analysis (870 faculty, 4470 staff, and 210 postdoctoral fellows). The overall response rate was 34.3% for unique surveys. Response rates were higher for staff than for faculty (40% versus 19.7%).

Figure 1. Survey response flowchart.

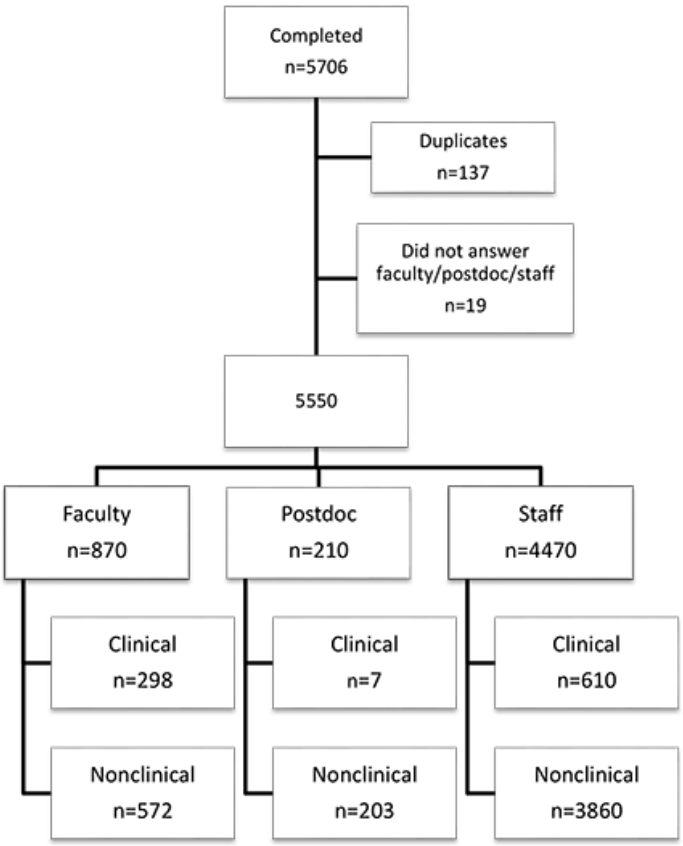


Table 1 compares demographics, work factors, and outcomes between faculty, staff, and postdoctoral fellows. Overall, 34.3% of faculty and 13.6% of staff reported working on-site in clinical operations while a majority of the faculty (60.6%) and staff (76.5%) were working from home. Smaller numbers worked on-site in nonclinical roles and few were not working. A majority of faculty (50.4%) reported that their workload increased after the COVID-19 workplace changes, as compared to 40.4% of staff and 21% of postdoctoral fellows. Overall, a majority of respondents reported being stressed (more than "a little bit") about personal finances, keeping their jobs, and about themselves or friends or family becoming infected. Of those with children at home, a majority reported feeling stressed about homeschooling; most of those providing care to older relatives reported stress about their care. Distributions of most perceived stressors were significantly different across the faculty, staff,

and postdoctoral fellows, with postdoctoral fellows more frequently reporting stress about childcare, homeschooling, and access to food and essential supplies. Faculty, staff, and postdoctoral fellows all reported a high prevalence of worsened overall well-being (58.3%) related to COVID-19 work or life changes. Moderate to high levels of stress were reported by 13%, anxiety by 13%, depression by 15.9%, and high work exhaustion by 43%.

Table 1. Comparison of demographics, personal factors, work factors, and outcomes between faculty, staff, and postdoctoral fellows[a].

	Faculty (n=870)	Staff (n=4470)	Postdoctoral fellows (n=210)	Total (N=5550)	*P* value
Personal and family factors					
Age above 40 years, n (%)	624 (72.0)	2652 (59.5)	22 (10.5)	3298 (59.6)	<.001
Gender, n (%)					<.001
Male	333 (38.4)	772 (17.3)	77 (37.0)	1182 (21.4)	N/A[b]
Female	523 (60.3)	3624 (81.3)	127 (61.1)	4274 (77.3)	N/A
Gender diverse	4 (0.5)	18 (0.4)	2 (1.0)	24 (0.4)	N/A
Prefer not to say	8 (0.9)	41 (0.9)	2 (1.0)	51 (0.9)	N/A
Underrepresented groups[c], n (%)	68 (7.8)	482 (10.8)	26 (12.4)	576 (10.4)	.02
Annual household income <$70,000, n (%)	68 (8.2)	1551 (36.5)	133 (64.3)	1752 (33.2)	<.001
Living alone, n (%)	111 (12.8)	645 (14.5)	62 (29.8)	818 (14.8)	<.001
Two adults in health care with children, n (%)	68 (7.8)	58 (1.3)	2 (1.0)	128 (2.3)	<.001
Stressed about childcare[d], n (%)	193 (46.5)	652 (36.7)	26 (53.1)	871 (38.9)	<.001
Stressed about home schooling[e], n (%)	216 (61.7)	846 (56.8)	22 (84.6)	1084 (58.1)	.006
Stressed about relatives[f], n (%)	87 (73.7)	560 (75.9)	12 (75.0)	659 (75.6)	.88
Stressed about essential supplies, n (%)	199 (23.0)	1341 (30.2)	77 (36.7)	1617 (29.3)	<.001
Stressed about being infected, n (%)	491 (56.5)	2556 (57.5)	101 (48.1)	3148 (57.0)	.03
Stressed about friends or family getting infected, n (%)	665 (76.7)	3347 (75.2)	130 (61.9)	4142 (75.0)	<.001
Stressed about keeping job, n (%)	288 (33.3)	2786 (62.6)	116 (55.5)	3190 (57.8)	<.001
Stressed about personal finances, n (%)	422 (49.2)	2698 (60.8)	110 (53.1)	3230 (58.7)	<.001
Number of stressors, mean (SD)	2.9 (1.9)	3.3 (1.9)	2.8 (1.8)	3.2 (1.9)	<.001
Any exposure to COVID-19, n (%)	142 (16.3)	272 (6.1)	11 (5.2)	425 (7.7)	<.001
Work factors					
Current work, n (%)					<.001
Working onsite, clinical operations	298 (34.3)	610 (13.6)	7 (3.3)	915 (16.5)	N/A
Working onsite, nonclinical operations	33 (3.8)	339 (7.6)	18 (8.6)	390 (7.0)	N/A
Working at home	527 (60.6)	3421 (76.5)	183 (87.1)	4131 (74.4)	N/A
Not working	12 (1.4)	100 (2.2)	2 (1.0)	114 (2.1)	N/A
Supervisor support scale (range 1-5), mean (SD)	2.5 (1.0)	2.2 (1.1)	2.3 (1.1)	2.3 (1.1)	<.001
Increased workload since COVID-19 restrictions began, n (%)	426 (50.4)	1747 (40.4)	43 (21.0)	2216 (41.2)	<.001
Outcomes					
Worse overall well-being due to COVID-19–related work or life changes, n (%)	588 (67.8)	2490 (56.2)	130 (62.2)	3208 (58.3)	<.001
Worse financial well-being due to COVID-19–related work or life changes, n (%)	381 (43.9)	1291 (29.1)	60 (28.6)	1732 (31.4)	<.001
Worse physical well-being due to COVID-19–related work or life changes, n (%)	387 (44.6)	1938 (43.7)	88 (41.9)	2413 (43.8)	.77
Worse mental well-being due to COVID-19–related work or life changes, n (%)	604 (69.7)	3027 (68.1)	142 (67.6)	3773 (68.4)	.63
Worse social well-being due to COVID-19–related work or life changes, n (%)	703 (81.2)	3482 (78.5)	168 (80.4)	4353 (79.0)	.18
Mean well-being score, mean (SD)	2.3 (0.5)	2.4 (0.5)	2.4 (0.5)	2.4 (0.5)	<.001

	Faculty (n=870)	Staff (n=4470)	Postdoctoral fellows (n=210)	Total (N=5550)	*P* value
Moderate to high depression (DASS), n (%)	133 (15.9)	676 (15.7)	39 (19.5)	848 (15.9)	.36
Moderate to high anxiety (DASS), n (%)	83 (10.0)	582 (13.5)	30 (14.9)	695 (13.0)	.02
Moderate to high stress (DASS), n (%)	105 (12.6)	552 (12.7)	39 (20.0)	696 (13.0)	.01
High work exhaustion, n (%)	419 (49.7)	1783 (41.3)	105 (51.2)	2307 (43.0)	<.001

[a]Missing values for each variable (range 0%-4.8%) were omitted from percentage calculations. Percentages may not total 100 due to rounding. Categorical variables are displayed as n (%), while continuous variables are displayed as mean (SD). The chi-square test was used for categorical variables, while analysis of variance was used for continuous variables.

[b]N/A: not applicable.

[c]Underrepresented groups were those identifying as Black/African American, Native American, Hawaiian/Pacific Islander, or Hispanic.

[d]Percentages are among those with children only.

[e]Percentages are among those with children above preschool age only.

[f]Percentages are among those with elderly parents or relatives only.

Multivariable analyses of associations between these outcomes and a common set of work and personal factors among all respondents showed that three factors were statistically significantly associated with a higher prevalence of all five outcomes (Table 2, univariable analyses in Multimedia Appendix 2): poor supervisor support, a higher number of family/home stressors, and age <40 years. Working on-site in clinical operations was associated with higher anxiety and lower mean well-being; being a staff member (rather than faculty or postdoctoral fellow) was associated with better well-being and lower prevalence of stress and work exhaustion. Reported exposure to COVID-19 (diagnosis in self or family, or exposure to someone likely to have COVID-19) was associated with higher stress, anxiety, depression, and work exhaustion. A household income of <$70,000 was associated with a higher prevalence of stress, anxiety, and depression. Women were more likely to report experiencing anxiety, work exhaustion, and decreased well-being. Unanticipated protective factors were also notable: having children at home was associated with a lower prevalence of anxiety and depression, and underrepresented racial/ethnic groups were less likely to report stress, depression, or decreased well-being.

Table 2. Multivariate associations between personal factors, work factors, and well-being among all participants (N=5550).

Variable	Moderate to high stress (DASS), PR (95% CI)[a]	Moderate to high anxiety (DASS), PR (95% CI)	Moderate to high depression (DASS), PR (95% CI)	High work exhaustion, PR (95% CI)	Decreased overall well-being, PR (95% CI)
Age >40 years	0.46 (0.40-0.54)	0.53 (0.46-0.62)	0.49 (0.43-0.56)	0.67 (0.63-0.72)	0.89 (0.86-0.93)
Female	1.16 (0.96-1.40)	1.36 (1.11-1.67)	0.94 (0.81-1.11)	1.18 (1.08-1.28)	1.06 (1.00-1.12)
Underrepresented groups[b]	0.79 (0.62-1.02)	0.99 (0.79-1.24)	0.74 (0.59-0.93)	0.92 (0.83-1.02)	0.91 (0.84-0.98)
Annual household income <$70,000	1.24 (1.06-1.44)	1.43 (1.22-1.67)	1.39 (1.21-1.59)	0.94 (0.87-1.00)	0.97 (0.93-1.02)
Children <18 years old living at home	0.96 (0.83-1.12)	0.85 (0.73-0.99)	0.75 (0.65-0.86)	1.01 (0.94-1.07)	0.98 (0.94-1.03)
High number of stressors[c]	2.17 (1.86-2.54)	2.18 (1.86-2.56)	1.51 (1.32-1.72)	1.37 (1.29-1.46)	1.43 (1.37-1.50)
Staff versus faculty and postdoctoral fellows	0.81 (0.68-0.97)	1.09 (0.89-1.33)	0.94 (0.80-1.11)	0.85 (0.79-0.92)	0.90 (0.85-0.95)
Exposure to coronavirus disease	1.48 (1.19-1.84)	1.37 (1.09-1.73)	1.28 (1.03-1.59)	1.24 (1.13-1.36)	1.04 (0.97-1.12)
Clinical	0.92 (0.76-1.11)	1.21 (1.01-1.45)	0.98 (0.82-1.16)	1.01 (0.93-1.10)	1.18 (1.12-1.24)
Poor supervisor support[d]	1.58 (1.37-1.83)	1.40 (1.21-1.62)	1.69 (1.48-1.92)	1.54 (1.44-1.64)	1.11 (1.07-1.16)

[a]Prevalence ratios (PRs) and 95% CIs were calculated using Poisson regression models.

[b]Underrepresented groups were those identifying as Black/African American, Native American, Hawaiian/Pacific Islander, or Hispanic.

[c]A high number of stressors was defined as a composite stress score >3 (median).

[d]Poor supervisor support was defined as a supervisor support scale score >2 (median).

A comparison of outcomes between faculty and staff working in clinical settings is shown in Table 3 (univariable analyses in Multimedia Appendix 3). Those working in high-risk settings (intensive care unit, emergency room, or performing procedures likely to generate respiratory aerosols) were more likely to report caring for patients with COVID-19 and experiencing an increased workload since COVID-19 restrictions began, had a worse mean score on changes in well-being, and were more likely to report moderate to high stress and depression, high work exhaustion, and burnout. Multivariable analysis of faculty

and staff working in clinical operations showed that caring for patients who had COVID-19 was associated with a higher prevalence of stress, anxiety, burnout, and work exhaustion (Table 4). High-risk clinical work (intensive care unit, emergency department, aerosol-generating procedures) showed similar, albeit weaker associations with these outcomes in multivariable analysis (data not shown). There were no statistically significant differences between clinically active staff and faculty for any outcome. Notably, low supervisor support was strongly associated with all mental health and well-being outcomes, and having a high number of family/home stressors was associated with all outcomes except depression.

Table 3. Comparison of work factors and outcomes among all clinicians and between high-risk and non–high-risk clinical groups[a].

Factors and outcomes	Not working in high-risk clinical settings (N=740)	Working in high-risk clinical settings (N=175)	All clinicians (N=915)	*P* value
Work factors				
Contact with outpatients, n (%)	534 (72.2)	77 (44.0)	611 (66.8)	<.001
Contact with inpatients, n (%)	143 (19.3)	112 (64.0)	255 (27.9)	<.001
Working in an intensive care unit, n (%)	0 (0.0)	68 (38.9)	68 (7.4)	<.001
Working in the emergency room, n (%)	0 (0.0)	51 (29.1)	51 (5.6)	<.001
Performing procedures that create respiratory aerosols, n (%)	0 (0.0)	106 (60.6)	106 (11.6)	<.001
Caring for patients with COVID-19[b], n (%)	123 (16.8)	127 (73.8)	250 (27.6)	<.001
Increased workload since COVID-19 restrictions began, n (%)	279 (38.0)	85 (49.4)	364 (40.2)	.006
Supervisor support scale (range 1-5), mean (SD)	2.5 (1.1)	2.4 (1.1)	2.5 (1.1)	.50
Outcomes				
Worse overall well-being due to COVID-19–related work or life changes, n (%)	500 (67.9)	127 (73.0)	627 (68.9)	.20
Worse financial well-being due to COVID-19–related work or life changes, n (%)	313 (42.6)	107 (61.5)	420 (46.2)	<.001
Worse physical well-being due to COVID-19–related work or life changes, n (%)	339 (46.1)	100 (57.1)	439 (48.2)	.009
Worse mental well-being due to COVID-19–related work or life changes, n (%)	564 (76.5)	141 (81.0)	705 (77.4)	.20
Worse social well-being due to COVID-19–related work or life changes, n (%)	629 (85.7)	149 (85.1)	778 (85.6)	.85
Mean well-being score, mean (SD)	2.2 (0.4)	2.1 (0.5)	2.2 (0.5)	.001
Moderate to high depression (DASS), n (%)	108 (15.1)	37 (21.6)	145 (16.4)	.04
Moderate to high anxiety (DASS), n (%)	125 (17.6)	27 (15.8)	152 (17.2)	.58
Moderate to high stress (DASS), n (%)	93 (13.0)	35 (20.3)	128 (14.5)	.01
High work exhaustion, n (%)	342 (46.8)	105 (60.7)	447 (49.5)	.001
High overall burnout, n (%)	233 (32.0)	74 (42.8)	307 (34.0)	.007

[a]The high-risk group reported working in an emergency room, intensive care unit, or performing procedures generating respiratory aerosols. Missing values for each variable (range 0%-3.5%) were omitted from percentage calculations. Percentages may not total 100 due to rounding. Categorical variables are displayed as n (%), while continuous variables are displayed as mean (SD). A chi-square test was used for categorical variables, while a *t* test was used for continuous variables.

[b]COVID-19: coronavirus disease.

Table 4. Multivariate associations between personal factors, work factors, and well-being among participants doing clinical work (N=915)[a].

Variable	Moderate to high stress (DASS), PR (95% CI)	Moderate to high anxiety (DASS), PR (95% CI)	Moderate to high depression (DASS), PR (95% CI)	High overall burnout, PR (95% CI)	High work exhaustion, PR (95% CI)	Decreased overall well-being, PR (95% CI)
Age >40 years	0.56 (0.39-0.81)	0.73 (0.53-1.00)	0.60 (0.43-0.84)	0.77 (0.64-0.93)	0.81 (0.71-0.93)	0.89 (0.82-0.96)
Female	1.26 (0.79-2.00)	1.47 (0.90-2.39)	1.19 (0.77-1.85)	1.18 (0.92-1.51)	1.20 (0.99-1.45)	1.08 (0.97-1.20)
Underrepresented groups[b]	0.56 (0.32-0.98)	0.74 (0.46-1.20)	0.60 (0.35-1.05)	0.66 (0.46-0.94)	0.96 (0.78-1.20)	0.90 (0.78-1.04)
Annual household income <$70,000	1.65 (1.11-2.47)	1.59 (1.11-2.29)	1.46 (1.02-2.11)	1.13 (0.89-1.44)	0.85 (0.72-1.01)	0.91 (0.82-1.01)
Children <18 years old living at home	0.97 (0.68-1.38)	1.07 (0.78-1.47)	0.91 (0.66-1.26)	1.09 (0.90-1.32)	1.06 (0.92-1.21)	0.90 (0.83-0.98)
High number of stressors[c]	1.92 (1.29-2.86)	1.76 (1.22-2.53)	1.23 (0.88-1.70)	1.47 (1.20-1.81)	1.33 (1.15-1.54)	1.27 (1.16-1.39)
Staff	0.97 (0.64-1.46)	1.51 (0.97-2.35)	1.10 (0.74-1.64)	0.88 (0.71-1.10)	1.11 (0.95-1.31)	0.92 (0.84-1.01)
Caring for patients with coronavirus disease	1.73 (1.22-2.46)	1.60 (1.14-2.23)	1.25 (0.88-1.79)	1.38 (1.14-1.67)	1.28 (1.11-1.46)	0.99 (0.91-1.09)
Poor supervisor support[d]	1.93 (1.33-2.81)	1.69 (1.22-2.35)	1.96 (1.39-2.76)	1.99 (1.61-2.47)	1.62 (1.39-1.88)	1.16 (1.06-1.26)

[a]Prevalence ratios (PRs) and 95% CIs were calculated using Poisson multiple regression.

[b]Underrepresented groups were those identifying as Black/African American, Native American, Hawaiian/Pacific Islander, or Hispanic.

[c]A high number of stressors was defined as a composite stress score >3 (median).

[d]Poor supervisor support was defined as a supervisor support scale score >2 (median).

Discussion

Principal Results

The EMPOWER study found a high prevalence of stress, anxiety, depression, work exhaustion, burnout, and worsened well-being among clinical and nonclinical university employees surveyed approximately 4 to 5 weeks after work-from-home policies were implemented for those performing work deemed "nonessential" during the crisis phase of the pandemic. These findings uniquely highlight the associations of health and well-being with additional personal and work factors beyond those addressed in existing studies of HCWs during the SARS-CoV-2 pandemic. Importantly, our study also reports on workers outside of clinical medicine, whose health and well-being has been minimally studied. A unique finding of this study is that the factors with the strongest consistent associations with all health and well-being outcomes in both clinical and nonclinical workers were items from the FSSB-SF, a measure of general perception of family-specific supervisory support [9], and a sum of 8 stressors related to family/home life and financial security. Perceived supervisor support for family is a pathway through which employees develop perceptions of organizational support [15], plays a major role influencing the health and well-being of workers [16], and is associated with reduction in work-family conflict, improved well-being, and increased job satisfaction [15,17]. Importantly, family-supportive supervisor behavior can be modified by employer policies and practices.

Limitations

Limitations of this study include its cross-sectional design, so associations between potential risk factors and outcomes of health and well-being may not be causal. In particular, participants with poorer well-being might differentially report supervisor behaviors. The overall response rate of 34.5% means that the respondents may not be fully representative of all university employees. Faculty were less likely to participate than were staff (19.7% versus 40%), and comparisons between these groups should be interpreted with caution. For instance, faculty were more likely to report increased workload and more work exhaustion since the onset of the pandemic; this difference may be due to differential reporting by faculty, or because faculty were in fact busier and more exhausted and thus less likely to respond. Since the survey was anonymous, our study relies entirely on self-reported data. We studied employees of one university, who may not be representative of other workforces, including lower-paid workers. The St. Louis region was an early adopter of physical distancing and has had a later peak of SARS-CoV-2 and a lower incidence of patients with COVID-19 than some other areas of the United States.

Strengths of the study include its large size, examination of employees who are not in health care, and evaluation of both family/home stressors and workplace factors including supervisor support. To our knowledge, this is the first large US study of multiple mental health and well-being outcomes related to the pandemic outside of a HCW population. We are conducting repeated surveys to track changes in individual health and well-being over time, and to allow more robust causal inferences.

Comparison With Prior Work

Our findings among clinical workers, both faculty (primarily physicians) and staff (primarily nurses), are broadly consistent with findings from other cross-sectional studies of HCWs caring for patients with COVID-19. A study of 1257 HCWs in China used different instruments and found a higher prevalence of depression and anxiety than seen in our study [2]. Their study reported that HCWs directly involved in the care of patients

with COVID-19 were at a greater risk of anxiety and depression, similar to our findings of increased risks of stress, anxiety, burnout, and work exhaustion. A study of 906 HCWs in Singapore and India [18], using the DASS-21, found moderate to severe stress in 3.8%, anxiety in 2.2%, and depression in 8.7%, much lower than the prevalence of 14.5%, 17.2%, and 16.4% seen in our study. Our finding that family/home stressors and supervisor support for family-work balance were strongly associated with mental health and well-being outcomes are consistent with the findings of a recent review of psychological reactions of HCWs during past epidemics [19]. Their analyses showed that responsibilities of caring for family members and lower household income were associated with poorer mental health outcomes among HCWs. Although HCWs caring for patients with COVID-19 had worse mental well-being than their fellow faculty and staff, those working from home or on-site in nonclinical roles also had appreciable rates of poor outcomes. Although we do not have baseline measures for the well-being and mental health outcomes in our study, respondents described altered well-being related to COVID-19–related work/life changes, with 14.6% reporting "much worse" and 68% reporting "much worse" or "somewhat worse" mental well-being. These findings are strikingly similar to those of an April 2020 poll by the Kaiser Family Foundation. Among those who had not experienced job or income loss, 15% reported major negative impacts on their mental health from worry or stress over coronavirus, and 54% reported some negative mental health impacts [20]. Our findings are also supported by results from a recent national online survey conducted among US adults, which compared responses in April 2020 to those from the National Health Interview Survey in 2018 [21]. This study found a higher prevalence of serious psychological distress (13.6% versus 3.9%) in 2020, with younger age and lower income predicting a higher prevalence of distress.

University staff and to some extent faculty are representative of the larger nonclinical workforce that is undergoing uniquely stressful circumstances that blur the boundaries between work and family as people work from home, find it difficult to work because their children's schools and daycares are closed, or worry about bringing an infection home to their families. Although frontline HCWs are at uniquely high risk due to their work, our study shows that effects of family and home stresses and of supervisor support play a large role in their health and well-being. Appreciation of these factors has been largely missing from studies of risk factors for mental health and well-being among HCWs during this pandemic. These same family and home stresses and supervisor support also influence the health of the broader working population. As the pandemic continues in the months and perhaps years to come, our concern over the mental health and well-being of HCWs must broaden to include other worker groups as well.

There are many possible interventions to address the health and well-being of the clinical and nonclinical workforces. A systematic review found that organizational and social support, clear communication, and having a sense of control were protective factors for adverse mental health outcomes among HCWs during prior epidemics [22]. Recent publications have stressed the importance of robust organizational responses to address the mental health and well-being of frontline HCWs [6,23]. Many of these interventions should be applicable outside of the health care setting. Although interventions aimed at improving resilience among individual workers may lead to improvements in burnout rates and other well-being measures, organizational-level interventions that reduce perceived work demands or increase resources are generally more effective [24]. Our data would suggest that organizations should explicitly focus on improving supervisor support for work-family issues. Evaluation of interventions training supervisors in family-supportive behaviors, including a study in HCWs, have suggested that such training is associated with improved reports of physical health, job satisfaction, job engagement, and decreased intent to leave the current job [25,26]. Future research should include longitudinal studies to follow mental well-being over time, include more workers outside of health care to better understand the effects on the broader population, and test both individual-level and institutional-level interventions to mitigate the effects of the pandemic on mental health.

Conclusions

Both health care and other workers have encountered worsened mental health and well-being as a result of the SARS-CoV-2 pandemic. Employers, health care systems, and public health agencies should begin interventions to improve mental health and overall well-being among HCWs and the broader workforce. In addition to traditional wellness interventions addressing resilience and mental health issues among individual workers, responses should include support for work/family balance and other organizational changes to improve work conditions for health care and other workers.

Acknowledgments

This study was supported, in part, by the Healthier Workforce Center of the Midwest grant U19OH008868 from the Centers for Disease Control and Prevention (CDC), and by the Washington University Institute of Clinical and Translational Sciences grant UL1TR002345 from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the CDC, NCATS, or NIH.

Authors' Contributions

BAE was responsible for study conception and design, survey development, analytic oversight, and drafting of this article. JRS contributed to study conception and design, survey development, coordinating data collection, and provided input on analytic

decisions, interpretation, and manuscript development. AMD provided input on study design, analytic decisions, interpretation of results, and manuscript development. LH contributed to survey development, managed online survey collection, conducted analyses, and assisted with manuscript development. EP assisted with study design, survey development, and data collection, contributed to the development of the manuscript, and communicated results back to university employees. JGD, TK, and DLG provided input on study design, survey development, analytic decisions, interpretation of results, and manuscript development. All authors reviewed and approved the final manuscript draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions and sources of personal factors, work factors, and well-being variables.
[DOCX File , 17 KB - jmir_v22i8e21366_app1.docx]

Multimedia Appendix 2
Univariable associations between personal factors, work factors, and well-being among all participants.
[DOCX File , 17 KB - jmir_v22i8e21366_app2.docx]

Multimedia Appendix 3
Univariable associations between personal factors, work factors, and well-being among participants doing clinical work.
[DOCX File , 17 KB - jmir_v22i8e21366_app3.docx]

References

1. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
2. Bettinsoli M, Di RD, Napier J, Moretti L, Bettinsoli P, Delmedico M, et al. Psychological Impact and Contextual Factors Associated With Physical and Mental Health Conditions of Italian Healthcare Professionals During the Covid-19 Disease Outbreak. PsyArXiv Preprints 2020:1-22. [doi: 10.31234/osf.io/w89fz]
3. Wu AW, Connors C, Everly GS. COVID-19: Peer Support and Crisis Communication Strategies to Promote Institutional Resilience. Annals of Internal Medicine 2020 Jun 16;172(12):822-823. [doi: 10.7326/m20-1236]
4. Ripp J, Peccoralo L, Charney D. Attending to the Emotional Well-Being of the Health Care Workforce in a New York City Health System During the COVID-19 Pandemic. Acad Med 2020 Aug;95(8):1136-1139. [doi: 10.1097/acm.0000000000003414]
5. Shanafelt T, Ripp J, Trockel M. Understanding and Addressing Sources of Anxiety Among Health Care Professionals During the COVID-19 Pandemic. JAMA 2020 Apr 07;323(21):2133. [doi: 10.1001/jama.2020.5893] [Medline: 32259193]
6. Galea S, Merchant RM, Lurie N. The Mental Health Consequences of COVID-19 and Physical Distancing: The Need for Prevention and Early Intervention. JAMA Intern Med 2020 Apr 10:2020. [doi: 10.1001/jamainternmed.2020.1562] [Medline: 32275292]
7. Petterson S, Westfall J, Miller B. Projected Deaths of Despair from COVID-19. Well Being Trust. 2020 Apr. URL: https://wellbeingtrust.org/wp-content/uploads/2020/05/WBT_Deaths-of-Despair_COVID-19-FINAL-FINAL.pdf [accessed 2020-08-20]
8. Kannampallil TG, Goss CW, Evanoff BA, Strickland JR, McAlister RP, Duncan J. Exposure to COVID-19 patients increases physician trainee stress and burnout. PLoS One 2020 Aug 6;15(8):e0237301. [doi: 10.1371/journal.pone.0237301] [Medline: 32760131]
9. Hammer LB, Ernst Kossek E, Bodner T, Crain T. Measurement development and validation of the Family Supportive Supervisor Behavior Short-Form (FSSB-SF). J Occup Health Psychol 2013 Jul;18(3):285-296 [FREE Full text] [doi: 10.1037/a0032612] [Medline: 23730803]
10. Lovibond SH, Lovibond PF. Manual for the depression anxiety stress scales. Sydney: Psychology Foundation of Australia; 1996.
11. Trockel M, Bohman B, Lesure E, Hamidi MS, Welle D, Roberts L, et al. A Brief Instrument to Assess Both Burnout and Professional Fulfillment in Physicians: Reliability and Validity, Including Correlation with Self-Reported Medical Errors, in a Sample of Resident and Practicing Physicians. Acad Psychiatry 2018 Feb 1;42(1):11-24 [FREE Full text] [doi: 10.1007/s40596-017-0849-3] [Medline: 29196982]
12. Serxner S, Kichlu R, Ratelis E. Consumer sentiment during a time of global crisis. Optum Inc. 2020. URL: https://www.optum.com/content/dam/optum3/optum/en/resources/PDFs/2020-comparative-insights-during-global-crisis.pdf [accessed 2020-08-20]

13. R Foundation for Statistical Computing. R: A language environment for statistical computing. URL: https://www.r-project.org/ [accessed 2020-08-20]
14. RStudio Inc. RStudio: Integrated Development for R. Boston, MA URL: https://rstudio.com/ [accessed 2020-08-20]
15. Kossek EE, Pichler S, Bodner T, Hammer LB. Workplace social support and work-family conflict: a meta-analysis clarifying the influence of general and work-family-specific supervisor and organizational support. Pers Psychol 2011;64(2):289-313 [FREE Full text] [doi: 10.1111/j.1744-6570.2011.01211.x] [Medline: 21691415]
16. King R, Karuntzos G, Casper L. Work-family balance issues and work-leave policies. In: Handbook of occupational health and wellness. Boston, MA: Springer; 2012:323-340.
17. Hammer LB, Kossek EE, Yragui NL, Bodner TE, Hanson GC. Development and Validation of a Multidimensional Measure of Family Supportive Supervisor Behaviors (FSSB). J Manage 2009 Aug;35(4):837-856 [FREE Full text] [doi: 10.1177/0149206308328510] [Medline: 21660254]
18. Chew NW, Lee GK, Tan BY, Jing M, Goh Y, Ngiam NJ, et al. A multinational, multicentre study on the psychological outcomes and associated physical symptoms amongst healthcare workers during COVID-19 outbreak. Brain Behav Immun 2020 Aug;88:559-565 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.049] [Medline: 32330593]
19. Kisely S, Warren N, McMahon L, Dalais C, Henry I, Siskind D. Occurrence, prevention, and management of the psychological effects of emerging virus outbreaks on healthcare workers: rapid review and meta-analysis. BMJ 2020 May 05;369:m1642 [FREE Full text] [doi: 10.1136/bmj.m1642] [Medline: 32371466]
20. Kirzinger A, Kearney A, Hamel L, Brodie M. KFF Health Tracking Poll - Early April 2020: The Impact of Coronavirus on Life in America. Kaiser Family Foundation. 2020 Apr 2. URL: https://www.kff.org/coronavirus-covid-19/report/kff-health-tracking-poll-early-april-2020/ [accessed 2020-08-20]
21. McGinty EE, Presskreischer R, Han H, Barry CL. Psychological Distress and Loneliness Reported by US Adults in 2018 and April 2020. JAMA 2020 Jun 03;324(1):94 [FREE Full text] [doi: 10.1001/jama.2020.9740] [Medline: 32492088]
22. De Brier N, Stroobants S, Vandekerckhove P, De Buck E. Factors affecting mental health of health care workers during coronavirus disease outbreaks: a rapid systematic review. PsyArXiv Preprints 2020:1-17. [doi: 10.31234/osf.io/w9uxs]
23. Dzau VJ, Kirch D, Nasca T. Preventing a Parallel Pandemic — A National Strategy to Protect Clinicians' Well-Being. N Engl J Med 2020 Aug 06;383(6):513-515. [doi: 10.1056/nejmp2011027]
24. Panagioti M, Panagopoulou E, Bower P, Lewith G, Kontopantelis E, Chew-Graham C, et al. Controlled Interventions to Reduce Burnout in Physicians: A Systematic Review and Meta-analysis. JAMA Intern Med 2017 Feb 01;177(2):195-205. [doi: 10.1001/jamainternmed.2016.7674] [Medline: 27918798]
25. Odle-Dusseau HN, Hammer LB, Crain TL, Bodner TE. The influence of family-supportive supervisor training on employee job performance and attitudes: An organizational work-family intervention. J Occup Health Psychol 2016 Jul;21(3):296-308. [doi: 10.1037/a0039961] [Medline: 26652264]
26. Hammer LB, Kossek EE, Anger WK, Bodner T, Zimmerman KL. Clarifying work-family intervention processes: the roles of work-family conflict and family-supportive supervisor behaviors. J Appl Psychol 2011 Jan;96(1):134-150 [FREE Full text] [doi: 10.1037/a0020927] [Medline: 20853943]

Abbreviations

COVID-19: coronavirus disease
DASS-21: Depression, Anxiety and Stress Scale - 21 Items
EMPOWER: EMPlOyee Well-Being during Epidemic Response
FSSB-SF: Family Supportive Supervisor Behavior Short-Form
HCW: health care worker
PFI: Professional Fulfillment Index
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by A Moorhead, G Eysenbach; submitted 12.06.20; peer-reviewed by A Wu, B Knezevic; comments to author 29.07.20; revised version received 03.08.20; accepted 06.08.20; published 25.08.20.

Please cite as:
Evanoff BA, Strickland JR, Dale AM, Hayibor L, Page E, Duncan JG, Kannampallil T, Gray DL
Work-Related and Personal Factors Associated With Mental Well-Being During the COVID-19 Response: Survey of Health Care and Other Workers
J Med Internet Res 2020;22(8):e21366
URL: http://www.jmir.org/2020/8/e21366/
doi:10.2196/21366
PMID:32763891

Original Paper

Information Disclosure During the COVID-19 Epidemic in China: City-Level Observational Study

Guangyu Hu[1,2], PhD; Peiyi Li[3], MSc; Changzheng Yuan[4], ScD; Chenglin Tao[5], MSc; Hai Wen[1], BS; Qiannan Liu[6], MPH; Wuqi Qiu[1,2], PhD

[1]Institute of Medical Information/Center for Health Policy and Management, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[2]World Health Organization Collaborating Center for Health and Biomedical Information, Beijing, China

[3]Institute of Hospital Management, West China Hospital of Sichuan University, Chengdu, China

[4]School of Public Health, Zhejiang University, Hangzhou, China

[5]Shenzhen Medical Information Center, Shenzhen, China

[6]National Institute of Hospital Administration, Beijing, China

Corresponding Author:
Guangyu Hu, PhD
Institute of Medical Information/Center for Health Policy and Management
Chinese Academy of Medical Sciences and Peking Union Medical College
No 3 Bldg 3th Fl
Yabao Road, Chaoyang District
Beijing, 100020
China
Phone: 86 01052328735
Email: hu.guangyu@imicams.ac.cn

Abstract

Background: Information disclosure is a top priority for official responses to the COVID-19 pandemic. The timely and standardized information published by authorities as a response to the crisis can better inform the public and enable better preparations for the pandemic; however, there is limited evidence of any systematic analyses of the disclosed epidemic information. This in turn has important implications for risk communication.

Objective: This study aimed to describe and compare the officially released content regarding local epidemic situations as well as analyze the characteristics of information disclosure through local communication in major cities in China.

Methods: The 31 capital cities in mainland China were included in this city-level observational study. Data were retrieved from local municipalities and health commission websites as of March 18, 2020. A checklist was employed as a rapid qualitative assessment tool to analyze the information disclosure performance of each city. Descriptive analyses and data visualizations were produced to present and compare the comparative performances of the cities.

Results: In total, 29 of 31 cities (93.5%) established specific COVID-19 webpages to disclose information. Among them, 12 of the city webpages were added to their corresponding municipal websites. A majority of the cities (21/31, 67.7%) published their first cases of infection in a timely manner on the actual day of confirmation. Regarding the information disclosures highlighted on the websites, news updates from local media or press briefings were the most prevalent (28/29, 96.6%), followed by epidemic surveillance (25/29, 86.2%), and advice for the public (25/29, 86.2%). Clarifications of misinformation and frequently asked questions were largely overlooked as only 2 cities provided this valuable information. The median daily update frequency of epidemic surveillance summaries was 1.2 times per day (IQR 1.0-1.3 times), and the majority of these summaries (18/25, 72.0%) also provided detailed information regarding confirmed cases. The reporting of key indicators in the epidemic surveillance summaries, as well as critical facts included in the confirmed case reports, varied substantially between cities. In general, the best performance in terms of timely reporting and the transparency of information disclosures were observed in the municipalities directly administered by the central government compared to the other cities.

Conclusions: Timely and effective efforts to disclose information related to the COVID-19 epidemic have been made in major cities in China. Continued improvements to local authority reporting will contribute to more effective public communication and efficient public health research responses. The development of protocols and the standardization of epidemic message templates—as

well as the use of uniform operating procedures to provide regular information updates—should be prioritized to ensure a coordinated national response.

(J Med Internet Res 2020;22(8):e19572) doi:10.2196/19572

KEYWORDS

information disclosure; COVID-19; website; risk; communication; China; disclosure; pandemic; health information; public health

Introduction

Prompt information disclosure is a top priority for preparedness and it enables a collective response to the COVID-19 pandemic [1]. In light of the important lessons learned from the global response to Ebola, reliable systems for sharing epidemiological and clinical data are essential for the timely production and dissemination of related knowledge [2]. Unfortunately, these systems had not been fully established before and during the COVID-19 outbreak in China. As the World Health Organization (WHO) declared the COVID-19 outbreak a public health emergency of international concern, people have the right to be clearly informed about the health risks that they and their communities face [3]. It is of great importance that authorities tackle the infodemic and improve response capacity, as they face a large amount of misinformation regarding the status of the pandemic. To help the general public and researchers bridge knowledge gaps and respond to the crisis in a timely manner, what information has been disclosed by the authorities was a crucial question. To address this issue, this study aimed to review and analyze the situation as it pertains to China.

Since the severe acute respiratory syndrome (SARS) emergency in 2003 and the avian influenza A (H7N9) epidemic in 2013, the Chinese government's provision of rapid, effective, and efficient disclosure of epidemic information has substantially improved [4,5]. Starting on January 3, 2020, information regarding COVID-19 cases has been reported to the WHO and the general public through China's National Health Commission on a daily basis [6]. The time interval from the first case description to the identification of the pathogen on January 7—as well as the availability of probes for PCR detection on January 21—was much faster for COVID-19 than for SARS [7,8]. The National Health Commission took prompt public health measures and soon classified COVID-19 as a new notifiable disease under the National Infectious Disease Law and the Frontier Health and Quarantine Law on January 20, 2020, thus authorizing by law the prompt disclosure of information about the epidemic at the subnational level [9]. However, the lack of technical norms, standards, and actionable guidance from the national health authorities that could have served as a reference was the main issue thwarting efforts by local authorities to disclose information. Therefore, their main challenge was to determine the type and level of detail of essential information that could be disclosed as part of their COVID-19 epidemic responses.

In this article, we collectively reviewed the information disclosures related to the COVID-19 outbreak of 31 Chinese cities aggregated from online resources up to March 18, 2020. The data were recorded from the official websites of local municipalities or their health commissions. We described and compared the officially released content regarding the local epidemic situations and then analyzed the characteristics of information disclosure through local transmission for both the individual and total population levels in our attempt to present a detailed overview of information disclosure related to the COVID-19 epidemic in China.

Methods

Study Design

In this city-level observational study, our sample included all capital cities of the 31 provinces, autonomous regions, and municipalities in mainland China. There are 4 municipalities administered directly by the central government, 4 capital cities of ethnic minority autonomous regions, and 23 provincial capital cities. Generally, these cities are the socioeconomic and health care centers in their respective regions, and they have the highest population densities and the largest transregional migrant communities, all of which represent potentially higher risks of exposure to COVID-19 than smaller, less-dense neighboring cities.

We developed a checklist for assessing COVID-19 information disclosure, which could be used as a rapid qualitative assessment tool (Multimedia Appendix 1). It includes four sections: local summary report, content covered in the COVID-19 webpages, epidemic surveillance summary and confirmed case reports. We screened the relevant official public COVID-19 information from local municipalities and health commission websites and assessed each situation to allow participants to complete the checklist based on their observations. All authors independently reviewed the COVID-19 webpages of each city following a reviewer training process. The results of the checklists were cross-checked by the members of the research team. The corresponding author reviewed all webpages as well as the results of the final version. Discrepancies were resolved by discussion until consensus was reached. It must be noted that all of this information is publicly available. Patient consent was unnecessary, and no approvals from corresponding ethics boards were required.

Data Sources

We queried the official websites of 31 sample cities to aggregate their information disclosure data related to the COVID-19 epidemic (Multimedia Appendix 2). As of March 18, 2020, there had been a total of 80,928 confirmed cases of COVID-19 on the Chinese mainland from the 31 provinces, autonomous regions, and municipalities. Meanwhile, no new domestically transmitted cases were reported for the first time since the outbreak [10]. This marked an important day that indicated that the intensity of the epidemic was diminishing and the increase in cases had slowed considerably as a result of effective national

containment efforts. Therefore, we observed and reported records from the publicly available sources up to the date of this notable turning point.

Data collection was completed between March 25 and April 8, 2020. The city-level information disclosure records obtained using the checklist were formatted into a line-list database for further analysis, and screenshots of the corresponding webpages were archived for further reference (Multimedia Appendix 3).

Outcome Measures

For the local summary report, the primary outcome was the number of cities that established specific public webpages for COVID-19 information. Cities were categorized as having provided a COVID-19 webpage if the local government or health commission published a webpage specifically for disclosure of information related to the local outbreak. Moreover, the secondary outcome was defined as the time interval from the date of the first confirmed case to the date of the release of this information.

For the content covered by the COVID-19 webpages, the primary outcome was the median number of content categories. We summarized a total of seven categories of prioritized information found on the front pages of the COVID-19 webpages and determined the distribution of these content categories.

For the epidemic surveillance summaries and the confirmed case reports, the primary outcome was the median frequency of daily updates of the epidemic surveillance summaries, and the secondary outcome was the median number of key facts disclosed in the case reports. We determined the specific indicators and related information recorded in the local epidemic situation updates either by March 18 or the latest date with available records.

Statistical Analysis

We summarized the results using descriptive statistics. There are three types of administrative cities in China, each with its own administrative system; therefore we divided the included cities into three groups: the MC group (municipalities administered by the central government), the AC group (capitals of autonomous regions), and the PC group (provincial capitals). Categorical variables were summarized as counts and percentages. Numerical variables were reported as medians and interquartile ranges. No sampling weights were used because this was not a probabilistic sample. The analysis was performed using R (Version 3.4.3; R Foundation for Statistical Computing) and data visualizations were performed using Tableau Desktop (Version 2020.1.2; Tableau Software).

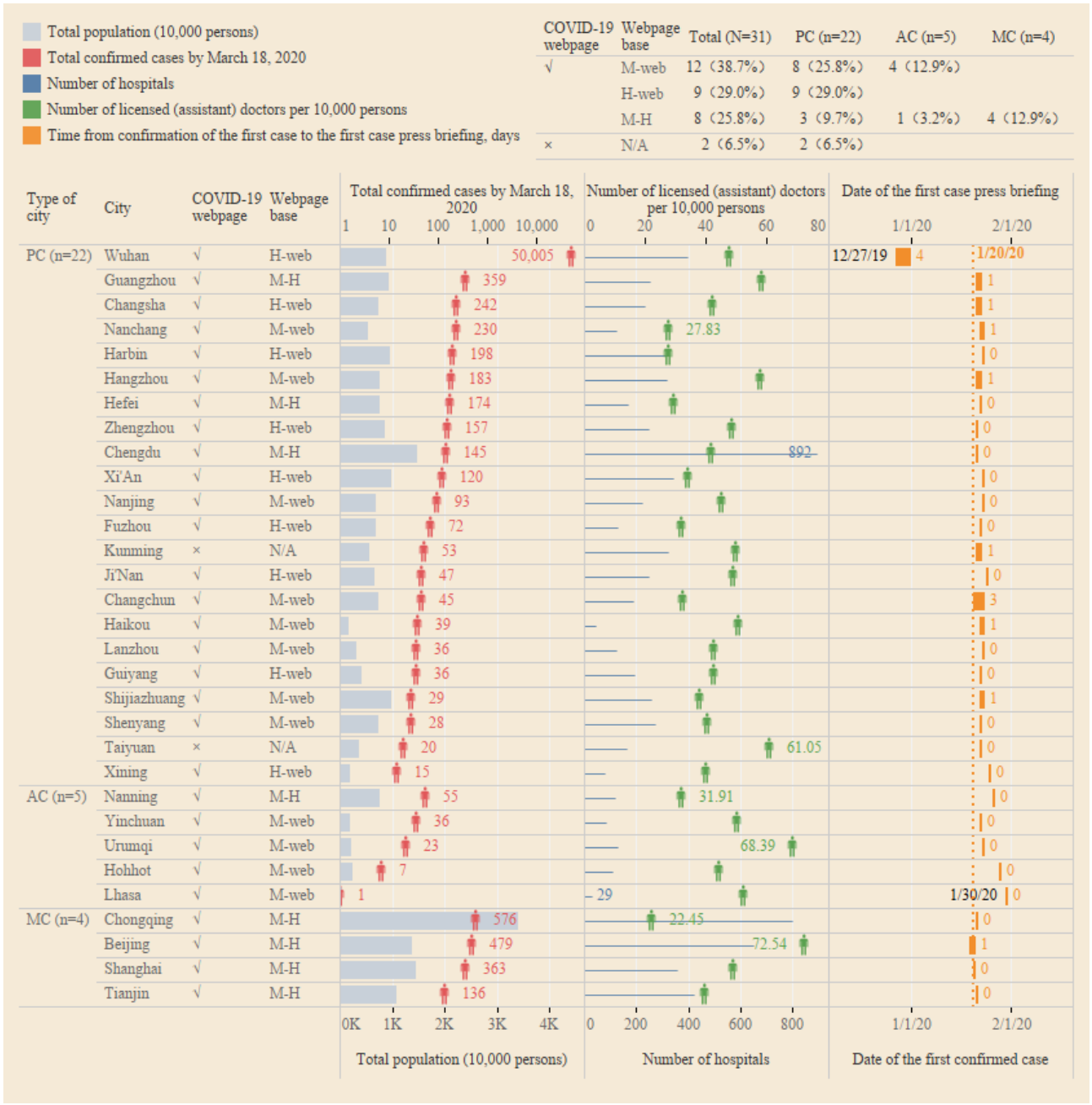
Results

Characteristics and Epidemic Summaries of the Cities

We reviewed the records from the official websites of 31 cities. Of these, 29 (93.5%) created specific COVID-19 webpages to disclose information as follows: 12 webpages were added to the respective municipal websites, 9 were added to health department websites, and 8 were published on both the municipal and health department websites (Figure 1).

As of March 18, 2020, a total of 50,005 COVID-19 cases had been reported in Wuhan, a city in Hubei province, significantly more than all of the other sample cities. The number of licensed (assistant) doctors per 10,000 persons in Wuhan was 47.87, while the median for all 31 cites was 42.62 (IQR 36.18-50.09). The first case with novel coronavirus symptoms in Wuhan was identified and reported to authorities on December 27, 2019. Lhasa reported their first confirmed case on January 31, 2020, making the capital of the autonomous region of Tibet the last city to report a COVID-19 case. The median time from confirmation of the first case to the press briefing that publicly acknowledged that case was 0 days (ie, the same day; IQR 0-1). In fact, the majority of capital cities (21/31, 67.7%) publicly reported their first case of COVID-19 on the day of confirmation (Multimedia Appendix 4).

Figure 1. Characteristics of cities with confirmed COVID-19 cases included in this analysis. Data on total population, number of hospitals, number of licensed (assistant) doctors per 10,000 persons as of 2018 were obtained from the China Statistical Database. Chinese scientists identified the pathogen as a novel coronavirus on January 7, 2020. The first case with symptoms of the novel coronavirus in Wuhan was identified and reported to the authorities on December 27, 2019, by Jixian Zhang, the director of the Department of Respiratory and Critical Care Medicine of Hubei Provincial Hospital of Integrated Chinese and Western Medicine. China's National Health Commission incorporated COVID-19 as a notifiable disease in the National Infectious Disease Law and the Frontier Health and Quarantine Law on January 20, 2020. AC: autonomous region capital; H-Web: health department website; MC: municipality administered by the central government; M-H: both municipality and health department websites; M-Web: municipality website; N/A: not applicable; PC: provincial capital.



Information Disclosure Highlights on Local COVID-19 Webpages

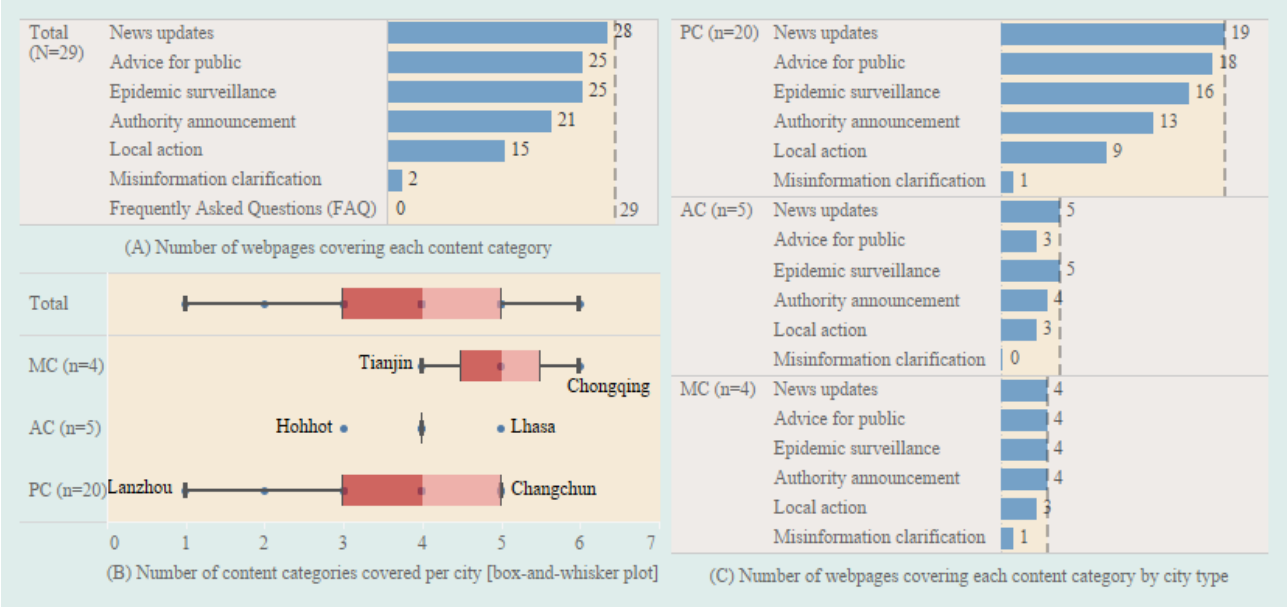
We summarized the information disclosure highlights gathered from the local COVID-19 webpages of 29 capital cities and identified the categories of content covered in these webpages (Figure 2A). News updates were published by almost all of the cities (28/29, 96.6%). This category included the most recent news from local media and press briefings from several government sectors. Of the 29 cities, 25 (86.2%) released epidemic surveillance and advice for the public on their sites. Announcements from authorities were also commonly highlighted (21/29, 72.4%), followed by local actions (15/29, 51.7%) that included the authorities' responses to local sporadic or widespread COVID-19 outbreaks. Only 2 of the 29 cities provided clarifications on previously published misinformation, and none of the cities provided lists of frequently asked questions.

Among the cites that developed COVID-19 webpages, the median number of content categories was 4 (IQR 3-5; Figure 2B). In terms of city type, the median for the MC group was

higher than the other two groups at 5 (IQR 4-6), and the six categories on Chongqing's webpage were the most of any city.

The percentage of content categories published on the COVID-19 webpages varied by city type (Figure 2C and Multimedia Appendix 5). The four municipalities administered by the Chinese central government all released at least four content categories on their official webpages as their responses to emergency information disclosure (Figure 2C). The disclosure performance of the provincial capitals and capitals of autonomous regions was worse compared with that of the centrally administered cities.

Figure 2. Description of content covered in the COVID-19 information webpages, March 2020. (A) Number of webpages covering each content category, summarized and identified from the cities' COVID-19 webpages, as of March 2020. (B) Number of content categories covered per city; distribution was reported using a box-and-whisker plot and the blue dots with black text labels indicate the maximums and minimums. (C) Number of webpages covering each content category by city type, summarized and identified from the cities' COVID-19 webpages, as of March 2020. AC: autonomous region capital; MC: municipality directly administered by the central government; PC: provincial capital.



Key Indicators Derived From the Epidemic Surveillance Summaries

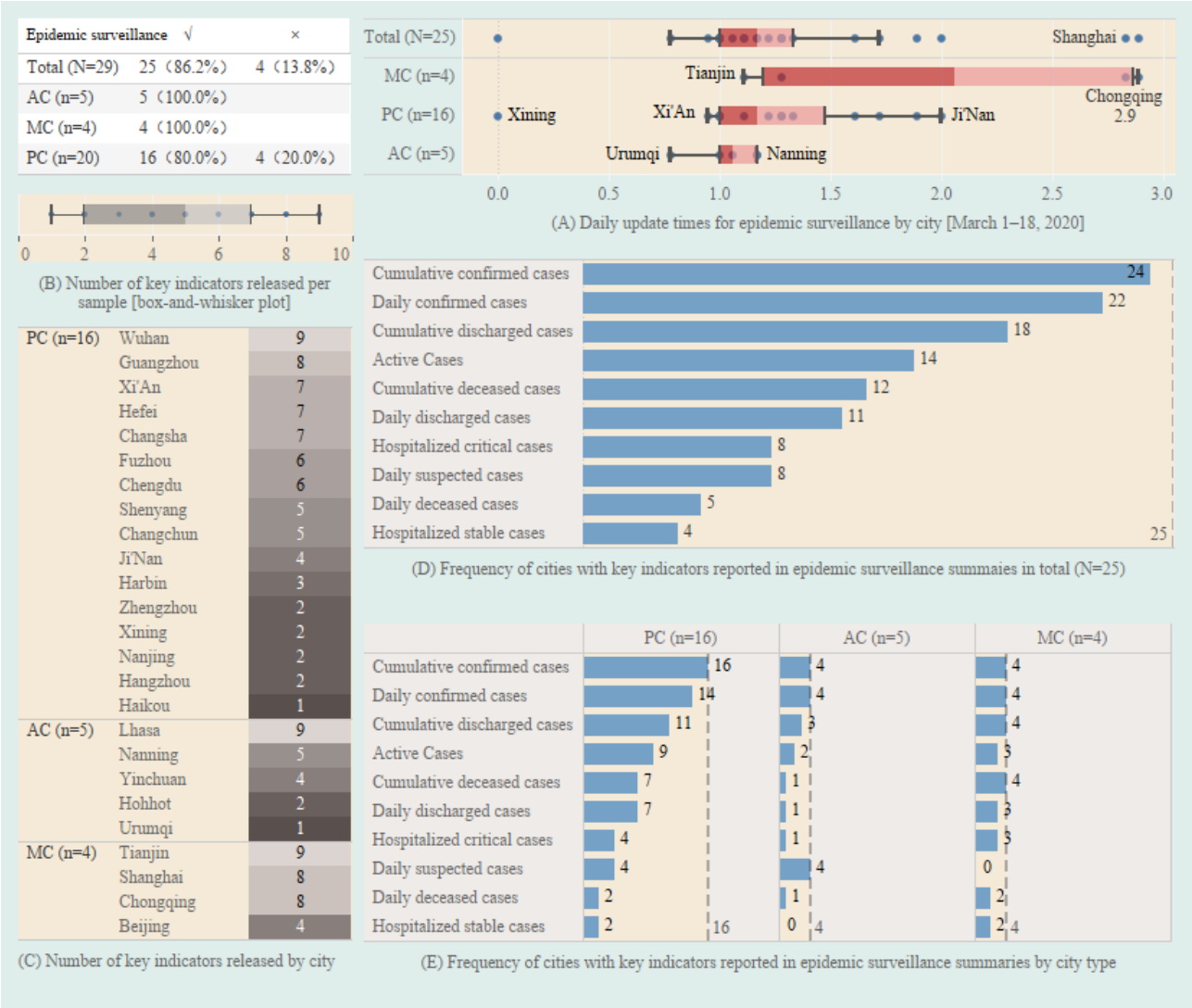
Of the 29 cites with COVID-19 webpages, 25 (86.2%) highlighted epidemic surveillance summaries, while 4 (13.8%) neither disclosed local epidemic surveillance summaries nor provided local bulletins of the epidemic (Figure 3).

During the period from March 1 to 18, we calculated the daily update frequency for epidemic surveillance in each city from the records of the summaries (Figure 3A). The median daily update frequency was 1.2 times in total (IQR 1.0-1.3 times), while the MC group updated 2.1 times per day (IQR 1.2-2.9 times), the PC group 1.2 times (IQR 1.0-1.5 times), and the AC group 1.1 times (IQR 1.0-1.2 times).

We reviewed the local epidemic surveillance summaries of the 25 cities either on March 18, 2020, or the latest date with available records. The median number of key indicators reported in the summaries was 5 (IQR 2-7; Figure 3B). Wuhan, Lhasa, and Tianjin published at least nine key indicators in their summaries, and they were the cities with the most disclosed epidemic indicators in the PC, AC, and MC groups, respectively (Figure 3C).

The most common key indicators reported were the cumulative confirmed cases that appeared in the epidemic surveillance summaries of 24 cities (Figure 3D and Multimedia Appendix 6). Daily confirmed cases were published by 22 cities, cumulative discharged cases by 18, and active cases by 14. Less than half of the cities disclosed cumulative deceased cases (12/25, 48.0%) and daily discharged cases (11/25, 44.0%). Both hospitalized critical cases and daily suspected cases were reported in only 8 cities each (32.0%). Daily deceased cases and hospitalized cases of stable patients were released by 5 and 4 cites, respectively. A comparison of the frequency of key indicators reported by the cities reveals that, on average, the MC group released more indicators in their summaries than the others (Figure 3E).

Figure 3. Key indicators reported in epidemic surveillance summaries as of March 18, 2020. (A) Daily update times for epidemic surveillance by city from March 1 to 18, 2020; distribution was reported using a box-and-whisker plot; blue dots with black text labels indicate maximums and minimums. (B) Number of key indicators released per sample; distribution was reported using a box-and-whisker plot. (C) Number of key indicators released by city; distribution was reported by heatmap. (D) Frequency of cities with key indicators reported in epidemic surveillance summaries in total (N=25). (E) Frequency of cities with key indicators reported in epidemic surveillance summaries by city type. AC: autonomous region capital; MC: municipality directly administered by the central government; PC: provincial capital.



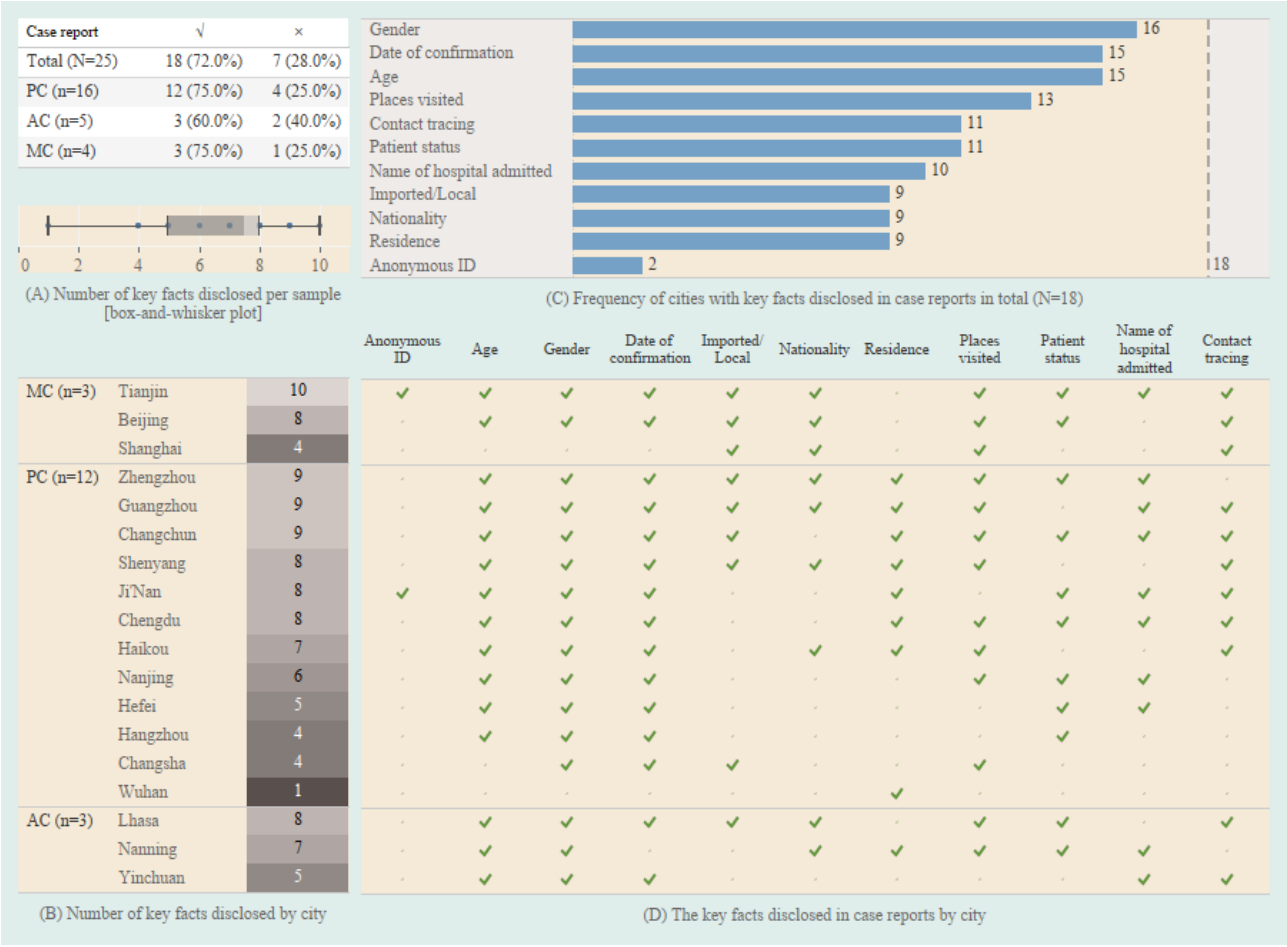
Details of the Confirmed Cases Reports

The majority of cities (18/25, 72.0%) provided detailed epidemic information of each confirmed case along with their epidemic surveillance summaries (Figure 4).

We identified the key facts disclosed in the detailed information about the latest confirmed cases as of March 18, 2020. The median number of key facts disclosed was 7.5 overall (IQR 5.0-8.0; Figure 4A); Tianjin disclosed the highest number of key facts at 10 and Wuhan the lowest at 1 (Figure 4B).

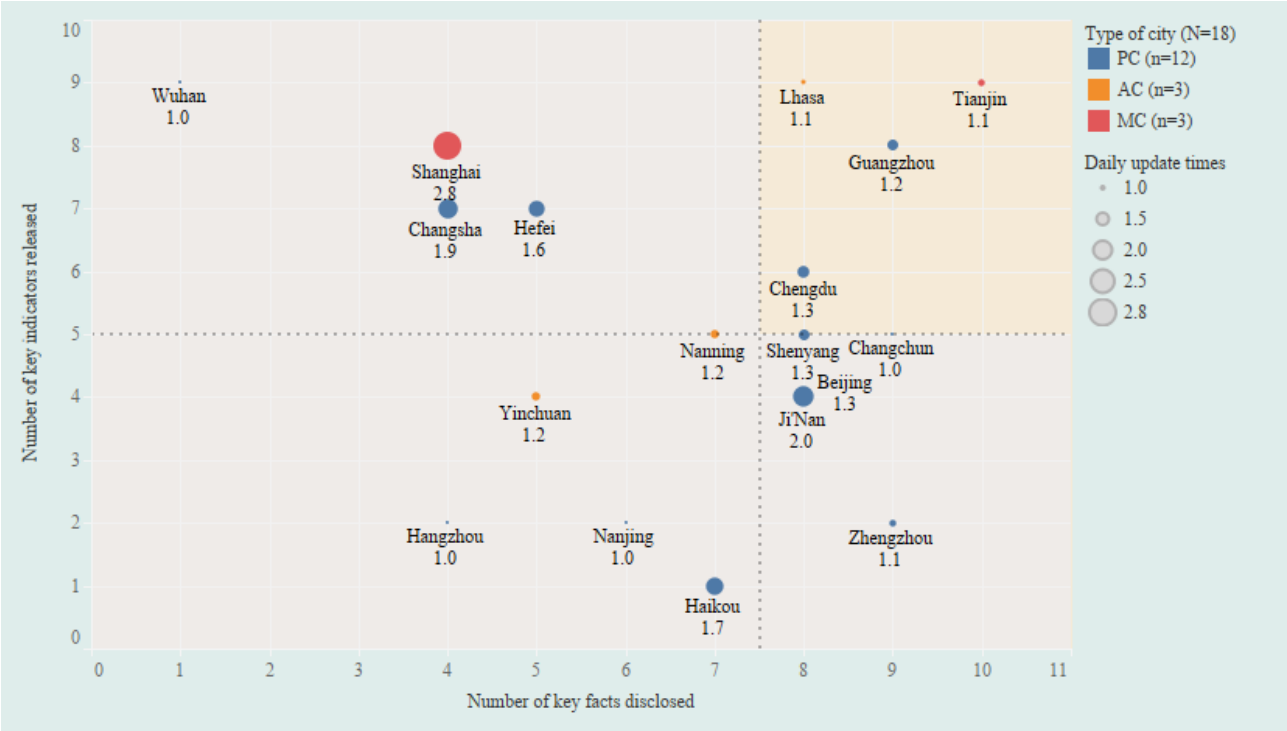
The genders of the patients with COVID-19 were disclosed in the case reports of most cities (16/18, 88.9%; Figure 4C). Dates of confirmation and age were usually included, with each being reported in 15 of the 18 cities (83.3%). Other information reported included the following: places visited (13/18, 72.2%), contact tracing (11/18, 61.1%), patient status (11/18, 61.1%), and the name of the hospital where each patient was admitted (10/18, 55.6%). Half of the cities that reported cases included whether the patient was a local resident or a migrant from somewhere else, their nationality, and their current residence. It is noteworthy that only two cities provided the anonymous ID of the confirmed patients per case (Figure 4D and Multimedia Appendix 7).

Figure 4. Key facts disclosed from the latest confirmed case reports as of March 18, 2020. (A) Number of key facts disclosed per sample; distribution was reported using a box-and-whisker plot. (B) Number of key facts disclosed by city; distribution was reported using a heatmap. (C) Frequency of cities with key facts disclosed in case reports in total (N=18). (D) The key facts disclosed in case reports by city. AC: autonomous region capital; MC: municipality directly administered by the central government; PC: provincial capital.



The 4 cities with both the largest number of disclosed key facts and the largest number of reported key indicators are displayed in the orange-shaded areas of Figure 5. The number of key facts disclosed was similar to the number of key indicators released in these cities. However, a much greater disparity in the facts disclosed compared with the number of indicators released was noted for other cities, such as Wuhan and Zhengzhou.

Figure 5. Scatterplots of the key facts disclosed in the case reports versus key indicators released in the epidemic surveillance summaries of 18 cities. The grey-shaded areas indicate cut-offs between the low number of key facts disclosed (lower than the median of 7.5) and the low number of key indicators released (lower than the median of 5.0). The dots representing Beijing and Ji'Nan coincide on the scatterplots. The name of the city and daily update frequencies for epidemic surveillance are displayed in black lettering.



Discussion

Principal Results

This study presents the first summary of disclosure performance of local authorities related to COVID-19 epidemic information reporting in major mainland China cities as of March 2020. We found that most cities responded proactively and in a timely manner to disclose key information regarding the COVID-19 outbreak by publishing theme-based COVID-19 contents on the websites of local authorities. After the national authority incorporated COVID-19 as a notifiable disease by law on January 20, 2020, almost all the capital cities publicly reported their first case of COVID-19 on the day of confirmation or the following day via the official websites. News updates, epidemic surveillance, and advice for the general public have been the most frequently released contents on these COVID-19 webpages. The rapid and transparent reports published in China surpass the responses of most countries during the current pandemic [11].

We performed an assessment of the content released from the epidemic surveillance summaries and confirmed case reports of each city. There were variations in the key indicators released and key facts disclosed as part of the publicly available information recorded by each city in the study. Given the recent communication regarding COVID-19 risks and the community engagement action plan guidance that was developed and recommended by the WHO [3], we suggest that the significant dissimilarities in message and data templates for compiling epidemic surveillance summaries and confirmed case reports may need to be addressed in some of the cities as they have important implications for information disclosure and risk communication during any pandemic or serious event.

The general public's misconceptions about COVID-19 in the United States and the United Kingdom in the early stages of the regional epidemic highlight the importance of timely and effective information disclosure by public health authorities [12]. This research revealed the efforts being made throughout the major municipalities of China. Regarding the scope of information released—as summarized and predefined in seven content categories—we revealed significant shortcomings in content related to misinformation clarification and frequently asked questions. Based on our review of the webpages archived in Multimedia Appendix 3 of this article, it is also worth noting that the information published on the webpages lacked clear content category labels for some cities (eg, Shijiazhuang and Lanzhou), and the content in some categories were mixed with unrelated material. As a resource that must provide prompt and accurate information, disorganization becomes an inevitable barrier for the dissemination of information to the general public. This finding underscores the need for clearly labeled content categories, proper sorting, and a clear focus on the needs of the audience when disclosing information during an emergency. In addition, the need for compliance with usability principles for information released on these types of websites has been highlighted in previous research published in the United States [13].

The epidemic surveillance summaries and confirmed case reports presented case-by-case in our research have been widely used as important resources for publicly available data in several clinical and epidemiological studies of COVID-19 [14-18]. As a valuable source of COVID-19 epidemiological data reported

in the context of both entire populations and individuals, epidemic surveillance summaries usually include key indicators like COVID-19 cases, deaths, and recoveries, all of which may be used to track epidemic trends. A prime example is one of the most commonly cited web-based interactive dashboards [19], hosted by Johns Hopkins University [15], that has contributed information that supports public health decision making and global communication. In the present study, our results revealed dissimilarities between the key indicators released in the summary reports of different cities. Which kind of indicators should be reported to the public? There is still a lack of consensus on this topic among the municipal governments in China during the epidemic, as we have revealed in the analysis of indicators involved in the epidemic surveillance summaries. Owing to the lack of consistent protocols and standards used to report findings and data related to the epidemic, it is difficult for the general public to identify and interpret the entire scope and magnitude of the health risks they face. Meanwhile, if given properly presented information, professional researchers would have been able to collect and curate data as rapidly and widely as possible rather than compile it manually at higher labor costs because of incompatible record styles. Further, the numbers of daily hospital admissions and discharges, which are less-biased indicators for detecting changes in COVID-19 transmission dynamics [20], were not fully investigated in the epidemic summaries.

Concerning the confirmed case reports, although the machine-accessible, detailed, real-time, and robust individual-level epidemiology data for COVID-19 are publicly available [21], the primary data records used in the official case reports face the same challenge as the epidemic surveillance summaries as presented in our results. Another issue in the public reporting of individual confirmed patients on a case-by-case basis is the overall workload. With the epidemic continuing to expand quickly, this surge can surpass the response capacity of local authorities. At times, they have been unable to disclose information on an individual level as quickly as desired during the emergency. This was most notable in the case of Wuhan. Given that the details disclosed in the confirmed case reports were crucial to the estimation of key epidemiological parameters like incubation periods [17], severity [16], transmission dynamics [18], and prediction models [22], the responsibility of maintaining the integrity and impartiality of each recorded case was more critical than ever. However, upon our assessment of the details in the confirmed case reports, we observed a disparity in the contents and the number of indicators released by each municipality, and it is worth noting that almost all case reports from the local authorities could be improved for better accuracy and completeness to some extent. Moreover, rapid information disclosure and data sharing are necessary for informed public health decision making and subsequent actions during public health emergencies [23]. The balance between public information disclosure and individual privacy concerns must be addressed. Maintaining the confidentiality of patients' names is challenging during a crisis like the COVID-19 pandemic. We found that anonymous identification was rarely used in the majority of confirmed case reports, while the surname, gender, age, and place of residence of confirmed cases were disclosed simultaneously in some cases (eg, Tianjin and Zhengzhou). The risk of personal information leaks in some cities' case reports needs to be addressed, and further effort is needed to ensure the anonymity of individual patients.

Limitations

There are several limitations to this study. First and foremost, capital cities are generally larger and denser, and they are unlikely to be representative of most Chinese cities. Therefore, the generalizability of our findings is limited. Second, the checklist we developed as a rapid qualitative assessment tool is somewhat subjective, and it is possible that some variations of information disclosure across the samples may be either overlooked or underestimated. Additional quantitative research would allow for a more accurate triangulation of the results. Third, local information disclosure performance might vary over time because of the rapidly evolving COVID-19 epidemiologic status in different regions. A retrospective longitudinal case study could further explain the evolution of the authorities' responses during the outbreak. An examination of trust and feedback from the local population regarding the authorized sites would provide direct evidence for information communication improvement of the sites. Fourth, internet social media as another available major public resource was not included in this study. We assumed that the official websites of the cities in question would be the most authoritative sources of disclosed information. However, real-time and rolling COVID-19 updates available through social media channels have the potential to provide more information to the public. A more comprehensive future study could include social media as an alternative data source.

Conclusions

Promoting the disclosure of information related to public health emergencies and providing the public with regular channels through which authoritative up-to-date information is disclosed are both essential for the timely communication of risk information and guidance. Our results augment the awareness of information and data disclosed on a city-by-city basis during the COVID-19 outbreak in China. The local authorities in major Chinese cities universally established COVID-19 webpages on their official websites to ensure the effective disclosure of epidemic information. Nevertheless, further improvements to local reporting practices will further contribute to effective public communication and more efficient public health research. This study offers insight into the deficiencies currently found in local information disclosure methods that were exposed during the COVID-19 epidemic in China. Therefore, the development of uniform protocols and standards of epidemic message templates must be encouraged, as should the use of standard operating procedures to regularly update all vital information in a manner that the public can easily interpret and researchers can effectively analyze. Our findings suggest that these issues should be considered a critical policy priority for the national health authorities in China and most countries worldwide.

Acknowledgments

We greatly appreciate the kind assistance of Shu Yan and Qing Wang from the Chinese Academy of Medical Sciences and Peking Union Medical College during this study. GH had full access to all the data in the study and takes responsibility for the integrity of the data. He is also grateful to his family and friends for their support and concern while he was quarantined in Beijing. The opinions reported in this paper are those of the authors alone and are not the official position of any affiliated institutions.

Authors' Contributions

GH conceived and designed the study. GH, PL, CT, and HW contributed to the data compilation. GH and PL contributed to data analysis and interpretation. GH contributed to design and the creation of figures, and wrote the first draft of the manuscript. CY, QL, and WQ contributed to the critical revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COVID-19 information disclosure checklist for cities where one or more cases have been identified.
[DOCX File , 18 KB - jmir_v22i8e19572_app1.docx]

Multimedia Appendix 2
Data sources for the information disclosure records of the COVID-19 epidemic.
[DOCX File , 16 KB - jmir_v22i8e19572_app2.docx]

Multimedia Appendix 3
Screenshots of the front page of the COVID-19 webpages.
[DOC File , 42169 KB - jmir_v22i8e19572_app3.doc]

Multimedia Appendix 4
Characteristics of cities with confirmed COVID-19 cases included in this analysis.
[DOCX File , 20 KB - jmir_v22i8e19572_app4.docx]

Multimedia Appendix 5
Percentage of cities with content categories covered on official COVID-19 websites, March 2020.
[DOCX File , 14 KB - jmir_v22i8e19572_app5.docx]

Multimedia Appendix 6
Percentage of cities with key indicators revealed in epidemic surveillance summaries, as of March 18, 2020.
[DOCX File , 14 KB - jmir_v22i8e19572_app6.docx]

Multimedia Appendix 7
Percentage of cities with key facts disclosed from the latest confirmed case reports, as of March 18, 2020.
[DOCX File , 14 KB - jmir_v22i8e19572_app7.docx]

References

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. The Lancet 2020 Feb;395(10223):470-473 [FREE Full text] [doi: 10.1016/s0140-6736(20)30185-9]
2. Moon S, Sridhar D, Pate MA, Jha AK, Clinton C, Delaunay S, et al. Will Ebola change the game? Ten essential reforms before the next pandemic. The report of the Harvard-LSHTM Independent Panel on the Global Response to Ebola. Lancet 2015 Nov 28;386(10009):2204-2221. [doi: 10.1016/S0140-6736(15)00946-0] [Medline: 26615326]
3. World Health Organization. Risk Communication and Community Engagement (RCCE) Action Plan Guidance COVID-19 Preparedness and Response. 2020 Mar 16. URL: https://www.who.int/publications-detail/risk-communication-and-community-engagement-(rcce)-action-plan-guidance?accessed [accessed 2020-04-01]
4. Zhong N, Zeng G. What we have learnt from SARS epidemics in China. BMJ 2006 Aug 19;333(7564):389-391 [FREE Full text] [doi: 10.1136/bmj.333.7564.389] [Medline: 16916828]
5. Wang Y. The H7N9 influenza virus in China--changes since SARS. N Engl J Med 2013 Jun 20;368(25):2348-2349. [doi: 10.1056/NEJMp1305311] [Medline: 23782176]

6. World Health Organization. Report of the WHO-China Joint Mission on coronavirus disease 2019 (COVID-19). 2020. URL: https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)?accessed [accessed 2020-03-25]
7. Wilder-Smith A, Chiew CJ, Lee VJ. Can we contain the COVID-19 outbreak with the same measures as for SARS? Lancet Infect Dis 2020 May;20(5):e102-e107 [FREE Full text] [doi: 10.1016/S1473-3099(20)30129-8] [Medline: 32145768]
8. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24:1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
9. China National Health Commission. Inclusion of 2019 novel coronavirus diseases (COVID-19) into statutory infectious disease management. 2020. URL: http://www.nhc.gov.cn/jkj/s7916/202001/44a3b8245e8049d2837a4f27529cd386.shtml [accessed 2020-03-20]
10. China National Health Commission. Update on COVID-19 as of 24on March 18, 2020. 2020. URL: http://www.nhc.gov.cn/xcs/yqtb/202003/e644c2fc18b4448db7ed4b30f68b91a6.shtml?accessed [accessed 2020-03-25]
11. Azman AS, Luquero FJ. From China: hope and lessons for COVID-19 control. Lancet Infect Dis 2020 Jul;20(7):756-757 [FREE Full text] [doi: 10.1016/S1473-3099(20)30264-4] [Medline: 32251637]
12. Geldsetzer P. Use of Rapid Online Surveys to Assess People's Perceptions During Infectious Disease Outbreaks: A Cross-sectional Survey on COVID-19. J Med Internet Res 2020 Apr 02;22(4):e18790 [FREE Full text] [doi: 10.2196/18790] [Medline: 32240094]
13. Devine T, Broderick J, Harris LM, Wu H, Hilfiker SW. Making Quality Health Websites a National Public Health Priority: Toward Quality Standards. J Med Internet Res 2016;18(8):e211 [FREE Full text] [doi: 10.2196/jmir.5999] [Medline: 27485512]
14. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. The Lancet Digital Health 2020 Apr;2(4):e201-e208 [FREE Full text] [doi: 10.1016/s2589-7500(20)30026-1]
15. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
16. Verity R, Okell L, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis 2020 Jun;20(6):669-677 [FREE Full text] [doi: 10.1016/S1473-3099(20)30243-7] [Medline: 32240634]
17. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med 2020 May 05;172(9):577-582 [FREE Full text] [doi: 10.7326/M20-0504] [Medline: 32150748]
18. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. Lancet Infect Dis 2020 Jul;20(7):793-802 [FREE Full text] [doi: 10.1016/S1473-3099(20)30230-9] [Medline: 32247326]
19. Johns Hopkins University and Medicine. Coronavirus resource center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-08-25]
20. García-Basteiro AL, Chaccour C, Guinovart C, Llupià A, Brew J, Trilla A, et al. Monitoring the COVID-19 epidemic in the context of widespread local transmission. Lancet Respir Med 2020 May;8(5):440-442 [FREE Full text] [doi: 10.1016/S2213-2600(20)30162-4] [Medline: 32247325]
21. Xu B, Gutierrez B, Mekaru S, Sewalk K, Goodwin L, Loskill A, et al. Epidemiological data from the COVID-19 outbreak, real-time case information. Sci Data 2020 Mar 24;7(1):106 [FREE Full text] [doi: 10.1038/s41597-020-0448-0] [Medline: 32210236]
22. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
23. Moorthy V, Henao Restrepo AM, Preziosi M, Swaminathan S. Data sharing for novel coronavirus (COVID-19). Bull World Health Organ 2020 Mar 01;98(3):150 [FREE Full text] [doi: 10.2471/BLT.20.251561] [Medline: 32132744]

Abbreviations

AC: capitals of autonomous regions
H7N9: avian influenza A
MC: municipalities administered by the central government
PC: provincial capitals
SARS: severe acute respiratory syndrome

Edited by G Eysenbach; submitted 23.04.20; peer-reviewed by J Gao, MY Song; comments to author 02.06.20; revised version received 13.06.20; accepted 11.08.20; published 27.08.20.

Please cite as:
Hu G, Li P, Yuan C, Tao C, Wen H, Liu Q, Qiu W
Information Disclosure During the COVID-19 Epidemic in China: City-Level Observational Study
J Med Internet Res 2020;22(8):e19572
URL: http://www.jmir.org/2020/8/e19572/
doi:10.2196/19572
PMID:32790640

Viewpoint

COVID-19 Research: Navigating the European General Data Protection Regulation

Regina Becker[1], PhD; Adrian Thorogood[2], BA&Sc, BCL/LLB, LLM; Johan Ordish[3], BA, BA (Law), BA (Hons), MA; Michael J.S. Beauvais[2], BA, BA (Jurisprudence), MSc

[1]ELIXIR-Luxembourg, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg

[2]Centre of Genomics and Policy, McGill University, Montreal, QC, Canada

[3]PHG Foundation, University of Cambridge, Cambridge, United Kingdom

Corresponding Author:
Regina Becker, PhD
ELIXIR-Luxembourg, Luxembourg Centre for Systems Biomedicine (LCSB)
University of Luxembourg
Campus Belval
7, avenue des Hauts-Fourneaux, L-4362
Esch-sur-Alzette
Luxembourg
Phone: 352 46 66 44 61 50
Email: regina.becker@uni.lu

Abstract

Researchers must collaborate globally to rapidly respond to the COVID-19 pandemic. In Europe, the General Data Protection Regulation (GDPR) regulates the processing of personal data, including health data of value to researchers. Even during a pandemic, research still requires a legal basis for the processing of sensitive data, additional justification for its processing, and a basis for any transfer of data outside Europe. The GDPR does provide legal grounds and derogations that can support research addressing a pandemic, if the data processing activities are proportionate to the aim pursued and accompanied by suitable safeguards. During a pandemic, a public interest basis may be more promising for research than a consent basis, given the high standards set out in the GDPR. However, the GDPR leaves many aspects of the public interest basis to be determined by individual Member States, which have not fully or uniformly made use of all options. The consequence is an inconsistent legal patchwork that displays insufficient clarity and impedes joint approaches. The COVID-19 experience provides lessons for national legislatures. Responsiveness to pandemics requires clear and harmonized laws that consider the related practical challenges and support collaborative global research in the public interest.

(J Med Internet Res 2020;22(8):e19799) doi:10.2196/19799

KEYWORDS

COVID-19; GDPR; health research; pandemic; data privacy; data protection; regulation

Introduction

The world continues to wait expectantly for health researchers to develop effective prevention tools, tests, vaccines, and treatments for COVID-19. The collection, analysis, and timely sharing of rich health data is a key component of this unprecedented international research effort [1]. However, data protection laws are not suspended during emergencies such as a pandemic [2]. Many forms of COVID-19 health research involve the processing of personal data, including human health or genetic data. The General Data Protection Regulation (GDPR) [3] regulates the processing of personal data in the European Economic Area (EEA), which includes the 27 European Union Member States as well as Iceland, Liechtenstein, and Norway. (The United Kingdom, following the Brexit transition period, will retain the GDPR in modified form; however, these modifications have not altered the core principles of the GDPR [4,5].) In principle, the GDPR provides a toolset for the processing of personal data in a health crisis, including for health-related research [6]. One of the central reasons for adopting the GDPR and thus replacing its predecessor, the Data Protection Directive [7], was to create a harmonized data protection regime across the EEA [8]. However, a principle weakness of the GDPR is that the interpretation of multiple public health provisions and other health-related provisions is left open to national legislatures. Consequently, countries have

their own solutions and requirements in many aspects where personal data—especially health and genetic data—are processed in these contexts. In particular, research institutions should be aware of the triple authorization they require under the GDPR for international data sharing: a legal basis for processing personal data, legitimation of processing special categories of data (eg, health and genetic data), and a basis for data transfers outside the EEA. Other important considerations for research institutions include respecting the rights of individual data subjects and ensuring appropriate security safeguards. The aim of our analysis is to help research institutions navigate European data protection law within the COVID-19 crisis. We also encourage European and Member State legislators to adopt a more harmonized regulatory framework that supports the needs of public health and research in a pandemic.

A Legal Basis for Processing COVID-19 Data Under Article 6 of the GDPR

A key principle of data protection is that all personal data be processed lawfully. In other words, a research organization must have a legal basis for processing such data. The different legal bases available under the GDPR are listed in Article 6(1). The two bases most suitable for health research are the consent basis and the public interest basis. In the following, we review the strengths and drawbacks of relying on different legal bases for research in a pandemic.

Consent (Article 6[1][a])

At first glance, consent appears to be a straightforward solution for processing personal data in COVID-19 research, as it coincides with the research ethics principle of consent. Consent under the GDPR, however, is seen to be conceptually and operationally different from the informed consent generally required by research ethics [9]. Consent under the GDPR must be freely given, specific, informed, and unambiguous, and it can be withdrawn at any time per articles 4(11) and (7). While this essentially overlaps with requirements in ethics, European data protection authorities have interpreted these four constituent elements of consent more strictly [10].

Where consent is directly sought at the time of data collection in the health care context, patients may be seen as vulnerable people, and the consent may not be valid under the GDPR because of the imbalance between the controller and the data subject [10]. Recent guidance from the European Data Protection Board (EDPB) suggests that consent to noninterventional research obtained by researchers would be legitimate as long as there was no pressure or threat of disadvantage [6]. However, the EDPB does not specify if this would be affected by the severity of illness or by the consent being obtained by the treating physician (paragraphs 21-23) [6]. Health researchers often want to analyze previously collected data as part of routine health care, particularly during a pandemic. The issue of consent regarding the secondary use of health care data for research is challenging. If the data were originally collected only for health care purposes, it would generally be necessary to re-contact the data subjects to obtain their consent to processing of their data for research purposes. Where people are in a critical status of health, which will be the case for most hospitalized patients with COVID-19, obtaining consent may not be practically possible. In addition, health care professionals are overloaded in the situation of a pandemic. It will be difficult for them to find time to provide the necessary information for consent to be valid. Staff approaching patients for consent may also increase their own risk of infection, unless innovative solutions such as electronic consent (eConsent) are implemented.

The requirement that consent be specific is another challenge of this legal basis. Broad consent to pandemic-related research is possible in principle under GDPR Recital 33, which states that "...data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognized ethical standards for scientific research." However, some data protection authorities may not accept such broad purposes as satisfying consent requirements without a subsequent consenting of individual projects. Indeed, the EDPB's guidelines on consent state that when research purposes cannot be fully specified, other ways should be found to ensure the essence of consent is provided (eg, through additional consents for subsequent steps in the research [10]). In contrast, Dara Hallinan [11] argues that broad consent may still be supported in certain fields of research, such as genomics. With this position in mind, it is particularly unfortunate that the EDPB guidelines on COVID-19 and research make no mention that broader consent is suitable for pandemic research [6].

The spirit of the logic that informs data protection authorities' insistence on additional consents for broad consent spreads into the relationship of the purpose limitation principle and processing for research. Provided that processing for research purposes is not incompatible with the initial purpose, data may be further processed for such research with appropriate safeguards (Article 5[1][b]). Where consent is the legal basis relied upon, this widening of the purpose limitation principle for research purposes is key. Without this widening, the secondary use of data for research purposes always requires that consent be obtained from the data subject. However, in case of consent, the data protection authorities may require reconsenting, as stated for example by the UK Information Commissioner's Office [12].

Last but not least, where consent is the legal basis, the data subject retains the right to withdraw their consent at any time. Otherwise, the respect for the data subject's decisional autonomy that finds its articulation in consent would be meaningless. If the consent that serves as the legal basis for the data processing is withdrawn, processing generally must stop and the data must be erased. Processing may only continue for other purposes based on a separate legal basis that was already established. The EDPB has pointed out on multiple occasions that a data controller cannot swap the legal basis after a withdrawal of consent [10]. This position appears to contradict GDPR Article 17(3)(d), which limits the right to erasure where this would seriously impair research or render it impossible. For ongoing research studies, deleting individual data sets may not render the research impossible, although having to repeat analyses could be cumbersome. For completed research studies, however,

the situation is different, as altering the underlying database may undermine scientific reproducibility.

Given the difficulty of obtaining valid consent from patients with COVID-19 and the ambiguity around the consequences of consent withdrawal, an alternative legal basis is desirable in many situations when processing personal data for research during and after the COVID-19 crisis.

Performance of a Task Carried Out in the Public Interest (Article 6[1][e])

As scientific research on COVID-19 aims to benefit society as a whole, using the legal basis of a task performed in the public interest appears to be a natural choice. It is also the choice suggested by the EDPB as more appropriate than consent for research in clinical trials [13] and is one of the potential legal grounds mentioned in the EDPB's guidelines on COVID-19 and research. The availability of the public interest legal basis, however, must be established by Union or Member State law (Article 6[3]). Infectious disease or public health laws may provide the necessary legal basis as a task in the public interest. Where infectious disease legislation does not authorize a public research institution or university to process personal data for pandemic research, they may still be able to rely on their research mission given to them by law. Some countries, such as Finland [14] and Norway [15], have specified in their national legislation that the public interest legal basis may be relied upon where the processing is necessary for scientific purposes.

Overall, the performance of a recognized task in the public interest can also allow other derogations, such as those under Articles 20(3) and 21(6). As such, the public interest legal basis may allow more flexibility than other legal grounds, depending on the conditions of implementation. It is also important to realize that Article 6(1)(e), like any other legal basis outside consent, will require further national legislation implementing Article 9 for processing special categories of data (see below).

Other Legal Grounds in Article 6

Where institutions cannot rely on public interest, legitimate interest is another option, provided that a balancing exercise is performed to ensure that the interest in processing outweighs the privacy interests of the data subject and that the fairness and transparency of the processing is demonstrated (Article 6[1][f]).

Other options apply to the health care context but do not extend to research:

1. Article 6(1)(c) supports processing to fulfill a legal obligation to which the controller is subject. This basis will cover the reporting of cases and accompanying personal data based on infectious disease acts to the relevant authorities as well as processing of data by the health authorities.
2. Article 6(1)(d) authorizes processing in the vital interest of a data subject or another natural person. Vital interest is strictly construed to mean the immediate life interests of a data subject. Relying on this article for research is too speculative: research aims to produce generalizable knowledge and ultimately to benefit public health and society; thus, it cannot be pursued for the vital interest of an individual.

Article 6(1)(d) can be a valid basis in a treatment context and could even provide an option to track people and inform them about a potential infection risk if the disease is life-threatening. In its *Opinion on the notion of legitimate interests of the data controller*, the Article 29 Working Party [16] admits that "vital interest" covers personal data processing to warn people of a potential infection during an epidemic; however, they also warn that it should not be the basis for massive collection or processing of personal data. Along the same lines, this legal basis is not even discussed in the recent statement of the EDPB on location data and contract tracing [17].

Further Processing for Scientific Research

Recital 50 of the GDPR indicates that where further processing is compatible with the original purpose, "...no legal basis separate from that which allowed the collection of the personal data is required." This seems to suggest that data that were not initially collected for research purposes (eg, in the health care context) may also be processed for research based on the original legal basis under which they were collected. For the compatibility of purposes, Article 5(1)(b) suggests that further processing for scientific research "...shall...not be considered to be incompatible with the initial purposes" as long as appropriate safeguards are in place (eg, those specified by Article 89[1]).

The extent to which these provisions can be applied beyond the original controller are being controversially interpreted. The European Data Protection Supervisor states in its *Preliminary Opinion on data protection and scientific research* that Article 5(1)(b) does not give a general authorization for further processing for scientific research [9]. A compatibility test should be performed, although in principle, compatibility can be assumed for both original and subsequent controllers processing data from health care for scientific research as long as appropriate safeguards are in place. By contrast, Edward Dove [18] questions if new controllers may avail themselves of the presumption of compatibility. In addition, an investigation of the Swedish government came to the conclusion that with respect to the continuation of the legal basis, only the transfer to a subsequent controller is covered by the original legal basis; the new controller must find its own valid legal ground [19]. The EDPB remains silent on the subject in their guidelines on COVID-19 and research; however, dedicated guidance on the subject of further processing is expected [9]. Given the unclear situation on the usability of the further processing exemption for collaborative research, the establishment of a legal basis based on Article 5(1)(b) in combination with Recital 50 is precarious.

Legitimation for the Processing of Special Categories of Data (eg, Health and Genetic Data)

Personal data processed for COVID-19 research invariably includes health data. Health and genetic data are considered to be "special categories" of data. Article 9(1) of the GDPR

prohibits processing of special categories of data by default. Research institutions therefore need not only a legal basis but also an additional legitimation under Article 9(2) to process health and genetic data.

Again, consent is a possible option; however, it faces many of the limitations described above. The GDPR requires that consent be both explicit and specific in the case of Article 9(2)(a). As consent for research is traditionally obtained explicitly, the larger challenge is still in the interpretation of the term "specific". Recital 33, however, permits consent to broader areas of scientific research where purposes cannot be fully specified at the time when consent is obtained. Where processing involves special categories of data, however, the EDBP interprets this permission narrowly [10]. This may pose challenges for research on COVID-19, where a broad consent model would be necessary to cover the full range from disease mechanisms to transmission pathways to psychological or socioeconomic consequences of the disease. The information-giving duties of the data controller are increased, as is the way in which consent is recorded [10]. It is therefore useful to examine other options for legitimation.

There are two likely options for processing health and genetic data for research in the COVID-19 context. Article 9(2)(i) foresees processing for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health; therefore, this article can cover processing in the face of an epidemic. Infectious disease legislation and research into infectious diseases in an epidemic can be legitimated based on this paragraph, but only if the law provides an explicit reference to research for public health or in the event of an epidemic. Article 9(2)(j), on the other hand, covers processing that is necessary for scientific research in general, independent of the type of disease. This legitimation must also be based on Union or Member State law. Moreover, suitable and specific measures to safeguard the fundamental rights and the interests of the data subject are required. The GDPR is silent on what is meant by "suitable and specific measures to safeguard" in this context. However, these measures likely link to the requirements of proportionality, data minimization, and data security. Specific measures may include encryption, pseudonymization, minimization of sensitive data processed, training of personnel, and imposition of duties of confidentiality. The cumulative effect of these measures is to reduce the risks of processing sensitive personal data [20,21].

There is great heterogeneity among EEA countries as to whether and how they make use of these GDPR provisions. For example, regarding the implementation of Article 9(2)(j), the United Kingdom [22] and the Netherlands [23] have limited, among other conditions, the application of these provisions to research in the public interest; Sweden requires ethics approval [24]; and Finland has made defined requirements for technical safeguards [14]. The GDPR furthermore confers powers on Member States to pass additional restrictions on the processing of health and genetic data, thus heightening the potential for divergence (Article 9[4]). Some countries have even adopted separate rules for health data on one hand and for genetic data on the other. For example, Ireland has reintroduced explicit consent as a prerequisite (without a specific government declaration in narrow circumstances) [25]. Where health care data are used, professional secrecy rules must also be considered. These conditions follow national or regional health care legislation.

In consequence, a confusing patchwork of heterogeneous provisions for processing personal health and genetic data for pandemic research has been created across Europe. This inharmonious assortment of differing solutions is not helpful where global cross-border sharing and timely solutions are needed. This is also demonstrated by the EDPB's guidelines on COVID-19 and research [6]. The recommendations remain on a generic level and refer to Member State solutions without further discussing the context and requirements of the pandemic. For a problem that Europe confronts in unison, we are required to return to the Member State level to find solutions.

Personal Data Transfer Outside the EEA

In the face of a truly global pandemic, there is a clear need for researchers to collaborate and rapidly share data internationally. Chapter V of the GDPR imposes limitations on the transfer of personal data outside the EEA, aiming to ensure that Europeans' personal data are subject to essentially equivalent levels of protection when sent to other countries [26].

Among the number of instruments that can be used in the context of a pandemic, an adequacy decision granted by the European Commission is the most straightforward, as it allows data exchange under the same conditions as within the EEA (Article 45). The utility of this instrument remains limited, however, as only thirteen jurisdictions have received recognition [27]. Adequacy decisions require careful study, as they sometimes cover only certain sectors. Moreover, the continued validity of adequacy decisions may be imperiled if recipient third countries adopt aggressive data collection and processing practices in response to COVID-19 [28,29].

Another option is to adopt one of the appropriate safeguards for transfer is introduced in Article 46. Legally binding and enforceable instruments between public authorities and bodies may be a satisfactory safeguard, particularly for the exchange between health authorities (Article 46[2][a]). For public research, however, few enforceable instruments currently exist. Institutions may also rely on the standard contractual clauses provided by the European Commission (Article 46[2][c]). A notable problem with these clauses, however, is that US government departments, public universities, and academic health centers cannot consent to dispute resolution in European courts [30]. Alternative contractual clauses or administrative arrangements between public authorities and bodies will require approval by the data protection authority and are not likely to provide ad hoc or fast solutions (Article 46[3][a]).

The pandemic, however, may justify the reliance on derogations for specific situations. Cross-border transfers can be legitimated by explicit consent under Article 49(1)(a); however, the same practical barriers must be overcome. For example, the data subject must be informed about the particular transfers envisaged; thus, it is not possible for a data subject to give blanket consent to any future, unspecified transfers outside the EEA [31]. The EDPB furthermore notes that explicit consent

is only suitable for certain situations, such as for private entities conducting COVID-19 research; however, it provides no further guidance other than its guidelines on COVID-19 and research (paragraph 67) [6]. Transfers necessary for important reasons of public interest, where this public interest is recognized in Union law or national law (Articles 49[1][d] and 49[4]), could provide a more appropriate solution in the acute situation of a pandemic, in particular because cross-border research collaboration is an important aspect in fighting a pandemic. However, the EDPB also cautions that transfers according to this derogation shall "not become the rule in practice," be restricted to "specific situations," and be "strictly necessary" for the purposes for processing [31]. In the specific situation of the COVID-19 outbreak, the EDPB has conceded "that the fight against COVID-19 has been recognized by the EU and most of its Member States as an important public interest," grounding this claim in provisions of EU law interpreted through the prism of national measures by Member States adopted in response to the crisis (paragraphs 62-67) [6]. Time will tell if grounding international transfers in the public interest will be an accepted solution across Europe.

Where none of these options can be applied, researchers who are not acting within a public authority in the exercise of its mandate may at least claim as an immediate measure that the transfer is necessary for purposes of compelling legitimate interest that are not overridden by the interests or rights and freedoms of the data subject (Article 49[1]). Such transfers are subject to myriad restrictions; for instance, the transfer must not be repetitive, only concern a limited number of data subjects, implement suitable safeguards, and be underpinned by proof to the relevant data protection authority that no other option is available (Article 49[1]). This is, at best, a stopgap solution.

The EDPB does recognize the importance of international transfers for pandemic research, albeit perhaps half-heartedly (transfers are "probably" required) [6]. The EDPB also suggests that the existing provisions under the GDPR are sufficient to conduct such transfers. Given that the European research community is still waiting for solutions for international research data sharing to be developed outside the pandemic context, the lack of a clear route remains a serious concern.

Derogation of Data Subject Rights

An additional hurdle is created through the strict information-providing obligations research institutions have toward data subjects under the GDPR. Extensive information on the use of data must be provided to the data subject before collection, including information about the legal basis for processing and the intention to transfer personal data to third countries (Article 13). There are only narrow exceptions to the requirement that direct information be given to the data subject. Exceptions are foreseen in cases where data are not obtained directly from the data subject (Article 14). COVID-19 research involving secondary use of data collected during the care of seriously ill patients may fall under this exception. Where informing the subject of such research would be impossible or would involve disproportionate effort, general information may alternatively be shared on the controller's web page or in public announcements. The research must be subject to the appropriate safeguards described in Article 89. However, even for the case of secondary use of data collected in health care, the original controller (ie, the health care provider) would still be obligated to inform the data subject about the data sharing for further processing in the research context.

The critiques about the implementation of these requirements are the same as above: providing additional information in the clinical context during an emergency, with many patients in a serious state of illness and experiencing breathing problems, may not be feasible. As no derogation is foreseen, it is not clear how the data protection authorities will judge any retrospective information provided to the patients who survive their disease. A way out could be based on provisions allowing the restriction of all data subjects' rights in matters of important public interest, such as public health (Article 23). Again, national or European laws must provide the necessary framework. Only a few Member States have implemented such derogations in their national data protection legislation; many of these are superficial, stating the possibility of derogations for public health, among others [23,32]. No detailed provisions have been made specific to related research aspects for the preservation of public health. The resulting legal uncertainty hampers responsive research during a pandemic if appropriate information cannot be provided before starting the research.

Regulations in a State of Emergency

Many countries in the EEA provide for the possibility that in emergency situations, the government can enact laws to cope with a crisis such as a pandemic [33-36]. These provisions can restrict the rights of citizens, including their data protection rights, as foreseen in the GDPR itself (Article 23). In a pandemic, such regulations can derogate from data subjects' rights and provide a legal basis for processing beyond the existing legal framework. These emergency powers do not amount to *carte blanche*; the measures must be necessary, appropriate, and proportionate to the aim pursued. The EDPB also points out that measures implemented based on emergency situations should be strictly limited to the duration of the emergency [37]. Moreover, to plausibly rely on such powers for research, there must be a close, proximate connection between the research conducted and the pandemic response. Such measures are also subject to oversight by national constitutional or administrative courts, as well as the Court of Justice of the European Union and the European Court of Human Rights.

Debates around emergency derogations in the context of data processing have thus far focused largely on location tracking or contact tracing of people through mobile phones rather than on health research. Such measures can be used to identify people at risk and to monitor adherence to social distancing. The EDPB has issued a letter to the European Commission stating that an enactment of national laws would be a good way to provide a solid legal framework to define the scope and also the limited duration of the use of mobile phone information, disabling of tracking systems, and deleting of data once the crisis is over [38]. At the same time, the EDPB insists that the use of an app

should be voluntary and not made a mandatory measure or involve any disadvantages for those not opting into the use of the app [38]. The rights of the data subjects should not be compromised but should rather be upheld to retain the trust and buy-in of the citizens. Along these lines, a joint statement has been signed by more than 550 researchers worldwide calling for privacy preserving technologies, strict limitation to COVID-19 purposes, and a voluntary basis for the use of any contact tracing app [39].

Our discussion below brings attention to the need for similar emergency derogations for health research during the pandemic. The two debates are connected: information collected as part of tracking and contact tracing could support research into the course of the pandemic and the mechanisms of virus transmission.

An example of emergency legislation with specific provisions for research was introduced in Italy, where the obligation for prior consultation of the data protection authority where no consent can be obtained can be waived under certain circumstances [40]. This requirement is arguably disproportionate even in normal circumstances; however, even in the present state of the COVID-19 crisis, it is only applicable to clinical trials, observational medicinal product studies, and compassionate therapeutic use that is essential for combating the disease as well as for COVID-19 research projects of the Institutes of Scientific Hospitalization and Treatment funded by the Ministry of Health. Also, derogations from GDPR Article 13 have been made based on Article 23(1)(e). The Italian waiver demonstrates how the current pandemic exposes weaknesses in data protection legislation for research and how emergency regulations can be used as a quick, albeit temporary, remedy.

Additional Obligations Under Data Protection Law

The general obligations of controllers to process personal data apply to pandemic research. Article 5 comprises the data protection principles of lawfulness, fairness, transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity, and confidentiality, as well as accountability of the controller. Some of these aspects have been discussed above. The implementation of data protection by design and default is largely described in Chapter IV of the GDPR. Technical and organizational measures must be taken to ensure that the rights and freedoms of data subjects are not unduly compromised. Before they process special categories of data in the COVID-19 context, research institutions will most likely need to perform data protection impact assessments (DPIAs), which act as both a safeguard and a mechanism for enhancing transparency and accountability. Researchers can find guidance regarding the specific rules in their country from dedicated lists of processing requiring a DPIA on the webpage of their data protection authority. A DPIA will need to include further documentation on the processing and the safeguards introduced as well as a subsequent risk analysis covering the potential impact on the rights and freedoms of the data subjects. The French supervisory authority provides a useful guide and associated tool in English and other European languages [41].

In any case, the GDPR requires that security safeguards be proportionate to the risks of processing (Article 32). Research institutions must therefore adopt stricter security safeguards for processing sensitive health and genetic data. Some countries have introduced specific legislative provisions for security safeguards that the research institutions need to be aware of, including technical measures when processing personal data for scientific research (eg, Luxembourg [42]) and when processing health or genetic data (eg, Ireland [25]). The requested safeguards differ between countries.

Conclusion

The GDPR has foreseen mechanisms that enable research in a pandemic, which include processing sensitive data of vulnerable people, the need for fast action, and global data sharing. However, these mechanisms depend largely on national implementation. Our analysis demonstrates that variation across national implementations hampers a coordinated global research response in the fight against COVID-19. This is also reflected by the request for a mandate by the EDPB to develop guidelines on short notice [43]. The subsequently published guidelines on COVID-19 and research [6] contain only limited advice that takes into account specific challenges in the research contexts (eg, hospitalized people who are unable to give consent, challenges in information-giving obligations). They largely refer to the principal framework and the possibility of measures based on Member State implementation of the GDPR, which remain incomplete and heterogeneous. Research institutions must investigate their scope of action and requirements on a national level and, when in doubt, contact the competent data protection authorities. Without legal clarity from European and national legislators, certain research institutions may need to lead the response based on their best interpretations of current laws.

As lessons learned from this crisis, the EEA countries should review and adapt their legislation, making particular use of enabling articles such as those relating to processing health data for reasons of the public interest in public health (Article 9[2][i]) and derogations to data subject rights and controller obligations based on public interest (Article 23). To date, few countries have used these for the benefit of research to contribute to public health. This indicates that many legislators and data protection authorities still lack deep understanding of the needs of health research and the reason for the privileged status of scientific research in the GDPR. Learning from the COVID-19 crisis, they should develop an operational framework that is appropriate to the needs of global research in a pandemic and provides legal certainty for researchers to act upon. In establishing national frameworks, legal interoperability should be a key consideration during the legislative process to allow all researchers to participate in joint research efforts under compatible conditions and to efficiently set up and manage cross-border cohorts. The EDPB could act as a coordinator or convener of such a process. This will allow researchers to combat not only future pandemics but also other pressing public health priorities.

Acknowledgments

The authors would like to thank Irith Kist at the Netherlands Cancer Institute and Katrina Bramstedt from the Luxembourg Agency for Research Integrity for their review and helpful comments. AT and MB acknowledge funding support from Genome Canada, Genome Quebec, and the Canadian Institutes of Health Research.

Conflicts of Interest

None declared.

References

1. Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak. Wellcome. URL: https://wellcome.ac.uk/coronavirus-covid-19/open-data [accessed 2020-04-03]
2. Pierucci A, Walter JP. Joint Statement on the right to data protection in the context of the COVID-19 pandemic. Council of Europe. 2020 Mar 30. URL: https://www.coe.int/en/web/data-protection/statement-by-alessandra-pierucci-and-jean-philippe-walter [accessed 2020-06-26]
3. Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Publications Office of the European Union. 2016 Apr 27. URL: https://op.europa.eu/en/publication-detail/-/publication/3e485e15-11bd-11e6-ba9a-01aa75ed71a1/language-en [accessed 2020-08-20]
4. European Union (Withdrawal) Act 2018. legislation.gov.uk. 2018 Jun 28. URL: https://www.legislation.gov.uk/ukpga/2018/16/contents/enacted [accessed 2020-08-20]
5. The Data Protection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019. legislation.gov.uk. 2019. URL: https://www.legislation.gov.uk/uksi/2019/419/contents/made [accessed 2020-08-20]
6. European Data Protection Board. Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak. 2020 Apr 21. URL: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202003_healthdatascientificresearchcovid19_en.pdf [accessed 2020-06-26]
7. Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. EUR-Lex. 1995 Nov 23. URL: https://eur-lex.europa.eu/eli/dir/1995/46/oj [accessed 2020-08-20]
8. Kuner C, Bygrave L, Docksey C. Background and Evolution of the EU General Data Protection Regulation. In: Kuner C, Bygrave LA, Docksey C, editors. The EU General Data Protection Regulation (GDPR): A Commentary. New York, NY: Oxford University Press; 2020:1-47.
9. European Data Protection Supervisor. A Preliminary Opinion on data protection and scientific research. 2020 Jan. URL: https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf [accessed 2020-06-26]
10. European Data Protection Board. Guidelines 05/2020 on consent under Regulation 2016/679 (Version 1.1). 2020 May 04. URL: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf [accessed 2020-08-20]
11. Hallinan D. Broad consent under the GDPR: an optimistic perspective on a bright future. Life Sci Soc Policy 2020 Jan 06;16(1):1 [FREE Full text] [doi: 10.1186/s40504-019-0096-3] [Medline: 31903508]
12. Guide to the General Data Protection Regulation (GDPR). Information Commissioner's Office (ICO). 2020 Apr 25. URL: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/ [accessed 2020-06-26]
13. European Data Protection Board. Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR) (art. 70.1.b)). 2019 Jan 23. URL: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_opinionctrq_a_final_en.pdf [accessed 2020-06-26]
14. Tietosuojalaki 1050/2018. Webpage in Finnish. Finlex. 2018 Dec 04. URL: https://www.finlex.fi/fi/laki/alkup/2018/20181050 [accessed 2020-06-26]
15. Lov om behandling av personopplysninger (personopplysningsloven). Wepage in Norwegian. Lovdata. 2018 Jul 20. URL: https://lovdata.no/dokument/NL/lov/2018-06-15-38 [accessed 2020-06-26]
16. Article 29 Data Protection Working Party. Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC. 2014 Apr. URL: https://ec.europa.eu/justice/article-29/press-material/public-consultation/notion-legitimate-interests/files/20141126_overview_relating_to_consultation_on_opinion_legitimate_interest_.pdf [accessed 2020-06-26]
17. European Data Protection Board. Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak. 2020 Apr 21. URL: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_20200420_contact_tracing_covid_with_annex_en.pdf [accessed 2020-06-26]
18. Dove ES. The EU General Data Protection Regulation: Implications for International Scientific Research in the Digital Era. J Law Med Ethics 2019 Jan 10;46(4):1013-1030. [doi: 10.1177/1073110518822003]

19. Personuppgiftsbehandling för forskningsändamål SOU 2017. Webpage in Swedish. Regeringskansliet. 2019 Jun 09. URL: https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2017/06/sou-201750/ [accessed 2020-06-26]
20. Georgieva L, Docksey C. Processing of special categories of personal data. In: Kuner C, Bygrave LA, Docksey C, editors. The EU General Data Protection Regulation (GDPR): A Commentary. New York, NY: Oxford University Press; 2020:365-384.
21. Article 29 Data Protection Working Party. Advice paper on special categories of data ("sensitive data"). 2011 Apr 20. URL: https://ec.europa.eu/justice/article-29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf [accessed 2020-06-26]
22. Data Protection Act 2018. legislation.gov.uk. 2018. URL: https://www.legislation.gov.uk/ukpga/2018/12/contents [accessed 2020-06-26]
23. Uitvoeringswet Algemene verordening gegevensbescherming. Webpage in Dutch. Overheid.nl. 2019 Feb 19. URL: https://wetten.overheid.nl/BWBR0040940/2019-02-19 [accessed 2020-06-26]
24. Lag (2003:460) om etikprövning av forskning som avser människor. Webpage in Swedish. Regeringskansliet. 2003 Jun 05. URL: http://rkrattsbaser.gov.se/sfst?bet=2003:460 [accessed 2020-06-26]
25. Data Protection Act 2018 (Section 36(2)) (Health Research) Regulations 2018. 2018 Aug 10. URL: http://www.irishstatutebook.ie/eli/2018/si/314/made/en/pdf [accessed 2020-06-26]
26. Case C-362/14 Maximillian Schrems v Data Protection Commissioner. InfoCuria. 2015 Oct 06. URL: https://tinyurl.com/yyseln4l [accessed 2020-06-26]
27. Adequacy decisions. European Commission. URL: https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en [accessed 2020-05-07]
28. The Coronavirus Crisis and EU Adequacy Decisions for Data Transfers. European Law Blog. 2020 Apr 03. URL: https://europeanlawblog.eu/2020/04/03/the-coronavirus-crisis-and-eu-adequacy-decisions-for-data-transfers/ [accessed 2020-06-26]
29. Wiewiórowski W. EU Digital Solidarity: a call for a pan-European approach against the pandemic. European Data Protection Supervisor. 2020 Apr 06. URL: https://edps.europa.eu/sites/edp/files/publication/2020-04-06_eu_digital_solidarity_covid19_en.pdf [accessed 2020-06-26]
30. Peloquin D, DiMaio M, Bierer B, Barnes M. Disruptive and avoidable: GDPR challenges to secondary research uses of data. Eur J Hum Genet 2020 Jun;28(6):697-705. [doi: 10.1038/s41431-020-0596-x] [Medline: 32123329]
31. European Data Protection Board. Guidelines 02/2018 on derogations of Article 49 under Regulation 2016/679. 2018 May 05. URL: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_2_2018_derogations_en.pdf [accessed 2020-06-26]
32. Act on supplementary provisions to the regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (the Data Protection Act). 2018 May 23. URL: https://www.datatilsynet.dk/media/7753/danish-data-protection-act.pdf [accessed 2020-06-26]
33. Constituzione della Repubblica italiana. Senato della Repubblica italiana. 1948 Jan. URL: http://www.senato.it/1024 [accessed 2020-06-26]
34. Constitution de la République française. Assemblée Nationale. 1958 Oct 04. URL: http://www.assemblee-nationale.fr/connaissance/constitution.asp [accessed 2020-06-26]
35. Constitución española. 1978 Dec 06. URL: https://app.congreso.es/consti/constitucion/indice/index.htm [accessed 2020-06-26]
36. Constitution du Grand-Duché du Luxembourg. Mémorial A 406 du 15 mai 2020 Oct 17, 1868. Journal officiel du Grand-Duché de Luxembourg. 2020 May 19. URL: http://legilux.public.lu/eli/etat/leg/recueil/constitution [accessed 2020-06-26]
37. European Data Protection Board. Statement on the processing of personal data in the context of the COVID-19 outbreak. 2020 Mar 19. URL: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_statement_2020_processingpersonaldataandcovid-19_en.pdf [accessed 2020-06-26]
38. European Data Protection Board. EDPB Letter concerning the European Commission's draft Guidance on apps supporting the fight against the COVID-19 pandemic. 2020 Mar 19. URL: https://edpb.europa.eu/sites/edpb/files/files/file1/edpbletterecadvisecodiv-appguidance_final.pdf [accessed 2020-06-26]
39. Contact Tracing Joint Statement. 2020 Mar 07. URL: https://www.esat.kuleuven.be/cosic/sites/contact-tracing-joint-statement/ [accessed 2020-06-26]
40. Raccolta delle principali disposizioni adottate in relazione allo stato di emergenza epidemiologica da Covid-19 aventi implicazioni in materia di protezione dei dati personali. Garante per la protezione dei dati personali. 2020 Jun 25. URL: https://tinyurl.com/y7499zyg [accessed 2020-06-26]
41. Privacy Impact Assessment. Commission Nationale de l'Informatique et des Libertés. URL: https://www.cnil.fr/en/privacy-impact-assessment-pia [accessed 2020-06-26]
42. Loi du 1er août 2018 portant organisation de la Commission nationale pour la protection des données et du régime général sur la protection des données. Journal officiel du Grand-Duché de Luxembourg. 2018 Aug 01. URL: https://tinyurl.com/y442zgzz [accessed 2020-08-21]

43. European Data Protection Board; 2020 Apr. Mandate on the processing of health data for research purposes in the context of the COVID-19 outbreak. URL: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_20200407_rfm_healthdataresearchcovid19.pdf [accessed 2020-06-26]

Abbreviations

DPIA: data protection impact assessment
eConsent: electronic consent
EDPB: European Data Protection Board
EEA: European Economic Area
GDPR: General Data Protection Regulation

Edited by G Eysenbach; submitted 01.05.20; peer-reviewed by D Hallinan, D Peloquin; comments to author 12.06.20; revised version received 30.06.20; accepted 22.07.20; published 27.08.20.

Please cite as:
Becker R, Thorogood A, Ordish J, Beauvais MJ
COVID-19 Research: Navigating the European General Data Protection Regulation
J Med Internet Res 2020;22(8):e19799
URL: http://www.jmir.org/2020/8/e19799/
doi:10.2196/19799
PMID:32784191

Original Paper

Motivations for Social Distancing and App Use as Complementary Measures to Combat the COVID-19 Pandemic: Quantitative Survey Study

Kai Kaspar[1], Prof Dr Dr

Department of Psychology, University of Cologne, Cologne, Germany

Corresponding Author:
Kai Kaspar, Prof Dr Dr
Department of Psychology
University of Cologne
Richard-Strauss-Str. 2
Cologne, 50931
Germany
Phone: 49 221 470 2347
Email: kkaspar@uni-koeln.de

Abstract

Background: The current COVID-19 pandemic is showing negative effects on human health as well as on social and economic life. It is a critical and challenging task to revive public life while minimizing the risk of infection. Reducing interactions between people by social distancing is an effective and prevalent measure to reduce the risk of infection and spread of the virus within a community. Current developments in several countries show that this measure can be technologically accompanied by mobile apps; meanwhile, privacy concerns are being intensively discussed.

Objective: The aim of this study was to examine central cognitive variables that may constitute people's motivations for social distancing, using an app, and providing health-related data requested by two apps that differ in their direct utility for the individual user. The results may increase our understanding of people's concerns and convictions, which can then be specifically addressed by public-oriented communication strategies and appropriate political decisions.

Methods: This study refers to the protection motivation theory, which is adaptable to both health-related and technology-related motivations. The concept of social trust was added. The quantitative survey included answers from 406 German-speaking participants who provided assessments of data security issues, trust components, and the processes of threat and coping appraisal related to the prevention of SARS-CoV-2 infection by social distancing. With respect to apps, one central focus was on the difference between a contact tracing app and a data donation app.

Results: Multiple regression analyses showed that the present model could explain 55% of the interindividual variance in the participants' motivation for social distancing, 46% for using a contact tracing app, 42% for providing their own infection status to a contact tracing app, and 34% for using a data donation app. Several cognitive components of threat and coping appraisal were related to motivation measurements. Trust in other people's social distancing behavior and general trust in official app providers also played important roles; however, the participants' age and gender did not. Motivations for using and accepting a contact tracing app were higher than those for using and accepting a data donation app.

Conclusions: This study revealed some important cognitive factors that constitute people's motivation for social distancing and using apps to combat the COVID-19 pandemic. Concrete implications for future research, public-oriented communication strategies, and appropriate political decisions were identified and are discussed.

(J Med Internet Res 2020;22(8):e21613) doi:10.2196/21613

KEYWORDS

COVID-19; protection motivation theory; social distancing; contact tracing app; data donation app; social trust; data security

Introduction

Background

The World Health Organization has declared the outbreak of COVID-19 to be a global pandemic [1]. Development of therapeutics and vaccines started early and remains a high priority [2,3]; however, no effective vaccine or drug treatment is currently available [4]. Moreover, negative social and economic consequences of broader shutdowns in many countries are already visible [5]; therefore, measures are being taken to revive social and economic life. The most critical and challenging task is to revive public life while minimizing the risk of infection. In addition to hand hygiene and mask-wearing [6], reducing interactions between people by social distancing is an effective and prevalent public health measure to reduce the risk of infection and spread of the virus within a community [7]. Current developments in several countries show that this measure may be technologically accompanied by mobile apps. Indeed, app stores already offer a wide range of apps related to the current pandemic; meanwhile, privacy concerns are being intensively discussed [8]. Hence, current research is focusing on ethical aspects of contact tracing apps [9,10] and ethical guidelines for such apps have already been formulated [11]. At the same time, an increasing number of national governments are disseminating contact tracing apps to help contain the pandemic; these apps differ remarkably regarding their technological approaches [12]. Importantly, the utility of mobile apps has already been examined in the context of previous epidemics, such as the Ebola epidemic in West Africa between 2014 and 2016 [13,14]. Given the current relevance of social distancing and using mobile apps as a complementary measure to combat the COVID-19 pandemic, in this study, I aimed to examine central cognitive variables that may constitute people's motivation for social distancing, using an app, and providing health-related data requested by two apps that differ in their direct utility for the individual user. The results would help increase our understanding of people's concerns and convictions, which can then be specifically addressed by public-oriented communication strategies and appropriate political decisions.

One of the most prominent technological concepts during the pandemic is contact tracing, such as the app concept developed by the Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT) initiative. In general, the Bluetooth connection of a smartphone is used "in order to detect whether two people have come into close enough physical proximity to risk an infection [15]." The app notifies users when they have had critical contact (in terms of time span and spatial proximity) with a person infected with SARS-CoV-2, the virus that causes COVID-19, so that the users can take appropriate measures. Importantly, each individual app user voluntarily provides their infection status and should not be identifiable by other app users because a critical contact is signaled with a temporal delay.

A second app type is offered by the Robert Koch Institute, which is the central institution of the German Federal Government in the field of disease surveillance and prevention. The app is officially called the Corona Data Donation app (Data Donation app) and it is connected to the individual user's digital wearables (eg, smartwatches and fitness trackers); it continuously tracks data related to the user's health and daily activities (eg, heart rate and sleep rhythm) and additional personal data, including postal code, weight, age, and gender. According to the official app description, identification of app users is not possible. However, the individual user receives no direct benefit from using this app type, as it does not provide any feedback. Instead, all data are collected centrally by the Robert Koch Institute to create different maps that may indicate and facilitate specific local measures. Importantly, using the Data Donation app and providing the personal data requested by this app are basically the same. In contrast, using a contact tracing app and voluntarily providing one's own infection status to that app are independent measures. It appears to be important to focus on these very different functional accounts to examine whether users' evaluation and use motivation are specific or general.

The Protection Motivation Theory and Social Distancing

The present study refers to the protection motivation theory (PMT). This theory was originally developed on the basis of expectancy-value approaches in the context of health sciences to explain preventive behavior based on threat and coping appraisal processes [16]. Meta-analyses [17,18] showed that the components of the theory reliably explain protection motivation. More recent studies supported these findings in several contexts that are not limited to health-related issues, such as skin cancer prevention [19] or people's intention to receive a seasonal influenza vaccination [20]. For example, Tsai et al [21] successfully applied the PMT to internet users' motivation for enacting safety precautions, Marett et al [22] used it to explain adaptive and maladaptive responses to risks associated with posting personal information on social networking sites, and Vance et al [23] found that most components of the PMT were significantly related to employees' intention to comply with information security policies. Moreover, a meta-analysis [24] revealed that the PMT is particularly effective in the context of information security behavior if the behavior is voluntary and specific and the potential security threat is directed to the individual instead of other people or the person's organization. These conditions are met with respect to the use of a contact tracing app and the Data Donation app. Thus, the PMT is a very powerful and flexible theory that is adaptable to both health-related and technology-related motivations, which qualifies it for the present study.

In general, the PMT comprises threat appraisal of the potential risk (eg, infection with SARS-CoV-2) and coping appraisal of the recommended preventive behavior (eg, social distancing). Threat appraisal includes the perceived severity of and vulnerability to the negative consequences of maladaptive behavior. The more pronounced these two variables, the higher the motivation to perform the recommended behavior. Threat appraisal also includes the perceived rewards associated with not performing the recommended behavior, counteracting protection motivation. Coping appraisal includes the perceived self-efficacy and response efficacy of the recommended behavior, which positively affect an individual's intention to actively prevent risks. Coping appraisal also includes the

perceived response costs, which counteract protection motivation. With respect to the prevention of SARS-CoV-2 infection by social distancing, the following hypotheses resulted:

H1a: The perceived severity of an infection is positively related to the motivation for social distancing.

H1b: The perceived vulnerability to an infection is positively related to the motivation for social distancing.

H1c: The perceived rewards associated with avoiding social distancing are negatively related to the motivation for social distancing.

H2a: The self-efficacy regarding social distancing is positively related to the motivation for social distancing.

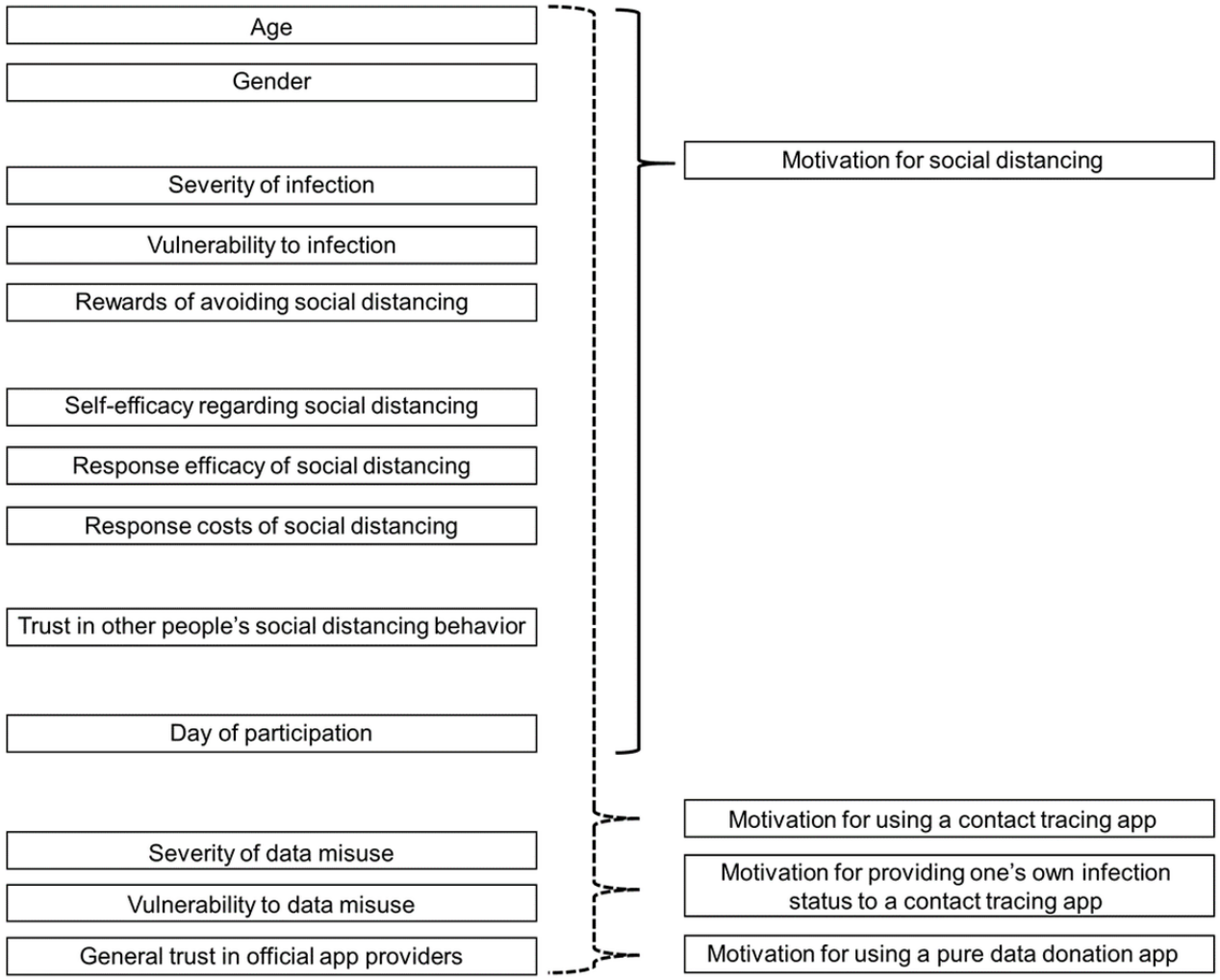
H2b: The perceived response efficacy of social distancing is positively related to the motivation for social distancing.

H2c: The perceived response costs of social distancing are negatively related to the motivation for social distancing.

In addition to these standard PMT variables, trust was added (see Figure 1). Rousseau et al [25] defined trust as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another." Indeed, because social distancing is a measure that only works effectively when performed collectively, trust in other people's social distancing behavior appears to be a critical component. However, two phenomena are conceivable: in terms of the social exchange theory [26] and the concept of reciprocity [27], higher trust in others' willingness to adequately perform social distancing may increase one's own motivation for social distancing. This relation would reflect the benefits of solidarity required to combat the current pandemic [28,29]. Alternatively, and in terms of a compensation mechanism to reduce one's own infection risk, a negative relationship between trust in others' social distancing behavior and one's own protection motivation is conceivable. Thus, the following undirected hypothesis was formulated:

H3a: Trust in other people's social distancing behavior is related to one's own motivation for social distancing.

Figure 1. The regression models examined in the present study, with independent variables on the left side and dependent variables on the right side.



Motivations for Using an App and Providing Personal Data

Regarding the use of an app and the provision of personal data, the question arises of whether the corresponding motivations are related to the cognitive variables assumed to constitute protection motivation or whether these are completely different evaluation processes. Depending on the answer to this question, an appropriate public communication strategy can be developed that supports realistic assessments and acceptance of different app types. Given that the main purpose of a contact tracing app is informing the user of critical contacts with infected persons, the perceived severity of and vulnerability to SARS-CoV-2 infection may be positively related to the motivation for using such an app. Perceived rewards of avoiding social distancing may also be associated with app use motivation. In contrast, coping appraisal of social distancing should not be related to the motivation for using a contact tracing app. In general, neither a contact tracing app nor a data donation app can actually help individual users to actively prevent infection. With respect to the motivation for providing personal data to both app types, threat and coping appraisal should not show a relationship because providing this information is only useful for other users (contact tracing app) or researchers and policy makers (data donation app). Thus, the following open research questions were formulated:

RQ1a: Are the motivations for app use and data provision related to the threat appraisal of SARS-CoV-2 infection?

RQ1b: Are the motivations for app use and data provision related to the coping appraisal of social distancing?

Moreover, and with respect to the idea of a compensation mechanism outlined above, reduced trust in other people's social distancing behavior may also be associated with increased motivation to use a contact tracing app, as this type of app can help the user monitor their own risk of being infected.

H3b: Trust in other people's social distancing behavior is negatively related to one's motivation for using a contact tracing app.

Additionally, the present PMT model was extended by threat and trust variables that are specifically tailored to app use (see Figure 1). In line with previous studies applying the PMT to data security issues [30,31], severity of data misuse and vulnerability to data misuse were added when focusing on the motivation for app use and the provision of personal data. Woon et al [31] found that perceived severity but not vulnerability was related to wireless network security measures. Banks et al [30] found that perceived threat was negatively associated with the intention to share personal information on web-based social media platforms, while perceived severity and vulnerability were both positively related to threat appraisal. Accordingly, the following hypotheses were tested:

H4a: The perceived severity of data misuse is negatively related to the motivations for using a contact tracing app and for using the Data Donation app.

H4b: The perceived vulnerability to data misuse is negatively related to the motivations for using a contact tracing app and for using the Data Donation app.

H4c: The perceived severity of data misuse is negatively related to the motivation for voluntarily providing one's own infection status to a contact tracing app.

H4d: The perceived vulnerability to data misuse is negatively related to the motivation for voluntarily providing one's own infection status to a contact tracing app.

Applying the trust construct to app use addresses the users' trust in the providers of an app that collects personally relevant information. Indeed, Lo et al [32] found that trust in social networking sites was positively related to users' willingness to provide personal information, leading to the following hypotheses:

H5a: General trust in official app providers with respect to the use, management, and protection of user data is positively related to the motivation for using both a contact tracing app and the Data Donation app.

H5b: General trust in official app providers with respect to the use, management, and protection of user data is positively related to the motivation for voluntarily providing one's own infection status to a contact tracing app.

The present study focuses on two fundamentally different app types. Although neither app type helps to reduce the user's individual risk of infection, a contact tracing app obviously has some personal utility, whereas the Data Donation app has no direct utility for its users. Consequently, motivation and acceptance should be higher for the more personally useful contact tracing app:

H6a: Motivation for use is higher for a contact tracing app compared to the Data Donation app.

H6b: Motivation for providing personal data is higher for a contact tracing app (infection status) compared to the Data Donation app (health, activity, and personal data).

H6c: The acceptance of mandatory use would be higher for a contact tracing app compared to the Data Donation app.

Finally, as shown in Figure 1, age was included in the regression models examined here due to older people's higher risk of a severe course of disease elicited by the novel coronavirus [33] and the ongoing public discussion about this risk [34]. Gender was included due to the well-known gender differences in health-related behavior [35]. Finally, the date of participation in this study was considered, as the pandemic was in progress; consequently, subjective risk assessment may change and habituation effects may occur over time.

Methods

Participants

The study included a final data set of 406 German-speaking participants (290 women, 71.4%) with a mean age of 32.56 years (SD 13.76). I previously excluded 9 participants: 3 (33%) were excluded due to incomplete data, 2 (22%) were younger

than the required minimum age of 18 years for participation, and 4 (44%) reported their gender as "diverse," which was an insufficient subsample for the gender-related statistical analyses. The highest educational attainment that was reported most often by the 406 participants was a higher education entrance qualification (173, 42.6%), followed by a master's degree or diploma (93, 22.9%), a bachelor's degree (70, 17.2%), completed vocational training (41, 10.1%), a secondary school certificate (21, 5.2%), no complete school leaving certificate (6, 1.5%), and main school graduation (2, 0.5%). At the time of the study, 385 of the 406 participants (94.8%) were not using any apps related to COVID-19, while 21 participants (5.2%) had already used the Data Donation app provided by the Robert Koch Institute. Importantly, a contact tracing app did not yet exist but had been officially announced for Germany at the time of the study. The participants were recruited through convenience sampling. The link to the study was broadly disseminated via mailing lists, social media, and a survey platform of a national journal (*Psychologie Heute*). Participation in the study was voluntary, and no incentives were provided. No identifying data were collected to guarantee the anonymity of the participants. At the start of the study, the participants were informed about the purpose of the study, that all data would be processed only for research purposes, that they would remain anonymous, and that they could prematurely stop the study at any point in time. In the latter case, the participant's data were deleted from the final data set before the analyses were performed. The participants finally indicated informed consent by clicking a corresponding box. In Germany, as stated by the German Research Association (Deutsche Forschungsgemeinschaft, DFG), ethics committee approval was not required for this survey because the research did not include a treatment, did not pose any threats or risks to the respondents, and was not associated with high physical or emotional stress; also, the respondents were informed about the objectives of the survey. The study ran for 30 days, starting on April 15, 2020.

Procedure

Participants initially provided their gender, age, and highest educational qualification. Afterward, the concept and functions of the two app types were presented in a detailed summary according to official descriptions of the PEPP-PT contact tracing app and the Data Donation app, as outlined above. With respect to each of the two apps, participants answered some questions related to app use. They subsequently assessed data security issues in terms of the perceived severity of potential misuse of their data, their perceived vulnerability to data misuse, and their general trust in official app providers with respect to the use, management, and protection of user data. Then, the participants reported their protection motivation by social distancing and their trust in other people's social distancing behavior. Finally, they assessed all PMT variables covered by threat appraisal (severity, vulnerability, rewards) and coping appraisal (self-efficacy, response efficacy, response costs) related to the prevention of SARS-CoV-2 infection by social distancing.

Measures

Questions Related to the Apps

Based on a 7-point scale (1="not motivated at all" to 7="very motivated"), participants indicated how much they were motivated to voluntarily use the described app and to voluntarily provide the personal data requested by the app. They also responded to the question "How much would you like it if the use of this app became a mandatory requirement for everyone?" (1="not at all" to 7="very") and whether they were already using the existing Data Donation app (yes/no). All subsequent measures were based on three items each with 7-point rating scales (1="completely disagree" to 7="completely agree"). All rating scales were continuously numbered from 1 to 7 and had verbal markers at the endpoints.

Data Security Issues

Items adapted from Banks et al [30] and Dang-Pham and Pittayachawan [36] were used to assess the participants' perceived severity of potential misuse of their data (eg, "If my personal information collected by a coronavirus app would be misused, it could harm me," Cronbach α=.80) and the perceived vulnerability to data misuse (eg, "I feel that I am vulnerable to misuse of my personal information collected by a coronavirus app," α=.76). Items adapted from Lo [32] were used to assess the participants' general trust in official providers of a COVID-19 app with respect to the use, management, and protection of user data (eg, "I believe that official providers of coronavirus apps are genuine and sincere in managing my personal information," α=.93).

Motivation for Social Distancing and Trust in Others' Social Distancing Behavior

Items adapted from Kaspar [37] assessed participants' protection motivation for social distancing (eg, "Over the next few weeks, I will avoid physical proximity to people who do not live in my household," α=.80). Trust in other people's intention to adequately perform social distancing behavior was measured by items adapted from Ross et al [38] (eg, "I think most people are currently trying their best to avoid getting too close to other people in public life so that the coronavirus cannot spread further," α=.80).

Threat and Coping Appraisal Regarding COVID-19

Items from previous studies [23,36,37,39,40] were adapted to social distancing behavior. Measures of threat appraisal included the perceived severity of an infection (eg, "If I became infected with the coronavirus, it would have a strong negative effect on my health," α=.89), the perceived vulnerability to an infection if no social distancing was performed (eg, "Other infected people will infect me with the coronavirus if I do not keep appropriate physical distance from them," α=.74), and perceived intrinsic rewards associated with not performing social distancing from people who do not live in the participant's household (eg, "I currently enjoy meeting with other people who do not live in my household," α=.91). Coping appraisal included the participants' perceived self-efficacy regarding social distancing (eg, "At the moment, it is easy for me to create physical distance to people who do not live in my household,"

α=.62), response efficacy in terms of the effectiveness of social distancing in averting an infection (eg, "Keeping sufficient distance from other people in public life protects me from the coronavirus," α=.82), and perceived response costs (eg, "At the moment, I find it exhausting to create sufficient spatial distances to other people in public space," α=.66).

Results

Intercorrelations and Mean Values of Independent Variables

In sum, intercorrelations among the independent variables of the regression models were rather low, with few exceptions (Table 1). Age and gender showed almost no significant correlation with the PMT or trust variables. The highest correlation was between age and the perceived severity of infection. Perceived response efficacy of social distancing showed several high correlations with other PMT variables. In accordance with the theoretical structure of threat and coping appraisal, perceived rewards and response costs were positively correlated with each other; however, they showed negative correlations with all other PMT variables (severity, vulnerability, self-efficacy, and response efficacy). Trust in other people's social distancing behavior was weakly positively correlated with self-efficacy and response efficacy. Perceived severity of and vulnerability to data misuse were highly positively correlated with each other but negatively correlated with general trust in app providers. Perceived vulnerability to infection and response efficacy of social distancing were positively correlated with general trust in app providers. Interestingly, the day on which the respondents participated in the survey, reflecting the temporal progress of the pandemic, was positively correlated with perceived rewards associated with avoiding social distancing but negatively correlated with self-efficacy regarding social distancing and trust in other people's social distancing behavior.

One-sample *t* tests showed that the mean value of most of the independent variables was above the midpoint of the 7-point scales (Table 2), except for the perceived severity of an infection (no deviation from the midpoint of the scale) and perceived rewards of avoiding social distancing (below the midpoint of the scale). The self-efficacy and response efficacy of social distancing were rated particularly high.

Table 1. Bivariate correlations (Pearson *r* and two-tailed *P* value) among all independent variables of the regression models.

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Age													
r	1	–.02	.29	.08	–.13	.23	.06	.04	.12	–.08	.09	.10	–.05
P value	—[a]	.66	<.001	.11	.01	<.001	.20	.45	.01	.13	.08	.04	.37
2. Gender[b]													
r	–.02	1	.04	.06	–.11	.06	.06	.07	.05	.02	.01	–.05	.01
P value	.66	—	.41	.23	.03	.24	.21	.19	.37	.66	.83	.28	.92
3. Severity of infection													
r	.29	.04	1	.44	–.17	.19	.25	–.14	–.07	.04	.17	.05	.11
P value	<.001	.41	—	<.001	.001	<.001	<.001	.004	.15	.39	.001	.32	.03
4. Vulnerability to infection													
r	.08	.06	.44	1	–.31	.25	.52	–.12	.05	.03	.01	–.05	.25
P value	.11	.23	<.001	—	<.001	<.001	<.001	.01	.32	.53	.79	.36	<.001
5. Rewards of avoiding social distancing													
r	–.13	–.11	–.17	–.31	1	–.37	–.43	.28	–.08	.17	.10	.12	–.14
P value	.01	.03	.001	<.001	—	<.001	<.001	<.001	.11	.001	.045	.02	.006
6. Self-efficacy regarding social distancing													
r	.23	.06	.19	.25	–.37	1	.50	–.25	.16	–.17	.05	–.03	.10
P value	<.001	.24	<.001	<.001	<.001	—	<.001	<.001	.002	<.001	.30	.58	.05
7. Response efficacy of social distancing													
r	.06	.06	.25	.52	–.43	.50	1	–.21	.15	–.04	–.06	–.16	.31
P value	.20	.21	<.001	<.001	<.001	<.001	—	<.001	.003	.43	.23	.001	<.001
8. Response costs of social distancing													
r	.04	.07	–.14	–.12	.28	–.25	–.21	1	–.07	.06	.12	.07	–.07
P value	.45	.19	.004	.01	<.001	<.001	<.001	—	.18	.20	.01	.15	.15
9. Trust in other people's social distancing behavior													
r	.12	.05	–.07	.05	–.08	.16	.15	–.07	1	–.17	.02	–.01	.10
P value	.01	.37	.15	.32	.11	.002	.003	.18	—	.001	.74	.87	.052
10. Day of participation													
r	–.08	.02	.04	.03	.17	–.17	–.04	.06	–.17	1	.05	.05	–.001
P value	.13	.66	.39	.53	.001	<.001	.43	.20	.001	—	.28	.33	.98
11. Severity of data misuse													
r	.09	.01	.17	.01	.10	.05	–.06	.12	.02	.05	1	.53	–.25
P value	.08	.83	.001	.79	.045	.30	.23	.01	.74	.28	—	<.001	<.001
12. Vulnerability to data misuse													
r	.10	–.05	.05	–.05	.12	–.03	–.16	.07	–.01	.05	.53	1	–.55
P value	.04	.28	.32	.36	.02	.58	.001	.15	.87	.33	<.001	—	<.001
13. General trust in official app providers													
r	–.05	.01	.11	.25	–.14	.10	.31	-.07	.10	–.001	–.25	–.55	1
P value	.37	.92	.03	<.001	.006	.05	<.001	.15	.052	.98	<.001	<.001	—

[a]—: not applicable.

[b]0=male, 1=female.

Table 2. Descriptive statistics and results of one-sample t tests against the scales' midpoint value (4) for independent variables of the regression models (age, gender, and day of participation were excluded due to the inappropriateness of the statistics in these cases).

Variable	Mean (SD)	t_{405}	*P* value	Cohen *d*
Severity of infection	4.01 (1.55)	0.139	.89	0.01
Vulnerability to infection	4.85 (1.46)	11.679	<.001	0.58
Rewards of avoiding social distancing	2.57 (1.60)	–17.976	<.001	0.89
Self-efficacy regarding social distancing	6.02 (1.07)	38.148	<.001	1.89
Response efficacy of social distancing	5.92 (1.17)	33.194	<.001	1.64
Response costs of social distancing	4.48 (1.48)	6.526	<.001	0.32
Trust in other people's social distancing behavior	5.10 (1.21)	18.372	<.001	0.91
Severity of data misuse	5.09 (1.52)	14.475	<.001	0.72
Vulnerability to data misuse	4.96 (1.43)	13.442	<.001	0.67
General trust in official app providers	4.20 (1.65)	2.410	.02	0.12

Motivation for Social Distancing

In the next step, multiple regression analyses were conducted (Table 3). Initially, the assumptions of the linear regression model [41] were checked and met for all models, except for one case of pronounced heteroscedasticity (see Multimedia Appendix 1). Hence, significance testing was based on the heteroscedasticity-robust HC3 estimator in this case [42], while the standard ordinary least squares (OLS) estimator was preferred in cases of homoscedasticity [43]. The participants' motivation for social distancing, serving as a dependent variable, showed a positive relation to the perceived severity of an infection (supporting H1a), no relation to the perceived vulnerability to an infection (contradicting H1b), and a negative relation to perceived rewards associated with avoiding social distancing (supporting H1c). Self-efficacy and response efficacy of social distancing were positively and strongly related to the motivation for social distancing (supporting H2a and H2b), whereas perceived response costs were unrelated (contradicting H2c). Finally, trust in other people's social distancing behavior was positively related to participants' motivation for social distancing (supporting H3a), whereas the day of participation, age, and gender were nonrelated. Overall, the model explained 55% of the interindividual variance in participants' motivation for social distancing. Importantly, all independent variables except gender showed significant bivariate correlations with the motivation for social distancing; however, several of these significant relationships disappeared in the complete regression model that simultaneously considered all independent variables (for bivariate correlations, see Multimedia Appendix 1).

Table 3. Results of the multiple regression analyses with standardized coefficients (β) and *P* values based on the heteroscedasticity-robust HC3 estimator (P_{HC3}) or standard OLS estimates (P_{OLSE}).

Independent variable	Motivation for social distancing (R^2=.547, P<.001)		Motivation for using a contact tracing app (R^2=.457, P<.001)		Motivation for providing the infection status to a contact tracing app (R^2=.423, P<.001)		Motivation for using the Data Donation app (R^2=.344, P<.001)	
	β	P_{HC3} [a]	β	P_{OLSE} [b]	β	P_{OLSE}	β	P_{OLSE}
Age	–.034	.34	–.021	.61	–.014	.74	–.088	.05
Gender[c]	–.015	.70	–.047	.22	–.007	.85	.027	.52
Severity of infection	.117	.003	.077	.09	.027	.56	.039	.43
Vulnerability to infection	–.014	.74	.072	.14	.042	.40	.015	.77
Rewards of avoiding social distancing	–.254	<.001	–.017	.70	–.051	.26	–.058	.23
Self-efficacy regarding social distancing	.211	<.001	.128	.006	.089	.06	.008	.88
Response efficacy of social distancing	.401	<.001	.103	.045	.098	.07	.092	.11
Response costs of social distancing	.029	.41	.137	.001	.056	.18	.040	.37
Trust in other people's social distancing behavior	.118	.003	–.078	.046	–.087	.03	–.103	.02
Day of participation	–.025	.53	–.042	.29	–.012	.77	–.027	.53
Severity of data misuse	N/A[d]	N/A	–.098	.03	–.075	.11	–.070	.16
Vulnerability to data misuse	N/A	N/A	–.219	<.001	–.224	<.001	–.195	.001
General trust in official app providers	N/A	N/A	.379	<.001	.384	<.001	.353	<.001

[a]P_{HC3}: *P* value based on the heteroscedasticity-robust HC3 estimator.

[b]P_{OLSE}: *P* value based on the standard ordinary least squares estimate.

[c]0=male, 1=female.

[d]Not applicable.

Motivations for Using a Contact Tracing App and the Data Donation App

Regarding participants' motivation for using an app (Table 3), the multiple regression analyses revealed some differences between the contact tracing app and the Data Donation app. Independently of the app type, there was no relation between use motivation and the components of threat appraisal of SARS-CoV-2 infection (severity, vulnerability, and rewards) (RQ1a). In contrast, all components of coping appraisal (self-efficacy, response efficacy, and response costs) were positively related to the motivation for using a contact tracing app but not related to the motivation for using the Data Donation app (RQ1b). Trust in other people's social distancing behavior was negatively related to the motivation for using a contact tracing app (supporting H3b) but also to the motivation for using the Data Donation app (not predicted). Perceived severity of data misuse was negatively and exclusively related to the motivation for using the contact tracing app (partially supporting H4a), while age was negatively and exclusively related to the motivation for using the Data Donation app (not predicted). Perceived vulnerability to data misuse was negatively and strongly related to the motivations for using both app types (supporting H4b). Also, general trust in official app providers showed a positive and the most pronounced relation to the motivations for using both app types (supporting H5a). The participants' gender and the day of participation in the study were not related to app use motivation. The models explained 46% and 34% of the interindividual variance in the participants' motivation for using a contact tracing app and the Data Donation app, respectively.

Motivation for Providing One's Own Infection Status to a Contact Tracing App

In the final regression analysis, the participants' motivation for voluntarily providing their own infection status to a contact tracing app served as the dependent variable (Table 3). Threat appraisal of an infection (RQ1a) and coping appraisal of social distancing (RQ1b) were not significantly related to the motivation for providing one's own infection status, as expected. Trust in other people's social distancing behavior was negatively related to the motivation for providing the infection status to the contact tracing app (not predicted). Perceived severity of data misuse did not show a significant relation to participants' willingness to share their infection status (contradicting H4c); however, perceived vulnerability to data misuse showed a negative relationship (supporting H4d). Moreover, general trust in official app providers showed a positive and strong relationship to the motivation for providing the infection status (supporting H5b). Again, the participants' gender and the day of participation in the study were not related to the motivational dependent variable. The explained variance was 42%.

Use Motivation and Acceptance of App Types

Finally, *t* tests for paired samples revealed that the motivation for using an app was higher for the contact tracing app (mean 4.11, SD 2.24) than for the Data Donation app (mean 3.76, SD 2.13; t_{405}=3.72, *P*<.001, Cohen *d*=0.18) (supporting H6a). The motivation for providing the personal data requested by the individual app type was also higher in the case of the contact tracing app (mean 4.48, SD 2.32) compared to the Data Donation app (mean 3.41, SD 2.23; t_{405}=10.86, *P*<.001, *d*=0.54) (supporting H6b). Also, participants were more receptive to the idea of the contact tracing app becoming mandatory (mean 3.06, SD 2.17) compared to the Data Donation app (mean 2.65, SD 1.98; t_{405}=6.57, *P*<.001, *d*=0.33) (supporting H6c).

Discussion

The results of the present study show that the PMT is a useful model to explain people's motivation to protect themselves from SARS-CoV-2 infection by social distancing and to use apps related to the COVID-19 pandemic. Although previous research [44] showed that a motivation-behavior gap can occur because "people do not always do the things that they intend to do," motivation supports effective health measures.

Social Distancing

Threat-appraisal of potential infection was related to the motivation for social distancing. Perceived severity of infection was positively related to the motivation for social distancing; however, perceived vulnerability was not. This result indicates that the perceived severity of an infection with SARS-CoV-2 is more important for social distancing motivation than the perceived vulnerability when both variables are simultaneously considered. Also, perceived intrinsic rewards of not performing social distancing showed a strong negative relationship to protection motivation, indicating that perceived social benefits from being physically surrounded by others can counteract prevention strategies. Hence, it appears to be important to develop alternative and appropriate means to satisfy people's social needs during the pandemic. Indeed, the longer the pandemic and the restrictions associated with it last, the more the commitment of the population to prevention strategies may decrease. The day of participation in this study, reflecting the temporal progress of the pandemic, was positively correlated with perceived rewards associated with not social distancing but negatively correlated with self-efficacy regarding social distancing and trust in other people's social distancing behavior. A more encouraging result is that coping appraisal of social distancing, namely self-efficacy and response efficacy, showed strong positive relationships to the motivation for social distancing and were rated above average. Hence, it appears it would be fruitful to foster these factors with health campaigns. At the same time, perceived costs of social distancing were not related to protection motivation but were also rated above average. In contrast, the perceived rewards of avoiding social distancing were rated below average; this draws a somewhat contradictory picture that should be scrutinized in further research. Finally, and in line with social exchange theory [26] and the concept of reciprocity [27], participants' trust in other people's social distancing behavior was positively related to their own motivation for social distancing. This result is promising, as it supports the relevance of solidarity required to combat the current pandemic [28,29].

App Use

It is important to note two aspects once more. First, while using the Data Donation app and providing the personal data requested by this app are basically the same, using a contact tracing app and voluntarily providing one's own infection status to this app are independent measures. Second, although neither app type helps reduce the user's individual risk of infection, a contact tracing app has some personal utility because it can help an individual monitor their own risk of being infected, whereas the Data Donation app has no direct utility for its users. Given these conceptual and technical differences, threat-appraisal of potential infection was not related to the motivation for using either app type or for providing one's own infection status to a contact tracing app. However, all components of coping appraisal of social distancing were positively related to the motivation for using a contact tracing app but were not related to using the Data Donation app. Self-efficacy and the response efficacy of social distancing were positively correlated to the motivation for using a contact tracing app, indicating that people who believe in their coping skills and the effectiveness of the recommended coping strategy tend to support further measures to combat the pandemic. At the same time, the participants appeared to be interested in effective ways to combat the pandemic overall, as the motivation for using a contact tracing app increased when perceived costs of one's own social distancing behavior increased but trust in other people's social distancing behavior decreased. Trust in others' social distancing behavior was additionally negatively related to the motivations for providing one's own infection status to a contact tracing app and for using the Data Donation app. Hence, when people have the impression that their fellow human beings are being less solidary by adhering less to recommended or even prescribed behaviors, they apparently attempt to compensate for this tendency by donating their personal data without receiving any direct counter value. However, participants' motivation for use, providing the requested data, and accepting mandatory use were higher for the contact tracing app than for the Data Donation app. Consequently, people are still more motivated to use the more personally useful app. Interestingly, age was negatively related to the motivation for using the Data Donation app. This result may indicate young people's generally higher willingness to use apps, as already shown by Cho [45]. However, when participants were explicitly asked how willing they were to provide the personal, health, and activity data requested by this app (although this was already highlighted in the app description), this willingness was not correlated with age (*r*=–.02, *P*=.74). Apparently, it is important to be as transparent as possible when indicating which specific data are collected by an app to create the basis for truly informed consent. Finally, the perceived severity of and vulnerability to data misuse were negatively related to the participants' motivation for using a contact tracing app; meanwhile, only vulnerability to data misuse was negatively related to the motivation for providing one's own infection status to a contact tracing app and for using the Data Donation app. Also, the participants' general trust in

official app providers was the most important independent variable with respect to app use motivation and data provision and donation. This result emphasizes the significant role of data security issues and trust in the context of app-based measures to combat the current pandemic.

Limitations

Some limitations of the present study should be mentioned. First, these are cross-sectional, correlational data that do not allow conclusions regarding causal relationships between the independent and dependent variables of the regression models. Second, although the present sets of psychological variables explained substantial interindividual variance in the participants' motivation for social distancing, using apps, and providing personal data, situational factors may also play important roles. For example, a high motivation for social distancing may be counteracted by insufficient space in urban infrastructure and public transport, while app use motivation also depends on the availability of the required hardware, usability of the software, and reliability of the data processing. Third, the participants' behavioral motivations were observed rather than their actual behavior. At the time of the study, an official contact tracing app was not yet available but had already been announced by the German Federal Government. The app has since been released (June 16, 2020); therefore, the actual download frequency and behavioral data of the app will be available for future research. Fourth, and relatedly, open questions remain as to whether different providers are assessed as having different levels of trust and how this could influence the relationships observed. Fifth, the present results are based on linear regression analyses, as the statistical assumptions were met, and all rating scales were treated as metric (except gender as a nominal variable), including the dependent variables. However, ordinal regression models may be considered as an alternative approach (with different limitations) which, however, largely replicated the present results for all models (see Multimedia Appendix 1). Finally, the participants in this sample had a mean age of approximately 33 years and were all residents of Germany; therefore, the generalizability of the present results to older people and other countries or cultures should be applied with caution.

Conclusion

The present study revealed four key findings. First, the present models revealed some important cognitive factors that constitute people's motivation for social distancing and using apps to combat the COVID-19 pandemic. However, the reduced model assessing the motivation for social distancing explained more interindividual variance than the extended model addressing app use, indicating that app use is more strongly constituted by other factors. Second, in addition to processes of threat and coping appraisal, social trust was found to be a relevant factor, highlighting the importance of both interpersonal solidarity and data security issues in the context of the ongoing pandemic. Third, the focus of the present study was on the joint contribution of several independent variables to the motivations for social distancing, using an app, and providing health-related data. As a consequence, several bivariate correlations between independent and dependent variables disappeared when independent variables were considered simultaneously in the regression models (cf. Multimedia Appendix 1). This result indicates that health campaigns addressing complex cognitive appraisal processes may fall short when focusing on selected correlations between individual variables. Finally, participants preferred the use of a contact tracing app compared to a pure data donation app that has no direct utility for its users.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bivariate correlations between independent and dependent variables of the regression models, robustness checks, and partial regression plots.
[PDF File (Adobe PDF File), 2595 KB - jmir_v22i8e21613_app1.pdf]

References

1. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
2. Ahmed S, Quadeer A, McKay M. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. Viruses 2020 Feb 25;12(3):254 [FREE Full text] [doi: 10.3390/v12030254] [Medline: 32106567]
3. Crosby J, Heimann M, Burleson S, Anzalone B, Swanson J, Wallace D, et al. COVID‐19: A review of therapeutics under investigation. J Am Coll Emerg Phys 2020 Apr 19;1(3):231-237 [FREE Full text] [doi: 10.1002/emp2.12081]
4. Ghosh A, Brindisi M, Shahabi D, Chapman M, Mesecar A. Drug Development and Medicinal Chemistry Efforts toward SARS-Coronavirus and Covid-19 Therapeutics. ChemMedChem 2020 Jun 04;15(11):907-932 [FREE Full text] [doi: 10.1002/cmdc.202000223] [Medline: 32324951]
5. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int J Surg 2020 Jun;78:185-193 [FREE Full text] [doi: 10.1016/j.ijsu.2020.04.018] [Medline: 32305533]

6. Ma Q, Shan H, Zhang H, Li G, Yang R, Chen J. Potential utilities of mask-wearing and instant hand hygiene for fighting SARS-CoV-2. J Med Virol 2020 Mar 31;92:1567-1571 [FREE Full text] [doi: 10.1002/jmv.25805] [Medline: 32232986]
7. Wilder-Smith A, Freedman D. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. J Travel Med 2020 Mar 13;27(2):taaa020 [FREE Full text] [doi: 10.1093/jtm/taaa020] [Medline: 32052841]
8. Bassi A, Arfin S, John O, Jha V. An overview of mobile applications (apps) to support the coronavirus disease-2019 response in India. Indian J Med Res 2020 [FREE Full text] [doi: 10.4103/ijmr.ijmr_1200_20]
9. Guinchard A. Our digital footprint under Covid-19: should we fear the UK digital contact tracing app? Int Rev Law Comput Tech 2020 Jul 15:1-14 [FREE Full text] [doi: 10.1080/13600869.2020.1794569]
10. Parker M, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020 Jul;46(7):427-431 [FREE Full text] [doi: 10.1136/medethics-2020-106314] [Medline: 32366705]
11. Morley J, Cowls J, Taddeo M, Floridi L. Ethical guidelines for COVID-19 tracing apps. Nature 2020 Jun;582(7810):29-31 [FREE Full text] [doi: 10.1038/d41586-020-01578-0] [Medline: 32467596]
12. Li J, Guo X. Global Deployment Mappings and Challenges of Contact-tracing Apps for COVID-19. SSRN Journal 2020 May 26:preprint [FREE Full text] [doi: 10.2139/ssrn.3609516]
13. Danquah L, Hasham N, MacFarlane M, Conteh F, Momoh F, Tedesco A, et al. Use of a mobile application for Ebola contact tracing and monitoring in northern Sierra Leone: a proof-of-concept study. BMC Infect Dis 2019 Sep 18;19(1):810 [FREE Full text] [doi: 10.1186/s12879-019-4354-z] [Medline: 31533659]
14. Tom-Aba D, Nguku P, Arinze C, Krause G. Assessing the Concepts and Designs of 58 Mobile Apps for the Management of the 2014-2015 West Africa Ebola Outbreak: Systematic Review. JMIR Public Health Surveill 2018 Oct 29;4(4):e68 [FREE Full text] [doi: 10.2196/publichealth.9015] [Medline: 30373727]
15. Abeler J, Bäcker M, Buermeyer U, Zillessen H. COVID-19 Contact Tracing and Data Protection Can Go Together. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e19359 [FREE Full text] [doi: 10.2196/19359] [Medline: 32294052]
16. Rogers RW. Cognitive and physiological processes in fear-based attitude change: A revised theory of protection motivation. In: Cacioppo J, Petty R, editors. Social Psychophvsiology: A Sourcebook. New York, NY: Guilford; 1983:153-176.
17. Floyd D, Prentice-dunn S, Rogers RW. A Meta-Analysis of Research on Protection Motivation Theory. J Appl Social Pyschol 2000 Feb;30(2):407-429. [doi: 10.1111/j.1559-1816.2000.tb02323.x]
18. Milne S, Sheeran P, Orbell S. Prediction and Intervention in Health-Related Behavior: A Meta-Analytic Review of Protection Motivation Theory. J Appl Social Pyschol 2000 Jan;30(1):106-143. [doi: 10.1111/j.1559-1816.2000.tb02308.x]
19. Babazadeh T, Nadrian H, Banayejeddi M, Rezapour B. Determinants of Skin Cancer Preventive Behaviors Among Rural Farmers in Iran: an Application of Protection Motivation Theory. J Cancer Educ 2017 Sep;32(3):604-612 [FREE Full text] [doi: 10.1007/s13187-016-1004-7] [Medline: 26922176]
20. Ling M, Kothe E, Mullan B. Predicting intention to receive a seasonal influenza vaccination using Protection Motivation Theory. Soc Sci Med 2019 Jul;233:87-92. [doi: 10.1016/j.socscimed.2019.06.002] [Medline: 31195194]
21. Tsai H, Jiang M, Alhabash S, LaRose R, Rifon N, Cotten S. Understanding online safety behaviors: A protection motivation theory perspective. Comput Secur 2016 Jun;59:138-150. [doi: 10.1016/j.cose.2016.02.009]
22. Marett K, McNab L, Harris R. Social Networking Websites and Posting Personal Information: An Evaluation of Protection Motivation Theory. AIS Trans Hum-Comp Interact 2011 Sep 30;3(3):170-188 [FREE Full text] [doi: 10.17705/1thci.00032]
23. Vance A, Siponen M, Pahnila S. Motivating IS security compliance: Insights from Habit and Protection Motivation Theory. Information & Management 2012 May;49(3-4):190-198. [doi: 10.1016/j.im.2012.04.002]
24. Sommestad T, Karlzén H, Hallberg J. A meta-analysis of studies on protection motivation theory and information security behaviour. Int J Inf Secur Priv 2015 Jan;9(1):26-46. [doi: 10.4018/ijisp.2015010102]
25. Rousseau D, Sitkin S, Burt R, Camerer C. Not So Different After All: A Cross-Discipline View Of Trust. Acad Manage Rev 1998 Jul;23(3):393-404. [doi: 10.5465/amr.1998.926617]
26. Homans G. Social Behavior as Exchange. Am J Sociol 1958 May;63(6):597-606. [doi: 10.1086/222355]
27. Buchan N, Croson R, Dawes R. Swift Neighbors and Persistent Strangers: A Cross - Cultural Investigation of Trust and Reciprocity in Social Exchange. Am J Sociol 2002 Jul;108(1):168-206. [doi: 10.1086/344546]
28. Ayalon L, Chasteen A, Diehl M, Levy B, Neupert S, Rothermund K, et al. Aging in Times of the COVID-19 Pandemic: Avoiding Ageism and Fostering Intergenerational Solidarity. J Gerontol B Psychol Sci Soc Sci 2020 Apr 16:A [FREE Full text] [doi: 10.1093/geronb/gbaa051] [Medline: 32296840]
29. Cheng K, Lam T, Leung C. Wearing face masks in the community during the COVID-19 pandemic: altruism and solidarity. The Lancet 2020 Apr:online [FREE Full text] [doi: 10.1016/s0140-6736(20)30918-1]
30. Banks M, Onita C, Meservy T. Risky Behavior in Online Social Media: Protection Motivation and Social Influence. In: AMCIS 2010 Proceedings. 2010 Presented at: 16th Americas Conference on Information Systems 2010; August 12-15, 2010; Lima, Peru.
31. Woon I, Tan G, Low R. A protection motivation theory approach to home wireless security. In: ICIS Proceedings. 2005 Jan Presented at: International Conference on Information Systems; December 11-14, 2005; Las Vegas, NV p. 367-380.

32. Lo J. Privacy concern, locus of control, and salience in a trust-risk model of information disclosure on social networking sites. In: AMCIS Proceedings. 2010 Presented at: 16th Americas Conference on Information Systems 2010; August 12-15, 2010; Lima, Peru. [doi: 10.4018/978-1-5225-7113-1.ch006]
33. Fratino L, Procopio G, Di Maio M, Cinieri S, Leo S, Beretta G. Coronavirus: Older Persons With Cancer in Italy in the COVID-19 Pandemic. Front Oncol 2020;10:648 [FREE Full text] [doi: 10.3389/fonc.2020.00648] [Medline: 32426285]
34. Grzelka M. Attitudes Toward Vulnerable Populations In The Time Of Covid-19: Critical Discourse Analysis Of Gazeta. SR 2020 Apr 07;4(2):121-132 [FREE Full text] [doi: 10.14746/sr.2020.4.2.09]
35. Kandrack M, Grant K, Segall A. Gender differences in health related behaviour: Some unanswered questions. Soc Sci Med 1991 Jan;32(5):579-590. [doi: 10.1016/0277-9536(91)90293-l]
36. Dang-Pham D, Pittayachawan S. Comparing intention to avoid malware across contexts in a BYOD-enabled Australian university: A Protection Motivation Theory approach. Comput Secur 2015 Feb;48:281-297. [doi: 10.1016/j.cose.2014.11.002]
37. Kaspar K. An embodiment perspective on protection motivation theory: the impact of incidental weight sensations on threat-appraisal, coping-appraisal, and protection motivation. Stud Psychol 2015 Jan;57(4):301-314 [FREE Full text] [doi: 10.21909/sp.2015.03.701]
38. Ross V, Fielding K, Louis W. Social trust, risk perceptions and public acceptance of recycled water: testing a social-psychological model. J Environ Manage 2014 May 01;137:61-68 [FREE Full text] [doi: 10.1016/j.jenvman.2014.01.039] [Medline: 24603028]
39. McMath B, Prentice-Dunn S. Protection Motivation Theory and Skin Cancer Risk: The Role of Individual Differences in Responses to Persuasive Appeals. J Appl Social Pyschol 2005 Mar;35(3):621-643. [doi: 10.1111/j.1559-1816.2005.tb02138.x]
40. Taylor AH, May S. Threat and coping appraisal as determinants of compliance with sports injury rehabilitation: An application of Protection Motivation Theory. J Sports Sci 1996 Nov 1;14(6):471-482. [doi: 10.1080/02640419608727734]
41. Poole M, O'Farrell P. The Assumptions of the Linear Regression Model. Trans Inst Br Geogr 1971 Mar(52):145-158. [doi: 10.2307/621706]
42. Hayes A, Cai L. Using heteroskedasticity-consistent standard error estimators in OLS regression: an introduction and software implementation. Behav Res Methods 2007 Nov;39(4):709-722 [FREE Full text] [doi: 10.3758/bf03192961] [Medline: 18183883]
43. Long J, Ervin L. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. Am Stat 2000 Aug;54(3):217-224. [doi: 10.1080/00031305.2000.10474549]
44. Sheeran P, Webb T. The Intention-Behavior Gap. Soc Personal Psychol Compass 2016 Sep 05;10(9):503-518. [doi: 10.1111/spc3.12265]
45. Cho J. Roles of Smartphone App Use in Improving Social Capital and Reducing Social Isolation. Cyberpsychol Behav Soc Netw 2015 Jun;18(6):350-355. [doi: 10.1089/cyber.2014.0657] [Medline: 26075923]

Abbreviations

COVID-19: coronavirus disease
DFG: Deutsche Forschungsgemeinschaft
OLS: ordinary least squares
PEPP-PT: Pan-European Privacy-Preserving Proximity Tracing
PMT: protection motivation theory
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 18.06.20; peer-reviewed by CF Yen, J Li; comments to author 13.07.20; accepted 24.07.20; published 27.08.20.

Please cite as:
Kaspar K
Motivations for Social Distancing and App Use as Complementary Measures to Combat the COVID-19 Pandemic: Quantitative Survey Study
J Med Internet Res 2020;22(8):e21613
URL: http://www.jmir.org/2020/8/e21613/
doi:10.2196/21613
PMID:32759100

Original Paper

Temporal and Location Variations, and Link Categories for the Dissemination of COVID-19–Related Information on Twitter During the SARS-CoV-2 Outbreak in Europe: Infoveillance Study

Monika Pobiruchin[1,2*], Dr Sc Hum; Richard Zowalla[2,3,4*], MSc; Martin Wiesner[2,3*], Dipl Inform Med

[1]GECKO Institute for Medicine, Informatics & Economics, Heilbronn University, Heilbronn, Germany

[2]Consumer Health Informatics SIG, German Association for Medical Informatics, Biometry & Epidemiology (GMDS e. V.), Cologne, Germany

[3]Department of Medical Informatics, Heilbronn University, Heilbronn, Germany

[4]Center for Machine Learning, Heilbronn University, Heilbronn, Germany

[*]all authors contributed equally

Corresponding Author:
Monika Pobiruchin, Dr Sc Hum
GECKO Institute for Medicine, Informatics & Economics
Heilbronn University
Max-Planck-Str 39
Heilbronn, 74081
Germany
Phone: 49 7131504633
Email: monika.pobiruchin@hs-heilbronn.de

Abstract

Background: The spread of the 2019 novel coronavirus disease, COVID-19, across Asia and Europe sparked a significant increase in public interest and media coverage, including on social media platforms such as Twitter. In this context, the origin of information plays a central role in the dissemination of evidence-based information about the SARS-CoV-2 virus and COVID-19. On February 2, 2020, the World Health Organization (WHO) constituted a "massive infodemic" and argued that this situation "makes it hard for people to find trustworthy sources and reliable guidance when they need it."

Objective: This *infoveillance* study, conducted during the early phase of the COVID-19 pandemic, focuses on the social media platform Twitter. It allows monitoring of the dynamic pandemic situation on a global scale for different aspects and topics, languages, as well as regions and even whole countries. Of particular interest are temporal and geographical variations of COVID-19–related tweets, the situation in Europe, and the categories and origin of shared external resources.

Methods: Twitter's Streaming application programming interface was used to filter tweets based on 16 prevalent hashtags related to the COVID-19 outbreak. Each tweet's text and corresponding metadata as well as the user's profile information were extracted and stored into a database. Metadata included links to external resources. A link categorization scheme—introduced in a study by Chew and Eysenbach in 2009—was applied onto the top 250 shared resources to analyze the relative proportion for each category. Moreover, temporal variations of global tweet volumes were analyzed and a specific analysis was conducted for the European region.

Results: Between February 9 and April 11, 2020, a total of 21,755,802 distinct tweets were collected, posted by 4,809,842 distinct Twitter accounts. The volume of #covid19-related tweets increased after the WHO announced the name of the new disease on February 11, 2020, and stabilized at the end of March at a high level. For the regional analysis, a higher tweet volume was observed in the vicinity of major European capitals or in densely populated areas. The most frequently shared resources originated from various social media platforms (ranks 1-7). The most prevalent category in the top 50 was "Mainstream or Local News." For the category "Government or Public Health," only two information sources were found in the top 50: US Centers for Disease Control and Prevention at rank 25 and the WHO at rank 27. The first occurrence of a prevalent scientific source was Nature (rank 116).

Conclusions: The naming of the disease by the WHO was a major signal to address the public audience with public health response via social media platforms such as Twitter. Future studies should focus on the origin and trustworthiness of shared resources, as monitoring the spread of fake news during a pandemic situation is of particular importance. In addition, it would be beneficial to analyze and uncover bot networks spreading COVID-19–related misinformation.

(J Med Internet Res 2020;22(8):e19629) doi:10.2196/19629

KEYWORDS

COVID-19; SARS-CoV-2; social media; public health; Twitter; infoveillance; infodemiology; infodemic; health informatics; disease surveillance

Introduction

Overview

The emergence of SARS-CoV-2 and the associated COVID-19 [1] was first observed and described in China [2-6]. The subsequent spread across Asia [7] and Europe [8], including Northern Italy [9-12], in early 2020 sparked a significant increase in public interest and media coverage [13] including on the social media platforms Weibo [14] and Twitter [15,16]. During the following weeks, several SARS-CoV-2 infections were reported in other European countries [17,18] including the United Kingdom [19], Germany [20,21], France [22], and Spain [23].

According to Merchant and Lurie [24], several aspects play an important role in coping with the COVID-19 pandemic situation, especially in the digital age. First, "directing people to trusted sources" stands out, and neither a vaccine or drug against SARS-CoV-2 exists as of the time of writing. Second, the authors describe "social media as a diagnostic tool and referral system." By monitoring related activities on different social media platforms, public authorities or research institutions can gather valuable insights into regional trends, country-specific trends, or even the global situation. Third, misinformation and rumors can quickly spread in a globally connected world [24,25]. Misbeliefs, fake news, and conspiracy theories pose a severe threat and might put people's lives in danger [26]. In this context, Merchant and Lurie [24] propose a strategy of "counteracting misinformation" actively. In this way, they argue that "enabling a culture of preparedness" could be achieved.

In this context, the origin of information plays a central role in the dissemination of evidence-based information about the SARS-CoV-2 virus and the associated COVID-19. On February 2, 2020, the World Health Organization (WHO) constituted a "massive infodemic" and argued that this situation "makes it hard for people to find trustworthy sources and reliable guidance when they need it" [27,28].

Several trustworthy sources seem to be of particular interest [15,29]: research, public health, and government institutions, as well as news agencies or broadcasting companies and digital or print newspapers.

Related Work

In 2009, Eysenbach [30] described the *infodemiology* and *infoveillance* concepts as a set of "public health informatics methods" to "analyze search, communication and publication behavior on the Internet." During the 2009 H1N1 flu pandemic, Chew and Eysenbach [31] applied this concept for a content analysis of topic-related posts on Twitter in which they analyzed diseases-related trends, the origin of shared resources, and the sentiment expressed in swine flu tweets as posted via the platform.

Fu et al [32] analyzed how people reacted to the Zika epidemic in the Americas from 2015 to 2016. The authors analyzed 132,033 tweets with the keyword "zika" written in the languages English, Spanish, and Portuguese via the Twitter application programming interface (API). The authors reported that the top ranked shared resources originated from social media platforms such as "Facebook, Instagram, Twitter, YouTube, LinkedIn, Tumblr, the blogging site WordPress, [...] which accounted for 26% of all domains." In the Zika study, the Centers for Disease Control and Prevention (CDC) and the WHO amounted to "0.06%" and "0.05%," respectively. This corresponded to a 90th and 140th rank, respectively.

However, people do not only share evidence-based or trustworthy content in social media environments [33]. Southwell et al [34] pointed out that misinformation and perils exist that can lead to a spread of incorrect information, ambiguous meanings, and misperceptions, which can persist for a long period of time, and it can be resource intensive to counter misinformation "once it has enjoyed wide exposure."

In the context of the current COVID-19 pandemic, an example of such incorrect information is the "5G conspiracy theory" [35], which led to phone masts being attacked in the United Kingdom [36].

Abd-Alrazaq et al [16] analyzed the content and sentiment of about 2.8 million COVID-19–related tweets retrieved via the Twitter standard search API written in the English language. They identified "four main themes: origin of the virus; its sources; its impact on people, countries, and the economy; and ways of mitigating the risk of infection" by applying topic modelling techniques using latent Dirichlet allocation. However, the analysis of shared resources and temporal and geographical variations of their 2.8 million tweets collection was not in the focus of their study.

Aims of the Study

For this *infoveillance* study, during the early phase of the COVID-19 pandemic, the authors decided to focus on the social media platform Twitter, as the platform allows monitoring of the dynamic pandemic situation on a global scale in real time for different aspects of a topic, languages, as well as regions and even whole countries.

In this context, the research questions (RQs) of this study were as follows:

1. What tweet volume was observed among COVID-19–related hashtags at the beginning of the pandemic before and after the WHO announced the name of the disease?
2. How did information on COVID-19 and its associated impact spread during the epidemic situation in Europe from early February to early April 2020?

3. What proportion of information originates from public institutions, media channels, and scientific journals, and which channels stand out?

To the best of the authors' knowledge, no similar COVID-19 study has been conducted on a comparable scale.

Methods

Study Design

This *infoveillance* study on the use of hashtags in the early onset of SARS-CoV-2 in the European countries consists of three stages. First, to answer RQ1, tweets with SARS-CoV-2– and COVID-19–related hashtags were collected. An analysis of which hashtags were used depending on 7-day intervals was conducted after the WHO announcement that named the disease on February 11, 2020. Second, based on the given geolocation information, the number of tweets from the European countries and their variations in time were analyzed (RQ2). Third, *European* tweets with online resources (ie, URL information) were extracted. The target of the URL was examined to determine its origin (eg, news agency, government institution, social media).

Study Setting

In the early onset of the SARS-CoV-2 epidemic, several hashtags emerged worldwide. Based on the global Twitter trends and media coverage in January 2020, eight hashtags were initially included for collecting tweets beginning on February 9, 2020. In late February and at the beginning of March 2020, several other hashtags were increasingly used and, therefore, included in the study setting (see Multimedia Appendix 1). The special European focus was initiated by monitoring the worsening of the severe SARS-CoV-2 outbreak in the Northern Italian regions Lombardy and Emilia Romagna [10,11,37]. For this reason, the authors decided to add the two Italy-specific hashtags #coronavirusitaly and #coronavirusitalia that were prevalent around the third week of February 2020, as reported by Twitter trends at that time.

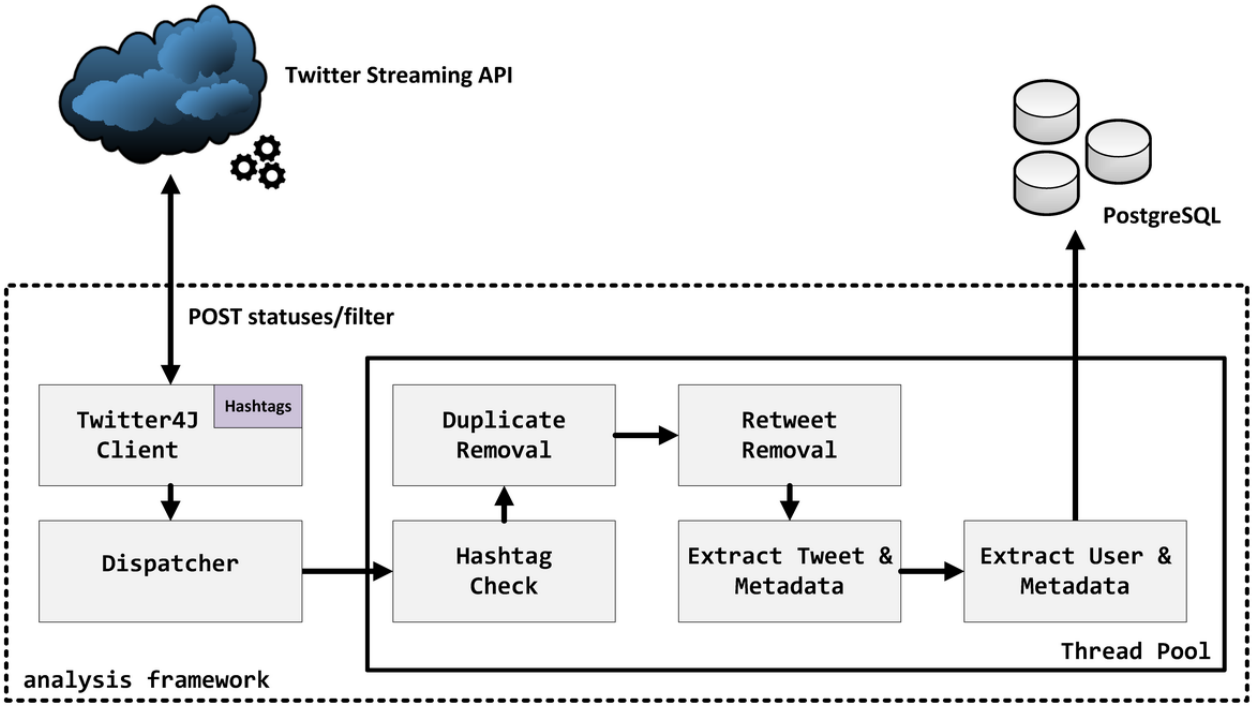
In total, 16 hashtags were selected for collecting COVID-19–related tweets for the purpose of temporal, geolocation, and link category analyses.

Data Acquisition

Twitter Data

For this study, tweets were collected between February 9 (midnight Central European Summer Time [CEST]) and April 11 (11:59 PM CEST), 2020, using the "Filter realtime Tweets" endpoint of the Twitter Streaming API [38] via the Java library *Twitter4J* [39] with the "standard" access level. To build the related filter query, the aforementioned hashtags (see Study Setting) were connected using the *OR* operator. Matching tweets were then processed by a self-implemented software framework written in Java (Oracle Corporation). Duplicate tweets as well as retweets were removed in this process. A tweet's text, its metadata (eg, URLs appended to a tweet), as well as the user's profile information were extracted. The results were then stored in a *PostgreSQL* in v10.12 [40]. The processing workflow is depicted in Figure 1.

Figure 1. Workflow of the processing steps and involved software components: lines with arrows indicate processing workflow for each tweet *t* returned by the Twitter streaming API under the given hashtags included in this study. Each *t* was processed in parallel by the analysis framework to reach high-throughput processing for the large volume of COVID-19–related tweets. API: application programming interface.



Twitter Analysis

In the context of this study, a tweet contains at least one of the 16 hashtags ([a] to [p] in Multimedia Appendix 1) as described in the section Study Setting. All analyses were conducted based on these hashtags. Additionally, detected hashtags mentioned in a specific tweet were not considered.

Twitter provides geographic information of a Twitter user's location (ie, latitude ϕ and longitude λ) [41]. According to the Twitter API, such geographic information can either be an exact point location or a bounding box (ie, a larger area or an entire region). Given such a bounding box, our analysis framework computed the geometric center of it and used this information as a point location.

However, Twitter users can deactivate sharing of their location information. If the geographic information was given, we leveraged this information to plot tweet locations on a map. The authors defined the European area with geographical limits ranging from $34.839° < \phi < 75.00°$ latitude (excluding the islands of Svalbard) and $-31.26192° < \lambda < 59.34569°$ longitude. Corresponding maps visualize the geographical and temporal spread of the pandemic via tweets in the European countries.

Link Category Analysis

Twitter users can share external resources to disseminate important information or to support an individual statement. In the context of RQ3, all URLs shared by users, excluding retweets or citations, were of particular interest. Before categorization could be conducted, URLs shortened by a corresponding service (eg, bit.ly, buff.ly) were resolved in an automatic procedure using the *crawler4j* framework [42]. In case shortened links could not be resolved, those URLs were left as originally captured via our analysis framework (see Figure 1). Next, domain aggregation was applied on each unique URL [43] (ie, "https://mhealth.jmir.org/" becomes "jmir.org"). The domain aggregation was conducted by a self-implemented software written in Java using the public suffix list provided by the Mozilla Foundation [44]. This transformation was conducted on a Ubuntu 18.04 LTS 64-bit computer running Java 11.0.7 on April 22, 2020.

The most prevalent (n=250) domain-aggregated URLs associated with a web site, as shared by Twitter users, were categorized according to the categories introduced by Chew and Eysenbach [31] in 2010. The category "No Reference" was not considered, as only tweets containing at least one URL were included in the link category analysis. Two additional categories with respect to RQ3 were introduced: (1) "Scientific resource" (eg, journal, magazine, preprint servers, or university provided COVID-19 dashboard) and (2) "URL Shortener."

The categorization was conducted manually by all three authors independently. Subsequently, the interrater reliability metrics percent agreement (PA) [45] and Fleiss κ [46] were computed. If there was a split situation, the authors discussed the specific case and resolved all unclear cases.

Statistical Analysis

Data were analyzed with the statistics software *R* (The R Foundation for Statistical Computing) in version 3.6.3 (February 29, 2020) on a Ubuntu 18.04 LTS 64-bit computer. The *R* package *ggplot2* [47] was used for visualization of tweets' and hashtags' temporal and geographic variations. In addition, *R* was used to compute PA and Fleiss κ.

Ethical Approval

This article does not contain any study of human participants performed by any of the authors. For this reason, no formal ethical approval is required.

Results

Principal Findings

Since the emergence of the first reports of human SARS-CoV-2 virus infections in China in late December 2019 and early January 2020 [5], the public interest and social media use grew steadily. The volume of #covid19-related tweets increased with the WHO announcement after February 11 [1] and stabilized at the end of March at a high level. Several hashtags were used in the early phase of the SARS-CoV-2 outbreak, such as #nCov2019, #nCov19, #nCov, and #2019nCov. Those earlier forms of referencing COVID-19 did not show substantial volume after the WHO announcement. Thus, the naming of the disease was a major signal to address the public audience with a public health response via social media platforms (ie, Twitter).

The situation in Europe changed with the #coronavirus outbreak in (Northern) Italy [37]. Public interest rose with climbing numbers of infections, as Italy became a hot spot of the epidemiological situation on the European continent. Country-specific hashtags were used to report on the Italian situation, and with the spread of the disease in Europe, users in other countries engaged with their individual hashtags such as #coronavirusES (Spain), #coronaFrance (France), and #coronavirusDeutschland (Germany). Nevertheless, neutral hashtags such as #covid19 or #COVID—19 showed constant use and corresponding high volume.

Many Twitter users expressed their engagement by sharing either image or multimedia content, or URLs to external references of which they believed provided important information to other users. A quarter of the observed tweets were posted with images, and another third provided links to external references. Of these references, 1 out of 5 cross-linked to posts in social media platforms (eg, YouTube, Instagram, or Reddit). Mainstream or local news resources were shared by 1 out of 8 posts. References to information provided by governmental or public health institutions and COVID-19–related scientific resources were posted rarely (1 out of 100).

Sample Characteristics

Between February 9 (midnight CEST) and April 11 (11:59 PM CEST), 2020, a total of 21,755,802 distinct tweets posted under the 16 hashtags were collected and stored in the study database. Those tweets were posted by 4,809,842 distinct Twitter accounts of which 83,560 were verified by the platform itself [48]. On average, each tweet contained 3.18 hashtags (min=1, max=47). The most prevalent languages were identified according to Twitter's language classification [49] and are listed in Table 1.

Of the 21,755,802 tweets, 25.78% (n=5,608,189) of tweets used (animated) images. Likewise, 4.95% (n=1,076,180) of all posts shared multimedia material (ie, videos). In total, 7,753,841 (34.16%) posts shared external resources. On average, a tweet referencing an external URL contained 1.04 URLs (min=1, max=10).

Table 1. Language distribution of the study sample.

Rank	Language	Observations (n=21,755,802), n (%)
1	English	11,829,991 (54.38)
2	Spanish	3,037,910 (13.96)
3	Undefined	1,325,729 (6.09)
4	French	1,246,211 (5.73)
5	Italian	898,979 (4.13)
6	Turkish	493,155 (2.27)
7	German	446,502 (2.05)
8	Portuguese	310,332 (1.43)
9	Indonesian	242,068 (1.11)
10	Hindi	228,966 (1.05)
11	Thai	227,665 (1.05)
12	Japanese	220,032 (1.01)
13	Arabic	195,541 (0.90)
14	Dutch	178,768 (0.82)
15	Catalan	155,535 (0.71)
≥16	Other	718,418 (3.30)

Twitter Analysis

Temporal Variations of Tweets

The total number of occurrences for each hashtag is presented in Table 2. The data shows a heterogeneous distribution of hashtag volume. The hashtags #WuhanVirus and #Wuhan were less frequently used than more *generic* hashtags such as #covid19. The four top hashtags in the study database represent 93.98% (24,203,025/25,754,619) of all hashtags mentioned: (1) #coronavirus, (2) #covid19, (3) #COVID—19, and (4) #Covid_19.

Figures 2 and 3 each depict the number of tweets per day that contained at least one of the COVID-19–related hashtags.

Overall, the number of daily tweets rose during the study period. The use of #covid19 increased throughout February and March. The trend was similar to the use of #coronavirus. However, the use of the hashtag #COVID—19 was fluctuating periodically. Similar peaks in usage could be detected for #Covid_19. Multimedia Appendix 2 provides a complete list of all depictions of temporal variations for each hashtag separately.

Table 2. Number of tweets per hashtag in ranked order within 7-day intervals.

Rank	Hashtag	Feb 9, n (%)	Feb 16[a], n (%)	Feb 23, n (%)	Mar 1, n (%)	Mar 8, n (%)	Mar 15, n (%)	Mar 22[b], n (%)	Mar 29[b], n (%)	Apr 5, n (%)	Total, n (%)
1	#coronavirus	337,478 (65.91)	159,013 (57.32)	1,000,498 (61.89)	1,287,233 (56.53)	2,057,022 (51.46)	1,609,870 (42.35)	1,537,502 (38.20)	1,553,761 (34.30)	1,587,109 (32.42)	11,129,486 (42.92)
2	#covid19	84,892 (16.58)	79,568 (28.68)	354,548 (21.93)	553,564 (24.31)	1,114,486 (27.88)	1,263,151 (33.23)	1,599,310 (39.73)	1,972,964 (43.55)	2,130,548 (43.52)	9,153,031 (35.30)
3	#COVID—19	—[c]	—	113,479 (7.02)	177,572 (7.80)	209,752 (5.25)	354,910 (9.34)	515,987 (12.82)	415,712 (9.18)	472,185 (9.65)	2,259,597 (8.71)
4	#Covid_19	—	—	—	—	152,901 (3.83)	297,663 (7.83)	230,274 (5.72)	443,837 (9.80)	536,257 (10.95)	1,660,932 (6.40)
5	#CoronaVirusUpdate	—	—	—	40,835 (1.79)	144,616 (3.62)	102,272 (2.69)	42,775 (1.06)	31,607 (0.70)	32,773 (0.67)	394,878 (1.52)
6	#CoronaVirusUpdates	—	—	56,143 (3.47)	25,440 (1.12)	47,877 (1.20)	76,200 (2.00)	19,024 (0.47)	29,460 (0.65)	27,665 (0.57)	281,809 (1.09)
7	#CoronaOutbreak	6237 (1.22)	2305 (0.83)	6070 (0.38)	96,691 (4.25)	105,210 (2.63)	16,081 (0.42)	10,052 (0.25)	11,611 (0.26)	14,131 (0.29)	268,388 (1.03)
8	#Wuhan	35,142 (6.86)	17,230 (6.21)	24,242 (1.50)	17,421 (0.77)	19,081 (0.48)	12,770 (0.34)	11,467 (0.28)	13,648 (0.30)	31,707 (0.65)	182,708 (0.70)
9	#WuhanVirus	10,301 (2.01)	4301 (1.55)	5522 (0.34)	5825 (0.26)	47,418 (1.19)	30,699 (0.81)	28,420 (0.71)	22,942 (0.51)	22,690 (0.46)	178,118 (0.69)
10	#coronavirusitalia	—	—	18,838 (1.17)	24,697 (1.08)	46,517 (1.16)	11,694 (0.31)	9828 (0.24)	9167 (0.20)	11,499 (0.23)	132,240 (0.51)
11	#sarscov2	369 (0.07)	4881 (1.76)	14,964 (0.93)	16,913 (0.74)	19,615 (0.49)	11,400 (0.30)	11,915 (0.30)	16,710 (0.37)	22,215 (0.45)	118,982 (0.46)
12	#2019nCov	18,041 (3.52)	5496 (1.98)	10,226 (0.63)	11,172 (0.49)	6330 (0.16)	2389 (0.06)	2471 (0.06)	3167 (0.07)	2939 (0.06)	62,231 (0.24)
13	#coronavirusitaly	—	—	3379 (0.21)	12,238 (0.54)	15,137 (0.38)	7082 (0.19)	3391 (0.08)	2562 (0.06)	1750 (0.04)	45,539 (0.18)
14	#nCov2019	11,426 (2.23)	2718 (0.98)	4255 (0.26)	3799 (0.17)	3951 (0.10)	2053 (0.05)	1043 (0.03)	1206 (0.03)	727 (0.01)	31,178 (0.12)
15	#nCov	6536 (1.28)	1265 (0.46)	2086 (0.13)	1831 (0.08)	2935 (0.07)	1385 (0.04)	565 (0.01)	856 (0.02)	535 (0.01)	17,994 (0.07)
16	#nCov19	1618 (0.32)	645 (0.23)	2382 (0.15)	1707 (0.07)	4151 (0.10)	1893 (0.05)	979 (0.02)	1166 (0.03)	768 (0.02)	15,309 (0.06)
Total	N/A[d]	512,040 (1.97)	277,422 (1.07)	1,616,632 (6.23)	2,276,938 (8.78)	3,996,999 (15.41)	3,801,512 (14.66)	4,025,003 (15.52)	4,530,376 (17.47)	4,895,498 (18.88)	25,932,420 (100.00)

[a]During this 7-day interval technical issues occurred for approximately 3 days.

[b]During this 7-day interval technical issues occurred for approximately 1 day.

[c]No data available.

[d]N/A: not applicable.

Figure 2. Number of tweets per day for the hashtags ranked 1-8 (see Table 2) between February 9, 2020, and April 11, 2020, on a logarithmic scale. The capital letter "A" represents the naming of the disease by the World Health Organization on February 11, 2020. Blue rectangles: No tweets were collected between February 20 and 22 as well as between March 28 and 29 due to technical issues.

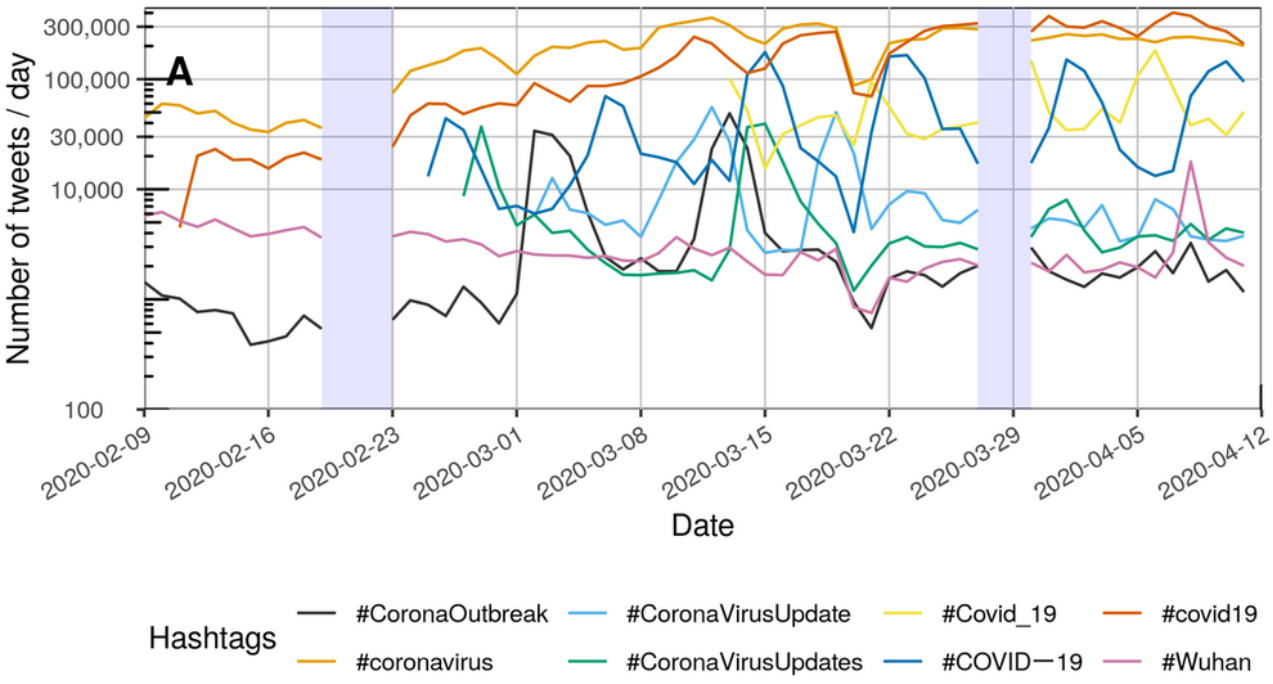
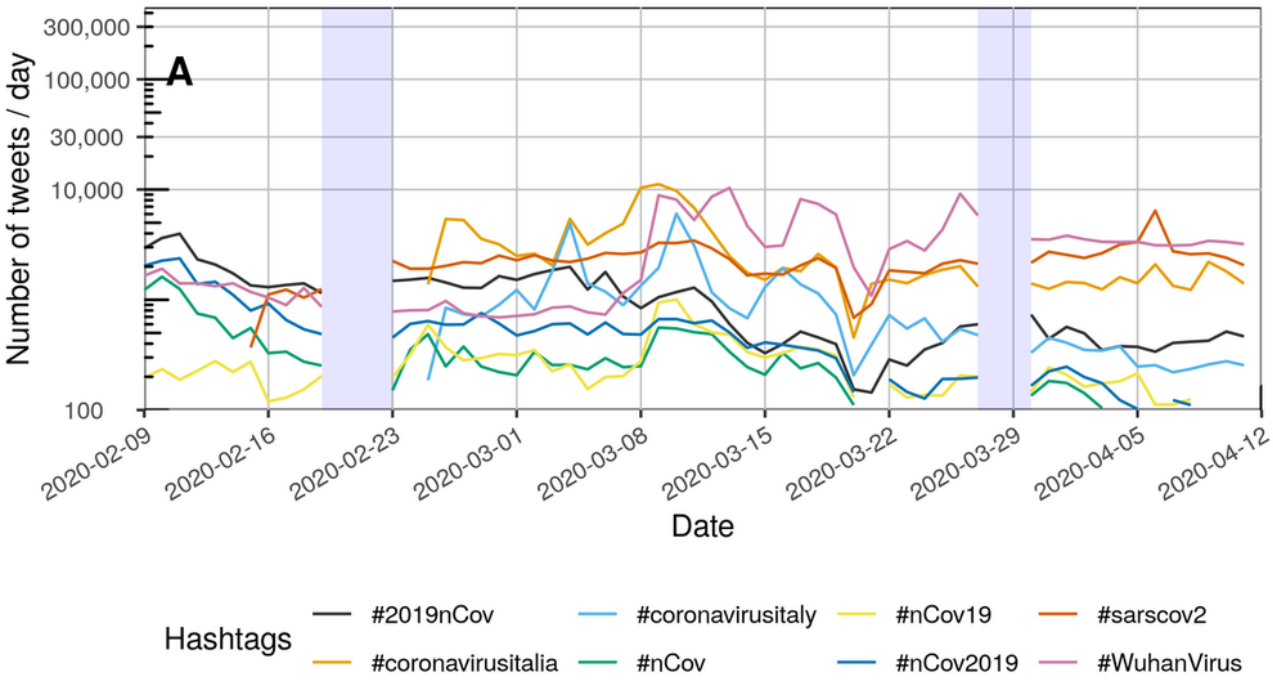


Figure 3. Number of tweets per day for the hashtags ranked 9-16 (see Table 2) between February 9, 2020, and April 11, 2020, on a logarithmic scale. The capital letter "A" represents the naming of the disease by the World Health Organization on February 11, 2020. Blue rectangles: No tweets were collected between February 20 and 22 as well as between March 28 and 29 due to technical issues.



Geographical Variations of Tweets

In the beginning of February 2020, the SARS-CoV-2 epidemic spread over Europe. The northern regions of Italy especially had a massive outbreak of COVID-19 [9]. To investigate the tweets' volume spread, all tweets that contained geographic coordinates were included in this subanalysis. Longitude and latitude information were available for 4.40% (957,947/21,755,802) of the tweets in the study database; filtered for the longitude and latitude representing the geographical borders of Europe, 29.83% (285,763/957,947) of tweets qualified. Each tweet was plotted in a geographical map of

Europe for each 7-day interval in the observation period (see Figure 4).

For an animated video that covers the observation period between February 9, 2020, and April 11, 2020, see Multimedia Appendix 3. For a high-resolution collection of the subplots in Figure 4, see Multimedia Appendix 4. In addition, Figure 5 presents a cumulative plot of all 285,763 tweets that provided geolocation information for the European continent. More tweets could be observed in the vicinity of countries' capitals (eg, Paris, Madrid, Vienna, or Berlin) or in densely populated areas such as the Benelux Union or South England. A higher number of tweets with geolocations was observed in Central and Western European countries than compared to Eastern Europe. Interestingly, tweet volumes in Turkey seemed to be higher than in surrounding countries.

Figure 4. Geolocation information of COVID-19–related tweets depicted for each 7-day interval. From top left (February 9, 2020, to February 15, 2020) to bottom right (April 5, 2020, to April 11, 2020). A single red dot denotes one tweet. Tweets with the same geographical information are plotted on top of each other. *No tweets were collected between February 20, 2020, and February 22, 2020, and between March 28, 2020, and March 29, 2020, due to technical issues.

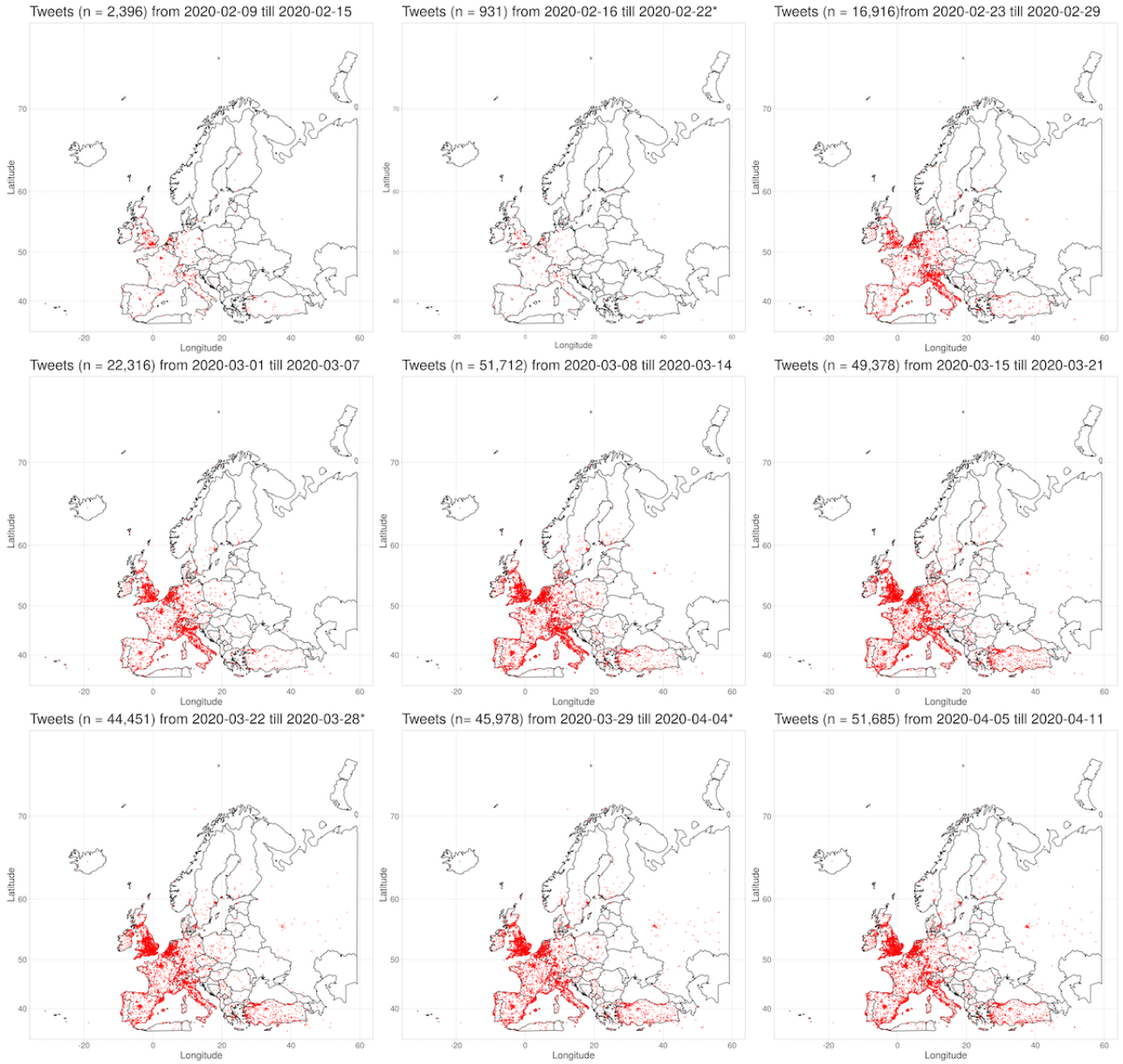
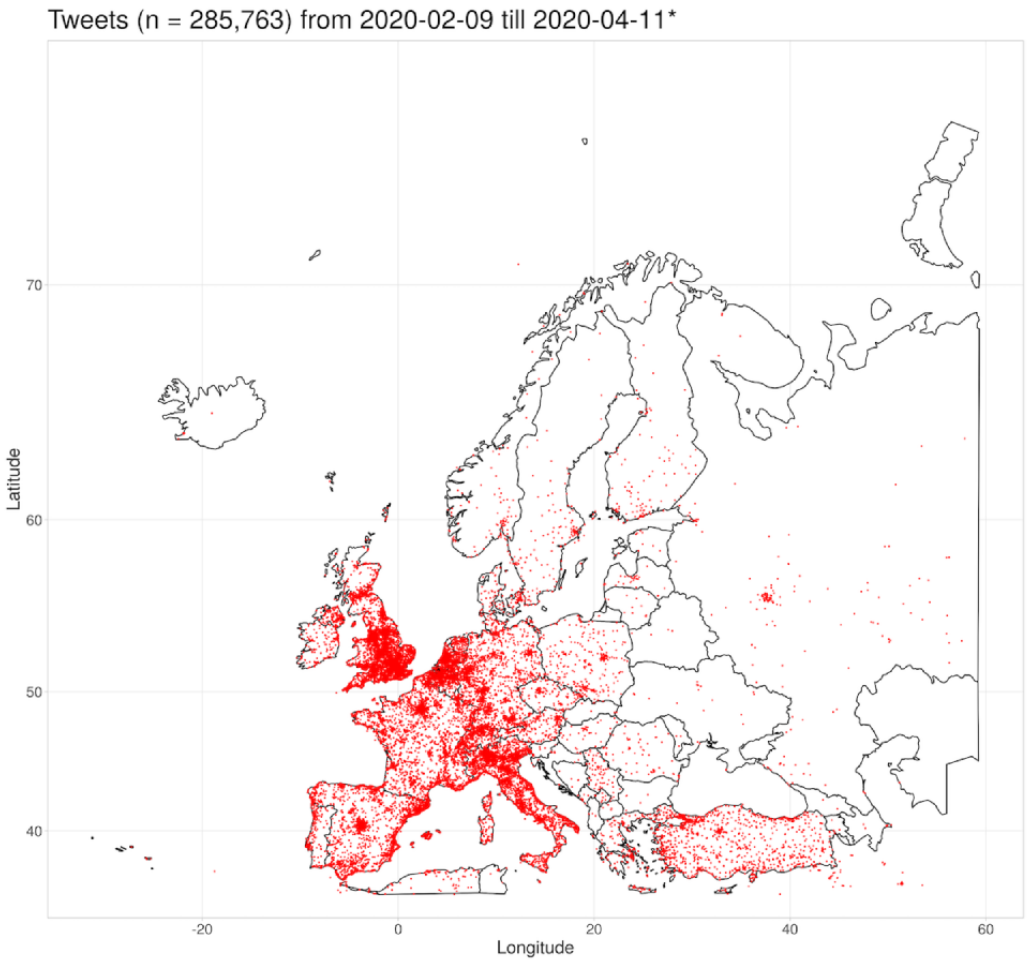


Figure 5. Cumulative depiction of all tweets in European countries between February 9 and April 11, 2020. Each red dot denotes one tweet. Tweets with the same geographical information are plotted on top of each other. *No tweets were collected between February 20, 2020, and February 22, 2020, and between March 28, 2020, and March 29, 2020, due to technical issues.



Link Category Analysis

The most prevalent (n=250) domain-aggregated URLs were categorized by three researchers independently according to the categories introduced by Chew and Eysenbach [31]. These URLs accounted for 46.38% (3,596,538/7,753,841) of all shared resources in our study database. The three researchers achieved a PA of 0.628 and a Fleiss κ of 0.639. According to Landis and Koch [50], these κ values correspond to a "substantial agreement." In 4.4% (11/250) of the cases, no majority vote was achieved and those were subsequently cleared by discussion among all of the authors. The link category "Not Accessible" was not selected, as all domains were accessible by every researcher.

Table 3 presents the top 50 shared domains and their link categories, and the occurrences of each domain in the study database. The complete list of the top 250 can be found in Multimedia Appendix 5. The most frequently shared resources originated from various social media platforms and are represented in the ranks 1-7. Cross-linking resources on social media (even on the same platform [ie, Twitter]) could be observed. The most prevalent category in the top 50 was "Mainstream or Local News." The resources of the newspapers The Guardian and the New York Times were the leading domains in this category, followed by the broadcasting services CNN and BBC. Only two domains qualified for the category "Government or Public Health" in the top 50: CDC at rank 25 and the WHO at rank 27. No scientific resource was contained in the top 50. The first occurrence was the British journal Nature at rank 116.

The relative proportion of tweets that shared references to external resources varied during the study period. A longitudinal subanalysis revealed a constant trend without major peaks for each day of the study (see Figure 6).

Table 3. Categorized top 50 shared website domains. Total number of occurrences of external references (n=7,753,841).

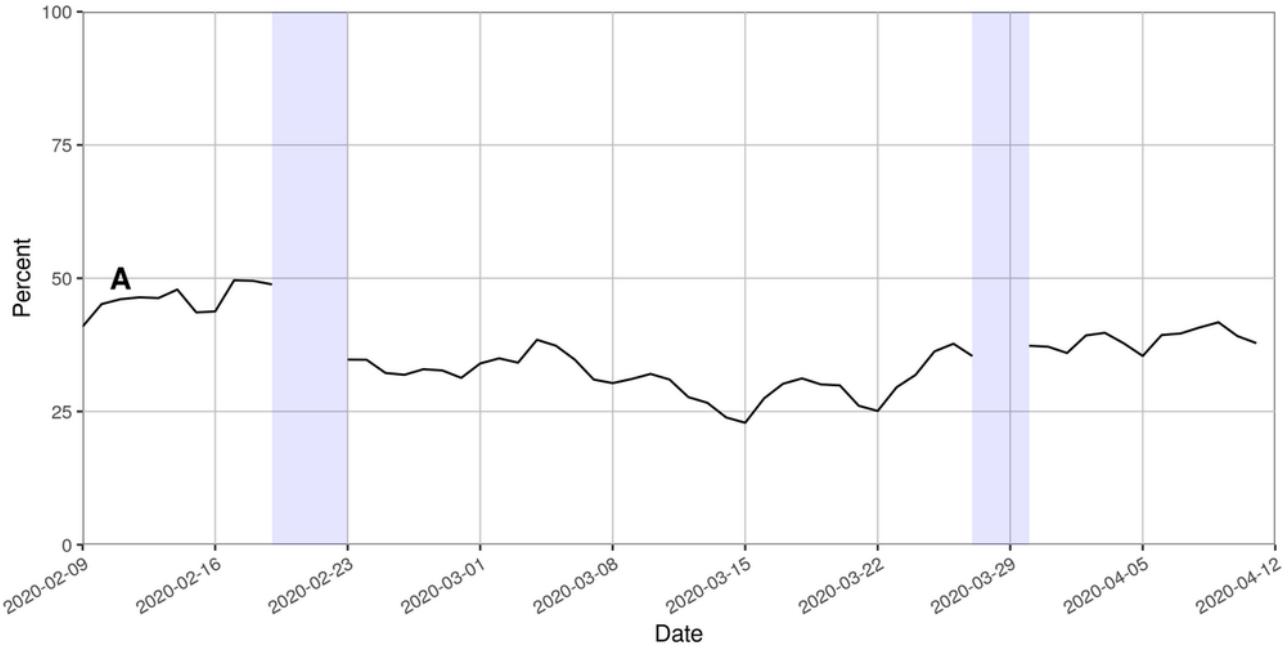
Link category and rank	Domain	Occurrences, n (%)
Mainstream or local news		
8	theguardian.com	52,733 (0.68)
11	nytimes.com	42,735 (0.55)
13	cnn.com	35,494 (0.46)
14	bbc.co.uk	28,286 (0.36)
15	washingtonpost.com	27,316 (0.35)
19	bbc.com	25,853 (0.33)
28	nyti.ms	17,720 (0.23)
30	reuters.com	17,589 (0.23)
33	cnbc.com	17,321 (0.22)
35	bloomberg.com	16,330 (0.21)
37	elpais.com	15,888 (0.20)
38	ouest-france.fr	14,976 (0.19)
40	francetvinfo.fr	14,609 (0.19)
41	scmp.com	14,072 (0.18)
43	reut.rs	13,637 (0.18)
46	forbes.com	13,242 (0.17)
48	nypost.com	12,464 (0.16)
49	businessinsider.com	12,433 (0.16)
News blog, feed, or niche news		
17	medium.com	26,201 (0.34)
42	zazoom.it	13,926 (0.18)
45	zazoom.info	13,303 (0.17)
47	topicza.com	12,960 (0.17)
Government or public health		
25	cdc.gov	19,729 (0.25)
27	who.int	18,298 (0.24)
Personal blog		
23	wordpress.com	22,376 (0.29)
Social network		
1	twitter.com	378,508 (4.88)
2	youtu.be	365,716 (4.72)
3	instagram.com	290,336 (3.74)
5	youtube.com	144,502 (1.86)
6	facebook.com	95,166 (1.23)
7	linkedin.com	79,787 (1.03)
16	pscp.tv	26,823 (0.35)
Online store		
20	amzn.to	25,378 (0.33)
Scientific resource[a]		
—[b]	—	—

Link category and rank	Domain	Occurrences, n (%)
URL shortener[a]		
10	tinyurl.com	44,768 (0.58)
24	trib.al	21,409 (0.28)
Other		
4	paper.li	204,077 (2.63)
9	google.com	47,184 (0.61)
12	chng.it	41,316 (0.53)
18	fiverr.com	25,905 (0.33)
21	ift.tt	23,304 (0.30)
22	avaaz.org	22,938 (0.30)
26	arcgis.com	18,604 (0.24)
29	worldometers.info	17,670 (0.23)
31	yahoo.com	17,588 (0.23)
32	apple.news	17,584 (0.23)
34	openstream.co	16,442 (0.21)
36	goo.gl	16,310 (0.21)
39	joinzoe.com	14,827 (0.19)
44	shoutcast.com	13,635 (0.18)
50	dy.si	12,343 (0.16)

[a]Link category as extension of the list given in Chew and Eysenbach [31].

[b]No domain qualified for a rank below or equal to 50. The full listing with all scientific resources under this category is found in Multimedia Appendix 5.

Figure 6. The relative proportion of tweets with links to external resources. The capital letter "A" represents the naming of the disease by the World Health Organization on February 11, 2020. Blue rectangles: No tweets were collected between February 20 and 22 as well as between March 28 and 29 due to technical issues.



Discussion

Principal Results

The COVID-19–related tweet volume observed in this study increased constantly during the weeks of February 2020. However, this study did not investigate whether tweet volumes correlated with infection or death rates in different European countries. It can be assumed that increasing SARS-CoV-2 infection figures correlate with increased public interest and engagement on social media platforms. In addition to rising infection rates, several other factors such as the death of a celebrity due to COVID-19 could have increased public interest in the progress of the pandemic. In this context, the hashtag-specific analysis revealed that #COVID—19 and #Covid_19 in particular were fluctuating periodically without a clear connection to specific events.

The analysis over time revealed that the first Twitter *hot spots* in Europe developed not only in the capital cities of London and Paris but also in the region of Milan, Italy. The northern regions of Italy showed a sharp increase in tweet volume in the beginning and middle of February.

As the epidemic spread further over Europe, an increase of tweet volume over most of Western and Central Europe could be observed. However, the increase of tweets was not that prevalent in Eastern European countries (eg, Czech Republic, Poland, Romania) and in Southeastern Europe (eg, Serbia, Croatia). The public in Turkey increased their Twitter activity around the second week of March (see Figure 4) when the first COVID-19 case was officially confirmed by the Turkish health authorities.

The most frequently shared resources linked to various social media platforms and were represented by the ranks 1-7. The CDC website reached the 25th rank and the WHO website the 27th rank in the top 250 shared domain analysis. By contrast, the first occurrence of a prevalent scientific source is Nature at rank 116. Nevertheless, it was surprising that a high-class journal such as Nature was only directly referenced in 0.08% (6043/7,753,841) of the links to external resources shared on Twitter. Likewise, this finding applied for other scientific sources: Science (rank 147; 4615/7,753,841, 0.06%), The New England Journal of Medicine (rank 154; 4405/7,753,841, 0.06%), medRxiv (rank 170; 4123/7,753,841, 0.05%), and Johns Hopkins University (rank 199; 3586/7,753,841, 0.05%). Even with these numbers at hand, it remains an open question whether direct references to scientific sources should be included more actively for the purpose of public health communication on Twitter or not, given that a broad media coverage, which translates scientific language for a broader audience, seems necessary to disseminate important COVID-19–related research results to the public.

Limitations

Many social media platforms were used to share personal opinions, information, and news or stories around a particular topic. In the context of the COVID-19 pandemic, different platforms were in the public interest. In the setting of this study, only contributions on the platform Twitter were investigated and public disease-related data was analyzed. For this reason, the reported findings may not be mapped and applicable to other social networks such as Facebook, Reddit, or YouTube.

Amplification of particular tweets can increase visibility of certain resources shared by users. In this context, retrospective queries and related analyses were limited by the capabilities as given by the “Post, retrieve, and engage with Tweets” endpoint [51] of the Twitter API with the “standard” access level. For this reason, the authors could not update the data collection of tweets at the end of the study period regarding retweets and likes. Consequently, no deep analysis of certain resources’ popularity could be conducted. It remains a future task to analyze these relationships, even though it seems impractical given the “standard” access level.

In addition, the Twitter API ensures, that privacy of nonpublic tweets is respected. This is why the “Filter realtime Tweets” endpoint [38] does not return privately posted tweets. Therefore, those users and tweets could not be included in this study. Yet, it is estimated that only a small proportion of Twitter users configure their account as fully private.

Most Twitter users configure their individual privacy settings to hide their personal geolocation. For this reason, the analyses of geographical variations was limited to a comparatively low amount of data. In the context of our study, geolocation data was only available for 4.40% (957,947/21,755,802) of the collected tweets. However, this subsample still accounts for around 1 million tweets in total. In this context, the study found Eastern European users of Twitter to be less engaged during the study period. This might originate from low Twitter adoption rates in Eastern Europe [52].

This study investigated the SARS-CoV-2 outbreak situation in Europe with a specific interest. This originated from the epidemic spread of the virus in Europe, starting in Italy [9-11]. This spread was accompanied by increased media coverage and public interest in Europe [21,53,54] and worldwide [55]. Researchers of different disciplines started analyzing regional differences among European countries such as Italy, Spain, France, Germany, or Austria [17]. In the context of this study, this motivated RQ2 and the specific analyses as reported. It is worth emphasizing, however, that the European languages cannot be mapped easily to very fine-grained country borders on classical maps. Therefore, the European region had to be approximated by using the geo-information and the bounding box, as described in the Methods section, defined by corresponding geo-coordinates. This resulted in a subsample of 285,763 tweets for the European region subanalysis.

Furthermore, the special European focus was initiated by monitoring the worsening of the severe SARS-CoV-2 outbreak in the Northern Italy regions of Lombardy and Emilia Romagna [11,37]. For this reason, the authors decided to add two Italy-specific hashtags that were prevalent around the third week of February 2020, as reported by Twitter trends at that time. However, it should be noted that those two hashtags account for only 0.18% (#coronovirusitaly; 45,439/25,932,420) and 0.51% (#coronavirusitalia; 132,240/25,932,420) of all hashtag usages in the study’s data collection. As the spread of the virus progressed over several countries in Europe, many other country-specific hashtags appeared in Twitter trend statistics.

The authors decided to avoid including all possible variations and country-specific subhashtags. This possibility limits comparisons among different countries in Europe. Nevertheless, the generic hashtags for COVID-19 remained stable over the full study period. Thus, tweets can be found in the data collection for every European country.

The collection of data was conducted in real time. Sadly, due to technical issues on February 20-22 and March 28 and 29, 2020, data could not be collected during these time spans. The issue in February originated from a loss of connectivity to the *PostgreSQL* study database, which was not discovered for around 48 hours during a weekend. A second, technical issue in late March resulted from an unexpected memory allocation problem on the processing server. Once the issue was resolved by a software patch, the system was capable of collecting and storing tweets correctly again.

Comparison With Prior Work

During the 2009 H1N1 flu pandemic, Chew and Eysenbach [31] applied the *infoveillance* concept for a content analysis for which they "archived over 2 million Twitter posts containing keywords 'swine flu', 'swineflu' and/or 'H1N1.'" The authors analyzed diseases-related trends, the origin of shared resources, and the sentiment expressed in swine flu tweets. In our study, more than 20 million COVID-19–related tweets were analyzed for temporal or geographical characteristics and trends as well as for the link category of external resources. In the 2009 study [31], the authors found that "government and health agencies were only linked 1.5% of the time." For a top 250 list, this low proportion is confirmed by our findings (78,786/7,753,841, 1.02%). Chew and Eysenbach [31] found that "news websites were the most popular sources (23.2%)." Likewise, our analysis revealed that the link category "Mainstream and local news" was represented by 11.97% (928,467/7,753,841), which was substantially lower than in 2009. In this context, our findings suggest that Twitter users cross-reference to Twitter itself or to other social media platforms (1,406,419/7,753,841, 18.14%), whereas this group was reported to represent only 2% of the corresponding category in the study by Chew and Eysenbach [31]. Moreover, the authors of the H1N1 flu study reported that "61.8% of all tweets had links [..]." In our study, this proportion was found to be 34.16% (7,431,226/21,755,802), which was substantially lower.

Fu et al [32] analyzed how people reacted to the Zika epidemic in the Americas from 2015 to 2016. The authors analyzed 132,033 tweets with the key word "zika" written in the languages English, Spanish, and Portuguese via the Twitter API. The authors reported, that the top ranked shared resources originated from social media platforms such as "Facebook, Instagram, Twitter, YouTube, LinkedIn, Tumblr, the blogging site WordPress, [..] which accounted for 26% of all domains." This could be confirmed by our results, as social media platforms were ranked on the positions 1-7 accounting for 18.14% (1,406,419/7,753,841) of all shared resources. In the Zika study, the CDC and the WHO accounted for 0.06% and 0.05%, respectively. This corresponded to a 90th and 140th rank, respectively, compared to a 25th (19,729/7,753,841, 0.25%) and a 27th (18,298/7,753,841, 0.24%) rank, respectively, in our analysis on shared resources. The comparison suggests that public health-related material provided via the CDC or the WHO was shared more frequently than during the Zika outbreak between 2015 and 2016. This increase might originate from multiple reasons: improved, timely provisioning of disease-related material by either the CDC, the WHO, or both; higher awareness of the public for quality aspects of material and evidence-based sources; or the use of easy language or easily comprehendible infographics by the public health teams of the CDC, the WHO, or both.

Abd-Alrazaq et al [16] analyzed the content and sentiment of about 2.8 million COVID-19–related tweets, retrieved via the Twitter standard search API, written in the English language. By contrast, our study design made use of Twitter's real-time Streaming API, which allows for a constant intake to the study database. In [16], the authors made use of the search terms "corona," "2019-nCov," and "COVID-19." In our study, we monitored 16 hashtags for a time span of 9 weeks. This resulted in a data collection with a total of approximately 21.8 million topic-related tweets. With our analysis framework, we were able to monitor specific regions (Europe) and countries, in particular the SARS-CoV-2 outbreak in Italy.

Future Directions

This study demonstrates how COVID-19–related tweets can be analyzed for a certain region (Europe). With the continuous progression of the pandemic situation, which is to be expected in the next months worldwide, further regions should be analyzed in-depth. Therefore, the authors encourage other researchers to contribute their analyses with a special focus on regions such as Africa, South and North America, or Asia. Moreover, different analysis techniques can be leveraged to learn more about what users share in the current pandemic situation. For this purpose, one could use sentiment analysis or conduct social network graph analysis to uncover patterns that might be hidden in the data. Sentiment analysis is of particular interest, as it could reveal differences between regions or even between several countries, such as demonstrated by Abd-Alrazaq et al [16] for tweets written in the English language [16].

Long-term Twitter monitoring based on geographical data could be a supporting tool for local health authorities. With an average tweet volume per city, region, or even country, significant peaks well above the 7-day average could be reported to official institutions quickly in an electronic, interoperable format. In this sense, an automated analysis tool could be an extension of our software components to capture pandemic-related tweets in real time.

Future studies should also focus on the origin and trustworthiness of shared resources. Monitoring the spread of fake news during a pandemic situation seems of particular importance [24,26]. Timely measures to fight and reduce the spread of COVID-19 misinformation could thus be supported. In addition, it would be beneficial to analyze and uncover bot networks spreading COVID-19–related misinformation. In this study, we could uncover periodicity of at least one hashtag (#COVID—19). This might be linked to a hidden bot network, which justifies further investigation.

In future work, the authors intend to publish the data collection according to the Developer Agreement and Policy of Twitter [56]. Other researchers might analyze this data collection with a different focus or with their own scientific perspective. By providing this data set, the requirement of providing one's own technical infrastructure would pose no barrier for non–computer science disciplines. We hope to provide this data set publicly, regularly updated in 1 week intervals.

Conclusions

The naming of the disease by the WHO on February 11, 2020 [1], was a major signal to address the public audience with a public health response via social media platforms. The volume of #covid19-related tweets increased after the WHO announcement and stabilized at the end of March at a high level.

During the spread of the SARS-CoV-2 virus in Europe between February 2020 and early April 2020, the public interest and media coverage increased rapidly. Consequently, the engagement of citizens on social media platforms rose accordingly. On April 16, 2020, Dr Hans Kluge, the WHO director for Europe, characterized the situation as "we remain in the eye of the [COVID-19] storm" [57]. The findings of this study allow for a detailed analysis for the European region and how citizens of different European countries shared their opinions, experiences, and concerns on Twitter. The detailed maps of Europe are available for each 7-day interval starting on February 9, 2020.

Social media platforms were ranked at the position of 1-7, counting for 18.14% (1,406,419/7,753,841) of all shared resources. The CDC website reached the 25th rank (19,729/7,753,841, 0.25%) and the WHO website the 27th rank (18,298/7,753,841, 0.24%) of the top 250 shared domain analysis. Future studies should focus on the origin and trustworthiness of shared resources, as monitoring the spread of fake news during a pandemic situation is of particular importance.

Acknowledgments

The authors would like to thank Prof em Dr Wilhelm Werner of Heilbronn University for his mathematical advice in the preparation of this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Listing of all hashtags included in the study.
[DOCX File , 16 KB - jmir_v22i8e19629_app1.docx]

Multimedia Appendix 2
Temporal variations of hashtag frequencies between February 9, 2020, and April 11, 2020.
[DOCX File , 2002 KB - jmir_v22i8e19629_app2.docx]

Multimedia Appendix 3
Animated video of the geographical variation of tweets in the European countries.
[MP4 File (MP4 Video), 730 KB - jmir_v22i8e19629_app3.mp4]

Multimedia Appendix 4
Seven-day interval plots of the geographical variation of tweets in all European countries.
[DOCX File , 2149 KB - jmir_v22i8e19629_app4.docx]

Multimedia Appendix 5
Categorized top 250 shared (web site) domains.
[DOCX File , 22 KB - jmir_v22i8e19629_app5.docx]

References

1. Naming the coronavirus disease (COVID-19) and the virus that causes it. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it [accessed 2020-03-09]
2. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020 Feb 3;579(7798):270-273. [doi: 10.1038/s41586-020-2012-7]
3. Wu F, Zhao S, Yu B, Chen Y, Wang W, Song Z, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020 Feb 3;579(7798):265-269. [doi: 10.1038/s41586-020-2008-3]

4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
5. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
6. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
7. Young BE, Ong SWX, Kalimuddin S, Low JG, Tan SY, Loh J, Singapore 2019 Novel Coronavirus Outbreak Research Team. Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. JAMA 2020 Mar 03:1488-1494 [FREE Full text] [doi: 10.1001/jama.2020.3204] [Medline: 32125362]
8. Day M. Covid-19: surge in cases in Italy and South Korea makes pandemic look more likely. BMJ 2020 Feb 25;368:m751. [doi: 10.1136/bmj.m751] [Medline: 32098875]
9. Spina S, Marrazzo F, Migliari M, Stucchi R, Sforza A, Fumagalli R. The response of Milan's Emergency Medical System to the COVID-19 outbreak in Italy. Lancet 2020 Mar;395(10227):e49-e50. [doi: 10.1016/s0140-6736(20)30493-1]
10. Grasselli G, Pesenti A, Cecconi M. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. JAMA 2020 Mar 13:A. [doi: 10.1001/jama.2020.4031] [Medline: 32167538]
11. Saglietto A, D'Ascenzo F, Zoccai GB, De Ferrari GM. COVID-19 in Europe: the Italian lesson. Lancet 2020 Apr;395(10230):1110-1111. [doi: 10.1016/s0140-6736(20)30690-5]
12. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? Lancet 2020 Apr;395(10231):1225-1228. [doi: 10.1016/s0140-6736(20)30627-9]
13. Garfin D, Silver RC, Holman EA. The novel coronavirus (COVID-2019) outbreak: amplification of public health consequences by media exposure. Health Psychol 2020 May;39(5):355-357. [doi: 10.1037/hea0000875] [Medline: 32202824]
14. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study. JMIR Public Health Surveill 2020 Apr 21;6(2):e18700 [FREE Full text] [doi: 10.2196/18700] [Medline: 32293582]
15. Llewellyn S. Covid-19: how to be careful with trust and expertise on social media. BMJ 2020 Mar 25;368:m1160. [doi: 10.1136/bmj.m1160] [Medline: 32213480]
16. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
17. Hodcroft EB. Preliminary case report on the SARS-CoV-2 cluster in the UK, France, and Spain. Swiss Med Wkly 2020 Feb 24;150(9-10):A. [doi: 10.4414/smw.2020.20212] [Medline: 32227799]
18. Lescure F, Bouadma L, Nguyen D, Parisey M, Wicky P, Behillil S, et al. Clinical and virological data of the first cases of COVID-19 in Europe: a case series. Lancet Infect Dis 2020 Jun;20(6):697-706. [doi: 10.1016/s1473-3099(20)30200-0]
19. Moss P, Barlow G, Easom N, Lillie P, Samson A. Lessons for managing high-consequence infections from first COVID-19 cases in the UK. Lancet 2020 Mar;395(10227):e46. [doi: 10.1016/s0140-6736(20)30463-3]
20. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N Engl J Med 2020 Mar 05;382(10):970-971. [doi: 10.1056/nejmc2001468]
21. Escritt T, Carrel P. New coronavirus cases in Germany, one man in critical condition. Reuters. 2020 Feb 25. URL: https://www.reuters.com/article/us-china-health-germany-idUSKBN20J2K2 [accessed 2020-04-24]
22. Danis K, Epaulard O, Bénet T, Gaymard A, Campoy S, Botelho-Nevers E, Investigation Team. Cluster of coronavirus disease 2019 (COVID-19) in the French Alps, February 2020. Clin Infect Dis 2020 Jul 28;71(15):825-832 [FREE Full text] [doi: 10.1093/cid/ciaa424] [Medline: 32277759]
23. Jones J. Spanish authorities confirm Briton is country's second coronavirus case. Reuters. 2020 Feb 09. URL: https://www.reuters.com/article/us-china-health-spain-idUSKBN20309T [accessed 2020-04-25]
24. Merchant RM, Lurie N. Social media and emergency preparedness in response to novel coronavirus. JAMA 2020 Mar 23:A. [doi: 10.1001/jama.2020.4469] [Medline: 32202611]
25. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
26. Chou WS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
27. Novel coronavirus(2019-nCoV) situation report - 13. World Health Organization. 2020 Feb 02. URL: https://apps.who.int/iris/bitstream/handle/10665/330778/nCoVsitrep02Feb2020-eng.pdf
28. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
29. Limaye RJ, Sauer M, Ali J, Bernstein J, Wahl B, Barnhill A, et al. Building trust while influencing online COVID-19 content in the social media world. Lancet Digital Health 2020 Jun;2(6):e277-e278. [doi: 10.1016/s2589-7500(20)30084-4]

30. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
31. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
32. Fu K, Liang H, Saroha N, Tse ZTH, Ip P, Fung IC. How people react to Zika virus outbreaks on Twitter? A computational content analysis. Am J Infect Control 2016 Dec 01;44(12):1700-1702. [doi: 10.1016/j.ajic.2016.04.253] [Medline: 27566874]
33. Guest JL, Del Rio C, Sanchez T. The three steps needed to end the COVID-19 pandemic: bold public health leadership, rapid innovations, and courageous political will. JMIR Public Health Surveill 2020 Apr 06;6(2):e19043 [FREE Full text] [doi: 10.2196/19043] [Medline: 32240972]
34. Southwell BG, Thorson EA, Sheble L. The persistence and peril of misinformation. Am Sci 2017;105(6):372. [doi: 10.1511/2017.105.6.372]
35. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
36. Mast fire probe amid 5G coronavirus claims. BBC. 2020 Apr 04. URL: https://www.bbc.com/news/uk-england-52164358# [accessed 2020-04-24]
37. Paterlini M. On the front lines of coronavirus: the Italian response to covid-19. BMJ 2020 Mar 16;368:m1065. [doi: 10.1136/bmj.m1065] [Medline: 32179517]
38. Filter realtime tweets. Twitter. URL: https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter [accessed 2020-03-09]
39. Twitter4J. URL: http://twitter4j.org/en/ [accessed 2020-01-04]
40. PostgreSQL. URL: https://www.postgresql.org/ [accessed 2020-01-04]
41. Geo objects. Twitter. URL: https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/geo-objects [accessed 2020-04-24]
42. crawler4j. GitHub. URL: https://github.com/yasserg/crawler4j [accessed 2020-01-04]
43. Lehmberg O, Meusel R, Bizer C. Graph structure in the web: aggregated by pay-level domain. In: WebSci '14: Proceedings of the 2014 ACM Conference on Web Science. 2014 Jun Presented at: WebSci '14: ACM Web Science Conference; June 2014; Bloomington, IN p. 119-128. [doi: 10.1145/2615569.2615674]
44. Public Suffix List. URL: https://publicsuffix.org/ [accessed 2020-04-23]
45. Lombard M, Snyder-Duch J, Bracken CC. Content analysis in mass communication: assessment and reporting of intercoder reliability. Human Comm Res 2002 Oct;28(4):587-604. [doi: 10.1111/j.1468-2958.2002.tb00826.x]
46. Fleiss JL. Measuring nominal scale agreement among many raters. Psychological Bull 1971;76(5):378-382. [doi: 10.1037/h0031619]
47. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer; 2009.
48. About verified accounts. Twitter. URL: https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts [accessed 2020-04-23]
49. Evaluating language identification performance. Twitter. 2015 Nov 16. URL: https://blog.twitter.com/engineering/en_us/a/2015/evaluating-language-identification-performance.html [accessed 2020-04-23]
50. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]
51. GET statuses/lookup. Twitter. URL: https://developer.twitter.com/en/docs/tweets/post-and-engage/api-reference/get-statuses-lookup [accessed 2020-04-24]
52. Leetaru K, Wang S, Padmanabhan A, Shook E. Mapping the global Twitter heartbeat: the geography of Twitter. FM 2013 May 06;18(5):A. [doi: 10.5210/fm.v18i5.4366]
53. Reynolds M, Weiss S. How coronavirus started and what happens next, explained. Wired. 2020 May 27. URL: https://www.wired.co.uk/article/china-coronavirus [accessed 2020-05-27]
54. COVID-19: Commission sets out European coordinated response to counter the economic impact of the Coronavirus. European Commission. 2020 Mar 13. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_459 [accessed 2020-04-24]
55. Sheikh K, Rabin RC. The coronavirus: what scientists have learned so far. The New York Times. 2020 Aug 05. URL: https://www.nytimes.com/article/what-is-coronavirus.html [accessed 2020-08-05]
56. Developer agreement and policy. Twitter. URL: https://developer.twitter.com/en/developer-terms/agreement-and-policy [accessed 2020-04-24]
57. United NRICFWE. COVID-19: “We remain in the eye of the storm”- WHO Europe. United Nations: Regional Information Centre for Western Europe. 2020 Apr 16. URL: https://unric.org/en/covid-19-we-remain-in-the-eye-of-the-storm-who-europe/ [accessed 2020-04-24]

Abbreviations

API: application programming interface
CDC: Centers for Disease Control and Prevention
CEST: Central European Summer Time
PA: percent agreement
RQ: research question
WHO: World Health Organization

Edited by G Eysenbach; submitted 25.04.20; peer-reviewed by J Zeng, J Vidal-Alaball; comments to author 20.06.20; revised version received 23.06.20; accepted 11.08.20; published 28.08.20.

Please cite as:
Pobiruchin M, Zowalla R, Wiesner M
Temporal and Location Variations, and Link Categories for the Dissemination of COVID-19–Related Information on Twitter During the SARS-CoV-2 Outbreak in Europe: Infoveillance Study
J Med Internet Res 2020;22(8):e19629
URL: http://www.jmir.org/2020/8/e19629/
doi:10.2196/19629
PMID:32790641

Original Paper

An Electronic Data Capture Framework (ConnEDCt) for Global and Public Health Research: Design and Implementation

Caleb J Ruth[1*], BS; Samantha Lee Huey[2*], BS, PhD; Jesse T Krisher[2], BS, CPH; Amy Fothergill[2], MPH; Bryan M Gannon[2,3], PhD; Camille Elyse Jones[2], MPH; Elizabeth Centeno-Tablante[2], MS; Laura S Hackl[2], PhD; Susannah Colt[2], PgDip, PhD; Julia Leigh Finkelstein[2,3], MPH, ScD; Saurabh Mehta[2,3], MBBS, ScD

[1]Data Performance LLC, Ithaca, NY, United States

[2]Division of Nutritional Sciences, Cornell University, Ithaca, NY, United States

[3]Institute for Nutritional Sciences, Global Health, and Technology, Cornell University, Ithaca, NY, United States

[*]these authors contributed equally

Corresponding Author:
Saurabh Mehta, MBBS, ScD
Division of Nutritional Sciences
Cornell University
314 Savage Hall
Ithaca, NY
United States
Phone: 1 (607) 255 2640
Email: smehta@cornell.edu

Abstract

Background: When we were unable to identify an electronic data capture (EDC) package that supported our requirements for clinical research in resource-limited regions, we set out to build our own reusable EDC framework. We needed to capture data when offline, synchronize data on demand, and enforce strict eligibility requirements and complex longitudinal protocols. Based on previous experience, the geographical areas in which we conduct our research often have unreliable, slow internet access that would make web-based EDC platforms impractical. We were unwilling to fall back on paper-based data capture as we wanted other benefits of EDC. Therefore, we decided to build our own reusable software platform. In this paper, we describe our customizable EDC framework and highlight how we have used it in our ongoing surveillance programs, clinic-based cross-sectional studies, and randomized controlled trials (RCTs) in various settings in India and Ecuador.

Objective: This paper describes the creation of a mobile framework to support complex clinical research protocols in a variety of settings including clinical, surveillance, and RCTs.

Methods: We developed ConnEDCt, a mobile EDC framework for iOS devices and personal computers, using Claris FileMaker software for electronic data capture and data storage.

Results: ConnEDCt was tested in the field in our clinical, surveillance, and clinical trial research contexts in India and Ecuador and continuously refined for ease of use and optimization, including specific user roles; simultaneous synchronization across multiple locations; complex randomization schemes and informed consent processes; and collecting diverse types of data (laboratory, growth measurements, sociodemographic, health history, dietary recall and feeding practices, environmental exposures, and biological specimen collection).

Conclusions: ConnEDCt is customizable, with regulatory-compliant security, data synchronization, and other useful features for data collection in a variety of settings and study designs. Furthermore, ConnEDCt is user friendly and lowers the risks for errors in data entry because of real time error checking and protocol enforcement.

(J Med Internet Res 2020;22(8):e18580) doi:10.2196/18580

KEYWORDS

data science; data collection; database management systems; global health; public health; data management; health information management; population surveillance; longitudinal studies; randomized controlled trial; Electronic Data Capture (EDC)

Introduction

Background

Our research on human nutrition and global health spans multiple countries, encompasses both interventional and observational study designs, and requires secure data collection and management. Noting the benefits of electronic data capture (EDC) over paper-based data capture [1,2], beginning in 2012, we searched for an EDC tool for our research projects on human health and were unable to find a package that met our needs. With several research projects on our roadmap that would be deployed in remote, international regions with complex longitudinal protocols, strict eligibility requirements, and distributed research teams, we aimed to use EDC over paper-based data capture to enforce our study protocols and provide better data security, data validation, and have instant access to data. In surveying EDC packages currently available on the market (Table 1), some had to be ruled out completely because of the unreliable internet coverage in areas we focused on for our studies (compared with Research Electronic Data Capture [REDCap], which had not yet released its mobile app when our initial research projects were being conducted in 2013 [3]). Other packages identified did not provide the required features specific to our study protocols, such as our strict data security specifications (compared with Open Data Kit [ODK]) [4]. To address this, we felt that we had the correct clinical and technical team in place to build our own EDC tools to solve our needs and potentially those of a large community of researchers in the future.

Objectives

As we developed our plans for creating the EDC tools for our research studies, we decided that instead of a custom tool for each study, a better investment would be one reusable platform that could be applied to each research project, with limited customizations for each iteration. The upfront investment would be higher, but subsequent deployments should build on the initial investment. Mobile phones and tablet devices are commonly used, even in lower-resource settings where our studies are located. We felt that an EDC tool for tablets would be easier for our local research staff to learn, more transportable, and less expensive to deploy than laptops. In this paper, we describe ConnEDCt, the EDC platform that we developed and successfully deployed for clinic-based cross-sectional studies, randomized controlled trials (RCTs), and surveillance projects.

Table 1. ConnEDCt, REDCap, and Open Data Kit electronic data capture systems.

Feature	ConnEDCt	REDCap[a]	REDCap Cloud	Open Data Kit 1 Suite, ODK-X[b] Suite
Requirement to use	FileMaker license	Nonprofit organization with sufficient IT[c] infrastructure. Join REDCap Consortium, license agreement with Vanderbilt University. Must submit new license to obtain a new REDCap system per each group of users	Fee-based hosting by a third-party company	Web access, user comfort in coding
Mobile app and platform	FileMaker Go on iPhone, iPad	REDCap mobile app; MyCap on iPhone, iPad, Android	REDCap Cloud mobile app on iPhone, iPad, Android	Yes: ODK Connect, Android app
Software and operating system	Claris FileMaker on Mac, Windows	Web server, database server, SMTP[d] email server, file server (optional) on any laptop	Web server, database server, SMTP email server, fileserver (optional) on any laptop	ODK Aggregate app, ODK Briefcase app, ODK Central app; XML documents created using ODK JavaRosa library/any laptop
Web based	Yes	Yes	Yes	Yes
Offline data capture	Yes	Yes	Yes	Yes
Data synchronization: software	Yes: MirrorSync [5]	Yes	Yes	Yes
Study designs	Customizable	Customizable	Customizable	Customizable
Study protocol enforcement for scheduled and unscheduled encounters	Yes	Not noted	Not noted	Not noted
Electronic-informed consent	Yes	Yes	Yes	Yes
Automated eligibility determination	Yes	Not noted	Not noted	Yes
Randomization	Yes: any form can be implemented	Yes: any form can be implemented	Yes: any form can be implemented	Yes: predetermined types allowed
Conditional CRFs[e] based on midline serial sampling or other criteria	Yes	Not noted	Not noted	Not noted
Open source	No	No	No	Yes
Regulatory-compliant security and encryption	Title 21 CFR[f] Part 11, HIPAA[g], GDPR[h]	HIPAA, Part 11, FISMA[i] standards (low, medium, or high), GDPR, depending on environment	Best practice security, Title 21 CFR Part 11 compliant, industry regulations, HIPAA compliant, data privacy technology	Security of third-party libraries are not vetted, require user's security staff to review libraries and source code on GitHub
Secure data collection	Yes	Yes	Yes	Encrypted form security; only transmissions over a secure HTTPS[j] connection are obscured from observers and prevent tampering in transmission
Export type	Excel, can be customized per user requirement	Excel, PDF, SPSS, SAS, Stata, R	Excel, PDF, SPSS, SAS, Stata, R	Export to CSV[k], JSON[l] (text only), KML[m] (for mapping applications)
Logic checks/concurrent error checking	Yes	Yes	Yes	Yes
Customizable user roles	Yes	Yes	Yes	Yes
Translation and cultural adaptation	Yes	Yes	Yes	Yes

Feature	ConnEDCt	REDCap[a]	REDCap Cloud	Open Data Kit 1 Suite, ODK-X[b] Suite
Signature capture	Yes	Yes	Yes	Yes
Ease of form construction	Create using FileMaker software on laptop interface; use Excel to record variable lists; methods adaptable per user's comfort level	Web-based designer, offline *data dictionary* file on Microsoft Word, Excel	Web-based designer, offline *data dictionary* file on Microsoft Word, Excel	Excel-based form creation (XLSForm), drag-and-drop form creation (ODK Build)
Customizable	Yes	Yes	Yes	Yes
Audit trail	Yes	Yes	Yes	Yes

[a]REDCap: Research Electronic Data Capture.

[b]ODK-X: Open Data Kit-X (formerly known as ODK-2).

[c]IT: information technology.

[d]SMTP: simple mail transfer protocol.

[e]CRF: case report form.

[f]CFR: Code of Federal Regulations.

[g]HIPAA: Health Insurance Portability and Accountability Act.

[h]GDPR: General Data Protection Regulation.

[i]FISMA: Federal Information Security Management Act.

[j]HTTPS: hypertext transfer protocol secure.

[k]CSV: comma-separated value.

[l]JSON: Javascript Object Notation.

[m]KML: keyhole markup language.

Methods

Technical Details

ConnEDCt consists of a distributed database, data entry interfaces, data management interfaces, scripted business rules, data synchronization, and data security features [6]. The main components of ConnEDCt are a client-side database, a cloud-based database server, a data synchronization engine, iOS mobile devices, and laptop personal computers. ConnEDCt is built with FileMaker Pro Advanced by Claris International Inc. [7], a commercial, cross-platform, relational database platform. FileMaker Pro Advanced provided the important advantages of rapid development and the ability to deploy cross-platform on iOS, Mac, and Windows platforms. One tool is therefore easily deployed on iPads and laptops, which are our primary computing devices. ConnEDCt's features, quality, and safety are reliable compared with existing guidelines and requirements, including Good Clinical Data Management Practices [8], the Food and Drug Administration (FDA) [9], and Title 21 Code of Federal Regulations (CFR) Part 21 [10].

Configuration of ConnEDCt

Study designers conduct a design process using Microsoft Excel templates to define the study schema. A *forms schedule* template is used to define case report form (CRF) usage over scheduled visits in a longitudinal study. The intersection of CRFs and scheduled visits in a two-dimensional matrix in this template shows which CRFs are instantiated per visit. In addition, CRFs can be defined for unscheduled instantiation. Study designers use *data dictionary* Excel templates to define variables within CRFs along with select lists, skip logic, etc. The study designers then document the eligibility criteria formulas within another Excel template.

The data dictionaries are then translated into FileMaker tables and fields in a FileMaker template file. FileMaker provides an easy-to-use interface for creating CRF tables and fields within the integrated development environment. A study designer can accomplish this step with minimal training. The final steps of implementation require some developer skills and are easily repeatable within a framework structure. These implementations include the forms schedule schema, eligibility criteria formulas, CRF validation, loading a randomization table, and synchronization. Once ConnEDCt has been configured for a study, the server files are then hosted on an internet-connected server, and client files are deployed to laptops and iPads.

We continue to develop the features of ConnEDCt further to enable study designers to complete the implementation of the majority of study protocols with minimal developer involvement. We anticipate that to maximize the flexibility of study protocols, we will always benefit from some custom or specialized development for novel EDC features.

Key Features

Management of Research Teams and Different Roles

Several *roles* were required to implement and operate ConnEDCt. *Software developers* and *systems integrators* performed programming and other technical tasks such as database and server management. *Data managers* helped translate clinical requirements into data definitions and used software interfaces to define the scheduled and unscheduled study events, CRFs, variables, validation criteria, and eligibility

criteria. Each study event involves capturing data in one or more CRFs. A CRF is a collection of related variables to be entered during a study event. The data manager also manages data exports for analysis by using external statistical and analytics software. *Research associates/assistants* perform data entry during interviews, direct observations, and also enter laboratory results. User-level privileges to various database features and data vary by role. For example, *research assistants* may create new participants and enter data, whereas research coordinators and principal investigators may additionally edit the entered data in the event of errors found as data managers. Different users can access ConnEDCt with their own account, secured by a custom username and password, which contains different levels of privileges depending on the user role.

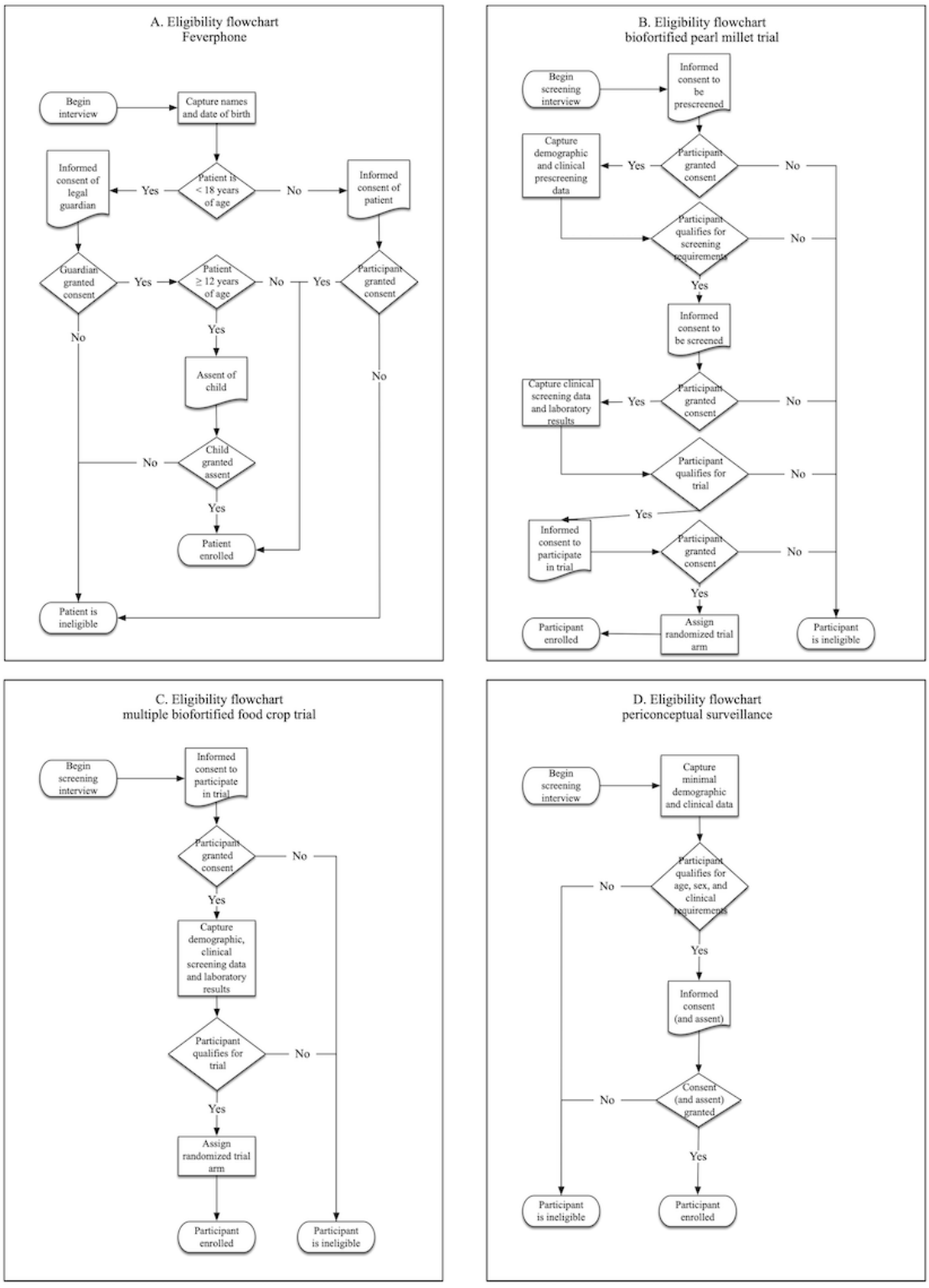
Asynchronous Data Synchronization

A key feature of ConnEDCt is the ability to capture data on a mobile device while offline and, at a later time, when an internet connection is available, send the captured data to a cloud-based server. In addition, study schema revisions can be deployed on mobile devices. The asynchronous data transport can be finely controlled such that data can be selectively deployed to individual mobile devices. In this manner, data can be synchronized in 1 direction only to minimize data traffic and to maintain data privacy. Data previously stored in mobile devices will remain in the FileMaker Go app on mobile devices to enable research assistants to track participant progress throughout the study. After study completion, the data are completely removed from mobile devices.

Real-Time Assessment of Eligibility

Assessing eligibility is a key process in participant screening and study enrollment. ConnEDCt provides automated features to determine participant eligibility based on criteria specified a priori. Eligibility algorithms are evaluated programmatically at runtime to provide feedback on participant eligibility. Figure 1 shows the workflow for evaluating the eligibility of a new participant. As eligibility criteria may be defined by including variables from different CRFs, eligibility is evaluated on completion of each predicate CRF. Participant exclusion can be determined based on a negative evaluation of a single criterion, whereas inclusion is determined on completion of all predicate CRFs.

Figure 1. Eligibility protocol for (A) clinical setting: FeverPhone, Ecuador; (B) RCT, Mumbai; (C) RCT, South India; (D) surveillance study, South India. RCT: randomized controlled trial.



Electronic Consent and Signatures

Electronic signatures and health records are regulated under the *FDA Title 21 CFR Part 11 of the CFRof the United States* [11]. *Part 11* requires strict data security, including user authentication, encryption, and auditability. Data security regulations vary by country, and ConnEDCt is compliant with currently known rules and adaptable to potential future data security requirements. For example, when study protocols require audiovisual evidence of consent, ConnEDCt can use a mobile device's built-in camera and microphone to capture the consent process.

Enforcement of Complex Study Protocols

We designed ConnEDCt to support complex, longitudinal study protocols and minimize training requirements for field research

staff. ConnEDCt provides features to support predefined longitudinal study schema, automated assignment of randomized study arms and midline serial sampling, and automated scheduling of additional CRFs based on captured responses. When ConnEDCt is configured for a study, visit records (eg, screening, baseline, midline, and endline) will be created as dictated by the study schema. CRFs are organized within appropriate visits. A visit is complete when all CRFs within the visit have been completed and signed. The interface clearly shows the schedule of visits (Figure 2) and the required CRFs for a particular visit (Figure 3). When the study includes an RCT, ConnEDCt can reference a randomization table to automatically assign a randomized arm to an eligible participant. When the study protocol includes midline serial sampling, ConnEDCt uses the internal randomization table to selectively include designated CRFs at the randomly designated midline visit (Figure 4). When responses to study questions determine the need for additional CRFs, they will be created on demand and presented in the visit schedule—for example, a response to a question on the number of pregnancies may trigger the same number of pregnancy CRFs, or a response on a sick form may trigger a blood or saliva sample CRF. These features reduce the learning curve for the research team. We believe that this improves the interview experience by allowing research assistants to focus on the study participant instead of on data management and study protocol compliance.

Designed for Flexibility and Reuse

The ConnEDCt architecture has a flexible schema and is designed to be adaptable to varying study protocols. Therefore, ConnEDCt is designed for flexibility and reuse. Events, event types, event forms, form schedules, form types, and eligibility criteria are defined in the data model and can be modified in the user interface. Validation rules and postprocessing triggers are built in a framework, minimizing the amount of custom code. In this manner, subsequent research studies can be implemented with minimal modification to the conceptual schema or programmatic code. When a new study protocol is implemented, entities for visits and CRFs are created, whereas the programming that controls participant management, navigation, eligibility, and study protocols responds to changes in study protocol design.

Language Customization and Data Location

The ConnEDCt system can also accommodate CRFs in multiple languages, as shown in Figure 5. Professionally translated questionnaires can be entered into FileMaker using native multilingual keyboards. Furthermore, data can be synchronized to and primarily stored on in-country servers, where available, to comply with local laws in many settings.

Figure 2. Schedule of visits for (A) clinical setting: FeverPhone, Ecuador; (B) RCT, Mumbai; (C) RCT, South India; (D) surveillance, South India. RCT: randomized controlled trial.

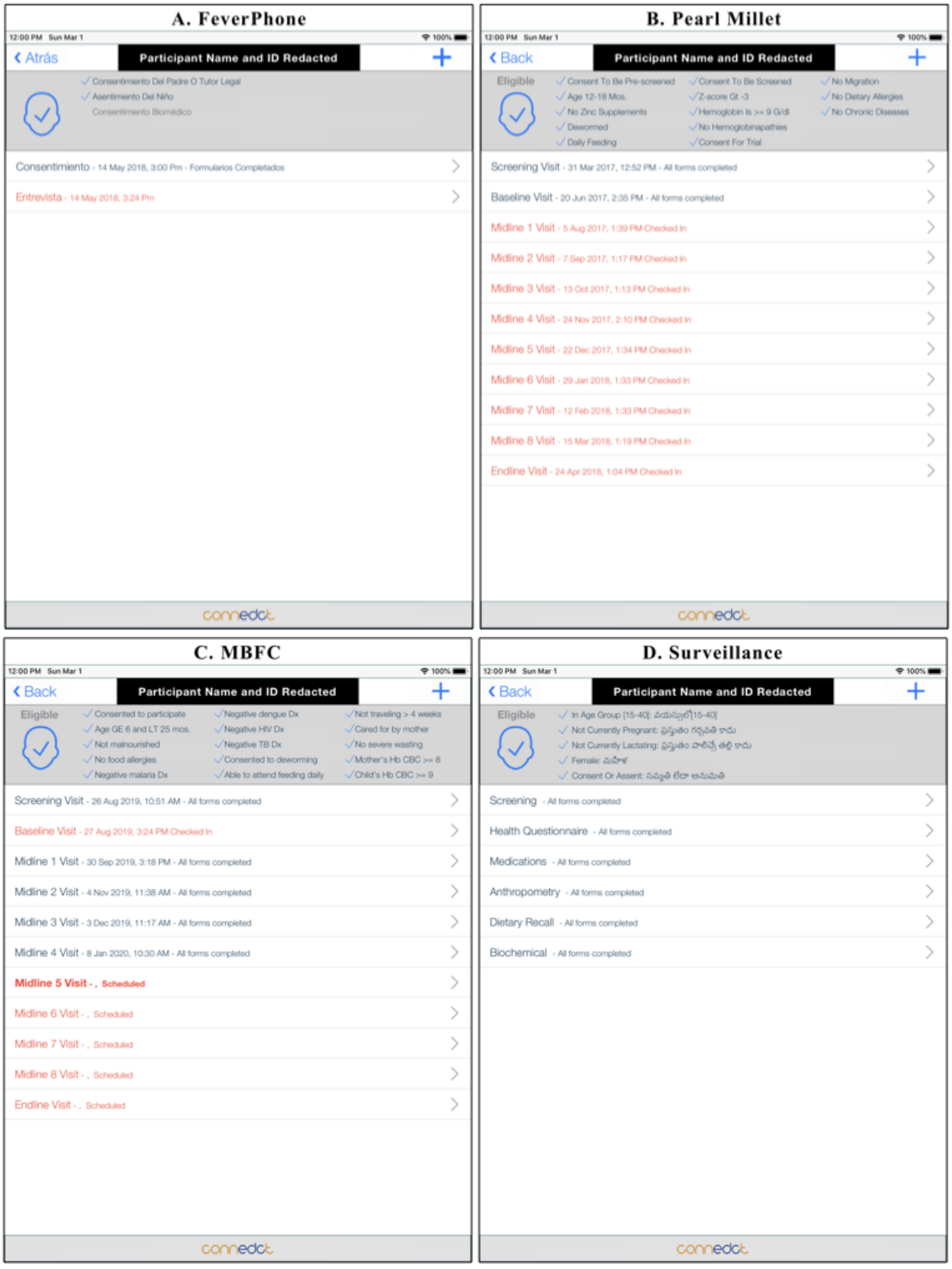


Figure 3. Schedule of visits for (A) clinical setting: FeverPhone, Ecuador; (B) RCT, Mumbai; (C) RCT, South India; (D) surveillance study, South India. RCT: randomized controlled trial.

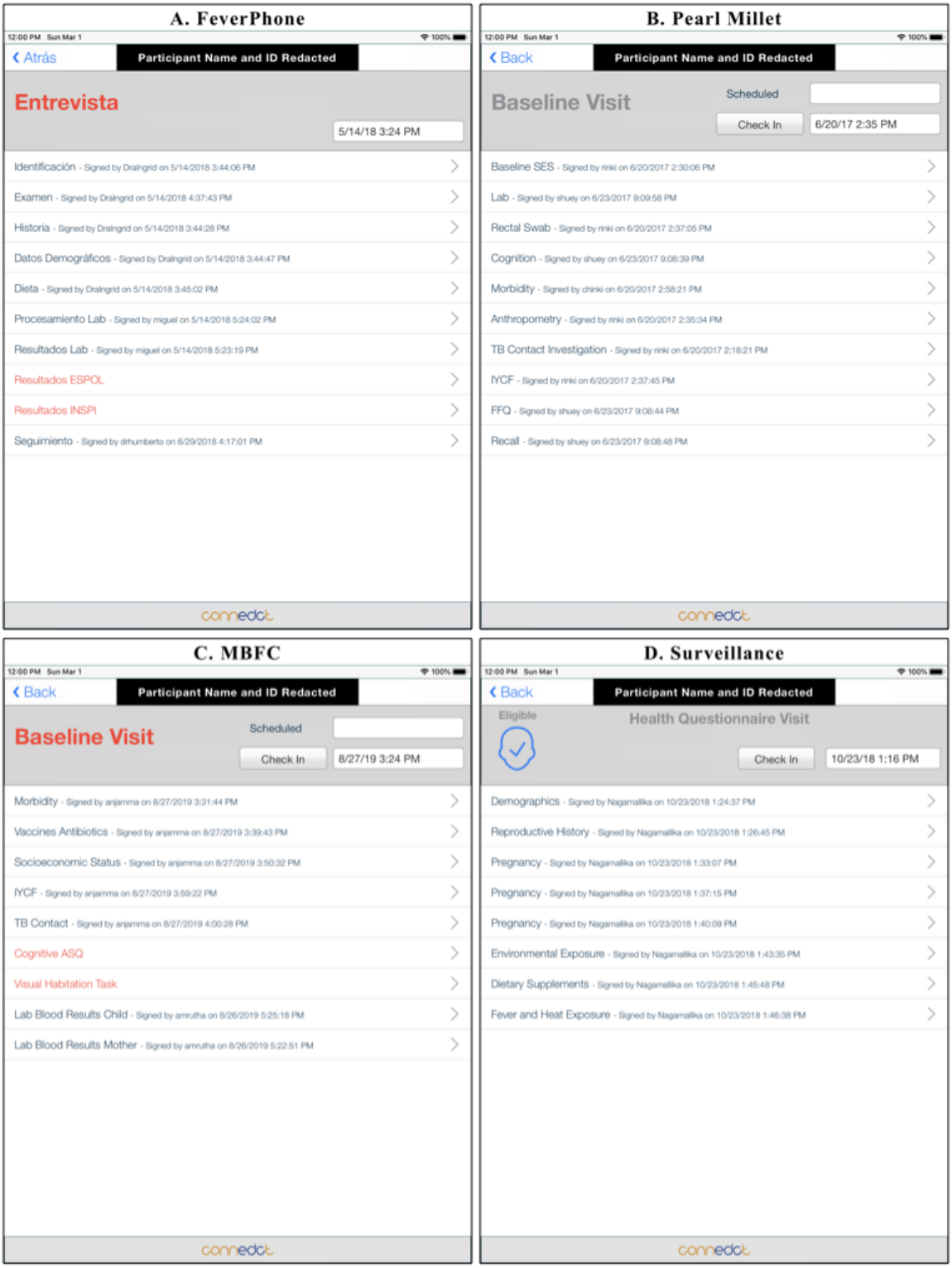


Figure 4. Example of display of all CRFs required for (A) baseline visit, for clinical setting: FeverPhone, Ecuador; (B) baseline visit for RCT, South India; and (C) health questionnaire visit for surveillance, South India. CRF: case report form; RCT: randomized controlled trial.

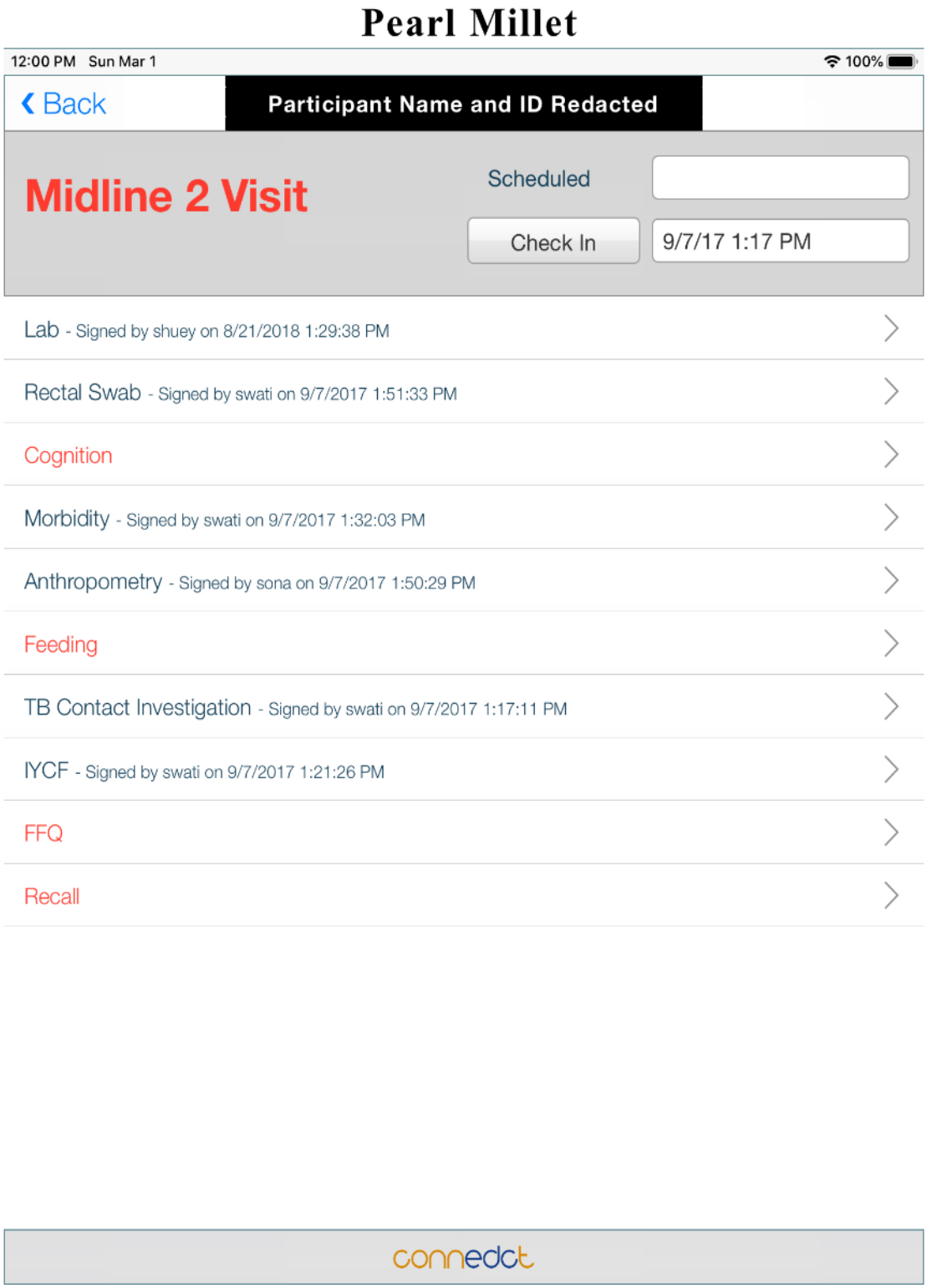
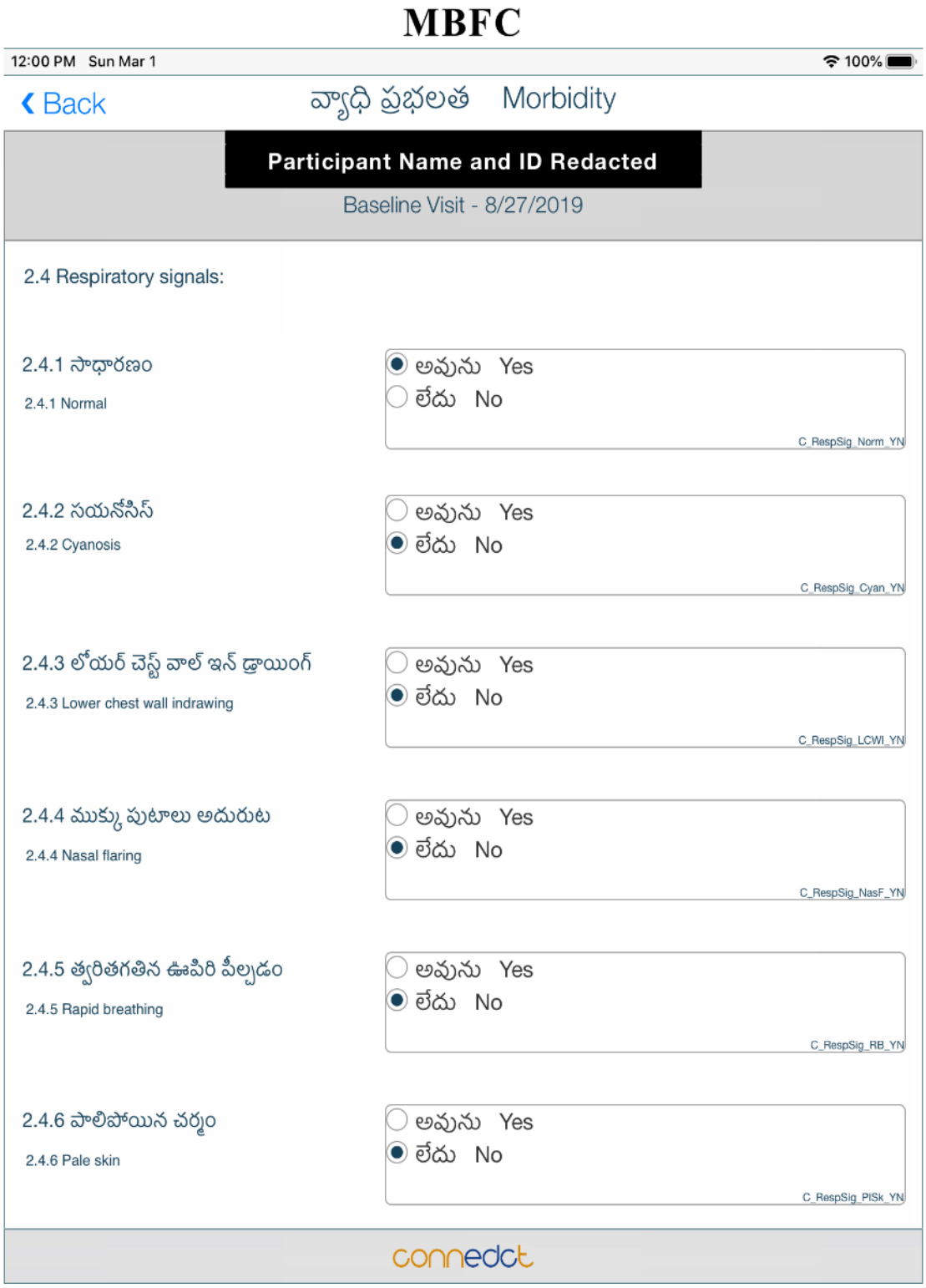


Figure 5. CRF shown in the local language, Telugu, for RCT, South India. CRF: case report form; RCT: randomized controlled trial.



Results

Description of Current Implementation: Case Studies

The ConnEDCt platform has been used in a variety of contexts within our research teams, including cross-sectional studies in clinical settings, surveillance studies (a repeated cross-sectional study in the clinic or the community, the latter surveying thousands of households), and RCTs (Table 2).

Table 2. ConnEDCt usage in case studies: clinic-based cross-sectional studies, surveillance studies, and randomized controlled trials.

Study design	Clinic-based cross-sectional study	Surveillance (repeated cross-sectional studies)	RCT[a]	RCT	RCT
Location	Ecuador	South India	Mumbai	South India	South India
Study reference	FeverPhone	Periconceptional program	Biofortified pearl millet	Multiple biofortified food crops	Quadruple fortified salts
ClinicalTrials.gov ID	N/A[b]	NCT04048330	NCT02233764	NCT02648893	NCT03853304
Status	Active	Active	Complete	Active	Planned
Synchronization methods	Continuously, using mobile SIM card, simultaneously from multiple locations	Daily, after data collection	Daily/weekly, after data collection	Daily, after data collection	Daily, after data collection
Number of participants entered into ConnEDCt	Overall 404 children and adults (including pregnant women)	2404 households (2876 women)	407 children	345 mother-infant dyads	1000 women of reproductive age
Informed consent process	Separate consent or assent forms depending on participant's age	Complex, depending on participant's age (estimated or actual)	3 forms (prescreening, screening, enrollment)	1 form, with multiple levels of consent	Complex, depending on participant's age (estimated or actual)
Eligibility	Evaluated in real time on completion of CRFs[c]	Evaluated in real time on completion of CRFs	Evaluated in real time on completion of CRFs	Evaluated in real time on completion of CRFs	Evaluated in real time on completion of CRFs
Unique features	Continuous synchronization across multiple locations simultaneously; ability for forms to be signed by multiple study personnel to trace missing information	Only requisite CRFs appear for research staff to complete; complex algorithm for study ID creation	Midline serial sample randomization scheme	2-level randomization to midline serial sample	Only requisite CRFs appear for research staff to complete; randomization scheme to establish 2×2 factorial design
Data types entered	Laboratory assay results, growth measurements, sociodemographic data, health history and current clinical signs, dietary frequency, and biological specimen collection dates and processing	Laboratory assay results; anthropometric measurements; sociodemographic data; complete reproductive history; general health history and clinical signs and symptoms; 24-hour recall; risk factors for birth defects (environmental exposures, medication use, and history of fever); and biological specimen collection volume, dates, and storage conditions	Laboratory assay results; growth measurements, sociodemographic data; infant and young child feeding; health history and clinical signs and symptoms; biological and specimen collection dates	Laboratory assay results; growth measurements, sociodemographic data; infant and young child feeding; health history and clinical signs and symptoms; and biological specimen collection dates	Laboratory assay results; anthropometric measurements; sociodemographic data; complete reproductive history; general health history and clinical signs and symptoms; 24-hour dietary recall; food frequency questionnaire; risk factors for birth defects (environmental exposures, medication use, and history of fever); and biological specimen collection volume, dates, and storage conditions
Language customization	Spanish	Telugu	Hindi	Telugu	Telugu

[a]RCT: randomized controlled trial.

[b]N/A: not applicable.

[c]CRF: case report form.

Clinical Research: Clinic-Based Cross-Sectional Studies

ConnEDCt has been the primary data collection tool for clinic-based cross-sectional studies in clinical settings of Ecuador for the National Institutes of Health-funded project *FeverPhone: Point of Care Diagnosis of Acute Febrile Illness using a Mobile Device* and was integral in the development of the Ecuador FeverPhone project protocols and manual of operations. Multiple types of users, based at different locations, input data into separate tablet computers with distinct CRFs on participant recruitment, eligibility, and informed consent; multiple participant study visits; and specimen management and laboratory results. The integration of ConnEDCt across

devices with daily synchronization of data to a cloud database and local computers allows research staff to easily review data for quality control, implement protocol revisions, and share progress with international collaborators.

Surveillance Research: Repeated Cross-Sectional Studies in a Clinic or the Community

The ConnEDCt platform provides the foundation for a periconceptional surveillance program, *Periconceptional Surveillance for Prevention of Anemia and Birth Defects in India*, funded by the US Centers for Disease Control and Prevention. This preintervention biomarker survey is being conducted among 1500 women of reproductive age (15-40 years) who are not pregnant or lactating along with their households in Southern India (ClinicalTrials.gov ID: NCT04048330; as reported in Finkelstein JL, Fothergill A, Johnson CB, et al., 2020). ConnEDCt allows enumerators to select from an imported list of households and to individually screen all women from each household while integrating all household-level data into each woman's final record. Synchronization is performed daily after data collection on all iPads. This allows for the maintenance of all household characteristics for women in the study while preventing enumerators from needing to re-enter data for all women within a household.

RCTs: Intervention Studies

The ConnEDCt platform for surveillance described above will be incorporated for an upcoming randomized efficacy trial, *A Randomized Trial of Quadruple Fortified Salt for Anemia and Birth Defects Prevention in Southern India* (ClinicalTrials.gov ID: NCT03853304; quadruple fortified salt [QFS]). This trial will be conducted among 1000 women of reproductive age who are not pregnant or lactating with their households using similar algorithms for eligibility and CRFs for data collection as the surveillance study.

ConnEDCt was also used for an RCT in Mumbai, *Effect of Iron/Zinc-biofortified Pearl Millet on Growth and Immunity in Children Aged 12–18 Months in India* (ClinicalTrials.gov ID: NCT02233764) that completed data collection in July 2018 [12]. The full protocol of the study was integrated into ConnEDCt, which included the eligibility evaluation of over 400 participants; randomization to one of 2 experimental arms, accounting for a random midline serial sample; and a diverse set of CRFs to collect data from participants at 9 monthly follow-up visits. After a training session, local research assistants independently scheduled visits and collected data throughout the trial, including synchronization to the secure server. Importantly, ConnEDCt was also able to be modified after pilot testing among research assistants and participants in the field to improve the workflow, such as improving the placement of certain buttons or options on each form screen. Real-time feedback of data throughout the trial allowed for error checking concurrently with data collection. Specifically, data were exported from ConnEDCt as a relational database in the form of an Excel spreadsheet (1 per each type of CRF) using a custom script. These data were then imported and merged into 1 full database in statistical software packages and subsequently analyzed for errors, such as biologically implausible values or typographical errors during data entry. One asset of the ability to analyze the data in SAS concurrently with the trial permitted examination of the raw variables for a particular eligibility criterion necessitating advanced statistical analysis, after which results could be re-entered into ConnEDCt. The ability to access the full database concurrently as the trial was ongoing was also crucial to summarizing data, such as technical reports to the study funders and progress reports, as well as any adverse events for our data safety and monitoring board.

The EDC framework established for the Mumbai trial was customized and expanded for a second ongoing RCT evaluating multiple biofortified food crops (MBFC) in children and their mothers, *Effect of a Biofortified Food Basket on Micronutrient Status and Immune and Cognitive Function among Infants in India* in Madanapalle, South India (ClinicalTrials.gov ID: NCT02648893) to incorporate a different type of participant (mother-child dyads) and CRFs specific to the trial. The protocol was modified to include (1) a 2-level randomization to 1 of 4 color-coded groups, with 2 groups corresponding to each treatment arm and randomization to a midline sampling point; (2) visit scheduling and tracking over a 9-month follow-up period; and (3) data collection across integrated CRFs to handle screening and eligibility determination, sample collection, and participant tracking. A randomization protocol was created that was executed by the study statistician and shared with the ConnEDCt programmer to incorporate into the database. Similar to Mumbai, the MBFC trial is synchronizing daily after data collection during screening, baseline, and follow-up appointments.

Errors During Implementation

During implementation of data capture, the more common errors included inputting the *wrong* CRF for a participant at the wrong visit or miscoding variables (eg, a numeric coding for a character variable). These were easily resolved by the database developer and/or data manager.

Discussion

Principal Findings

The benefits of using ConnEDCt as an EDC system greatly outweigh the few challenges. Below, we describe how we addressed potential or actual issues of connectivity, complexity in consent forms, participant confidentiality, and research staff compliance.

Problem 1: Connectivity

In all settings, data were synchronized to the server and the participant's data were harmonized by the MirrorSync software [5]. To account for lapses in internet connectivity and prevent data conflicts, synchronization was performed at the end of each day of data collection in the RCTs to avoid data conflicts and any need for internet connectivity throughout the data collection process. In practice, 1 iPad was assigned to a particular participant for the duration of the visit and, after completion of the visit, all iPads were synced to the data server.

In contrast to RCTs, the clinic-based cross-sectional study in Ecuador (FeverPhone) required multiple users to input data for

a single participant using separate devices in different clinical locations. Therefore, regular synchronization was important for study procedures. The clinical site did not offer wireless internet access, but each iPad was fitted with a prepaid mobile SIM card to allow for regular synchronization during participant visits.

For the periconceptional surveillance program, the database was designed such that iPads could remain at 1 station for data collection while participants moved from station to station. This greatly simplified data collection and allowed each enumerator to continue using the same iPad to help track enumerator productivity and workflow.

Problem 2: Complex Consent Forms

In Mumbai's RCT, informed consent was taken at 3 separate visits: prescreening (consent to noninvasive data collection), screening (consent to biospecimen collection and growth measurements), and enrollment (consent to participate in the trial), including informed consent signatures per the type of data collected. During follow-up, we included informed consent signatures on biological specimen collection forms to remind participants of their option to opt out of the trial at any time.

One RCT in Southern India, the MBFC trial, combined the informed consent process into a single form, with multiple levels of consent. Participants were asked separately for their consent to screening and participation in the trial, subsequent biological sample storage and analysis, and future analyses of their biological samples, which may include genetic analysis. This allowed documentation and tracking of different levels of consent for each participant as well as allowing for tracking of which participants consented to future analyses.

At the Ecuador clinical site, study procedures required separate consent or assent forms depending on the participant's age. ConnEDCt automatically provided the appropriate consent or assent forms based on the date of birth information provided in the recruitment CRF. ConnEDCt would further limit progress to subsequent CRFs if the appropriate consent or assent forms were incomplete.

The periconceptional surveillance study in Southern India has a complex informed consent process because of multiple factors influencing eligibility and additional factors to determine which set of informed consent forms need to be completed. In the periconceptional surveillance program, participants included adolescents aged 15 to 18 years (requiring an assent form and parental consent) and adults (≥18 years of age, requiring a consent form) are potentially eligible to participate. To ensure that participants complete the correct forms and that all required signatures are obtained, the participant's age (to the day) is required in real time. ConnEDCt was able to be adapted for this purpose to determine a participant's eligibility, and which set of forms they needed to complete to enroll. Additionally, to account for variable amounts of information being available for the determination of current age, we developed an algorithm that considered all available information (current age, exact date of birth, and estimated/partial date of birth). The database developer built the algorithm into ConnEDCt such that research assistants entered data available to them, and the software made visible the CRFs that they needed to proceed (eg, CRFs indicating the participant was eligible and which set of assent/consent forms they needed to sign). To simplify and streamline the process, we adapted ConnEDCt such that it only allowed necessary CRFs to appear on the iPad for the research assistants, making it impossible for research assistants to fill out the incorrect set of CRFs.

Problem 3: Confidentiality

Confidentiality was maintained in RCTs, clinic-based cross-sectional studies, and repeated cross-sectional studies, such as our surveillance program, through many layers of security. First, each field personnel had their own computer-generated unique username and password to access the database. Second, although participant names were included in the database to facilitate tracking of each participant locally, names were easily removed for deidentification after exporting the data. Participants were identified with ID numbers derived from a 4-digit sequentially generated code and a letter corresponding to the participant's color group study allocation.

The periconceptional surveillance program includes women of reproductive age and their families—and a hierarchical data structure. In the database, each participant has an identifying individual participant ID and a household ID, which in combination, uniquely identifies them as a participant in the periconceptional surveillance program; the latter being generated by ConnEDCt. Their personal ID incorporates a code that indicates that they are a participant in the surveillance program, a code indicating which iPad their data were collected on, and an additional 4-digit code (generated sequentially by ConnEDCt) to ensure that all participants have unique IDs. This identification system in ConnEDCt allows for an undetermined number of participants to be enrolled in the surveillance program while ensuring that no duplicate IDs are utilized. The structure of the IDs (3 concatenated codes) also provides flexibility to add and remove pieces to indicate the person's participation in potential future studies led by the Finkelstein research group in this area (eg, for the upcoming QFS randomized efficacy trial).

Problem 4: Compliance

The data collected in the periconceptional surveillance program are very detailed and complex. To improve data quality during collection, we built various safeguards into the data collection tools. For free-entry numeric variables, limits were placed around values entered that would flag responses and indicate enumerators that data entered seemed implausible. For example, if a participant's weight was entered as 1 kg, a flag would pop up on the iPad screen, indicating that the response was outside the anticipated range and ask the enumerator to verify the entry before moving forward. In addition to raw variable entry, these flags were also built in to read across multiple variables that were deemed to be of high importance. If a participant's response was left blank, or if a response to question 1 directly contradicted with an entry for question 3, a flag would pop up and ask the enumerator to verify before continuing. Finally, various time stamps were built into variables throughout the questionnaire to allow for remote monitoring of enumerator progress. Enumerators had to log in using their unique username and password before any data entry and had to sign each form

they entered data on using their unique username and password before they could continue.

The RCTs incorporated similar compliance measures into the ConnEDCt EDC system, including real-time calculations for replicate measurements, thresholds for variation to prompt additional replicates, and flagging of responses outside of expected ranges. Adaptive and forced protocols are also implemented into CRFs. For example, anthropometric measurements for the first replicate need to be entered before the next replicate fields will be displayed, meaning a full replicate of the measurement protocol is followed instead of basic remeasurements with participants and equipment in the same position. Adaptive protocols and instructions are also included. For example, for certain anthropometric measurements such as triceps skinfold, the research assistant determines the participant's dominant arm, triggering ConnEDCt to prompt the research assistant to use the participant's other, nondominant arm. Furthermore, warning pop-ups (logic checks) for impossible or nonbiologically plausible data (such as body weight entered or calculated as a negative number) were incorporated into ConnEDCt, allowing real-time data correction during data collection, minimizing challenges in ascertaining errors during data cleaning and analysis. Queries, raised by research assistants and/or data managers concurrently with the study or after data collection is completed, can be settled by data managers who have form modification privileges. ConnEDCt includes an audit trail to monitor when data were modified and by whom.

Conclusions

ConnEDCt, an EDC system, is an ideal tool for research studies, particularly those with complex protocols in settings where internet access is limited. In addition to mitigating the time required and error-prone nature of paper-based data collection methods, ConnEDCt represents a fixed framework that is adaptable to a variety of study designs. As demonstrated in a variety of settings, the ConnEDCt EDC system has been integral to carrying out several studies, all diverse in design and setting, types of participants, and overall goals for the research. Compared with other EDC tools [1,2,13-24], ConnEDCt's benefits, including its utility as a mobile system, the ability to collect data without internet access, customization options for specific study designs, and data security, are comparable with systems such as REDCap and other systems [3,25,26], while at the same time serving complex protocols more precisely than other systems. An EDC system can be made as a straightforward framework that is adaptable for the successful management and completion of almost all kinds of field-based research studies and allows for easy export and transfer of collected data into statistical processing software for further analyses.

Acknowledgments

The authors gratefully acknowledge Guanglan Zhang, Associate Professor of Computer Science at Boston University, who provided guidance and advice; Elaine A Yu, PhD from Cornell University, who designed the forms and piloted the initial ConnEDCt design in South India; Françoise Vermeylen from Cornell University for her expertise in generating the randomization list for our studies; our intern study coordinators from our RCTs in Mumbai, Laura Powis and Yadurshini Raveendran for field testing and troubleshooting; and, finally, all patients, participants, and research assistants and community health workers from our studies in Mumbai, South India, and Ecuador. SM and JF acknowledge support from HarvestPlus and the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under Award Number R01EB021331 for the research studies in which ConnEDCt was initially implemented. AF and CJ were supported by the National Institutes of Health under award 5 T32 HD087137. The content in this paper is solely the responsibility of the authors and does not necessarily represent the official views of National Institute of Biomedical Imaging and Bioengineering, the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health. The publication costs of this manuscript were supported by the Bill and Melinda Gates Foundation through their open access fund.

Conflicts of Interest

SM is an unpaid board member for a diagnostic start-up focused on developing point-of-care assays for nutritional status informed by his research as a faculty member at Cornell University. CR is the owner of Data Performance LLC, which owns the trademark and commercial rights to the ConnEDCt software. All other authors have no competing interests to declare.

References

1. Walther B, Hossin S, Townend J, Abernethy N, Parker D, Jeffries D. Comparison of electronic data capture (EDC) with the standard data capture method for clinical trial data. PLoS One 2011;6(9):e25348 [FREE Full text] [doi: 10.1371/journal.pone.0025348] [Medline: 21966505]
2. Malik I, Burnett S, Webster-Smith M, Morden J, Ereira S, Gillman A, et al. Benefits and challenges of electronic data capture (EDC) systems versus paper case report forms. Trials 2015 Nov 16;16(S2):-. [doi: 10.1186/1745-6215-16-s2-p37]
3. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
4. Hartung C, Lerer A, Anokwa Y, Tseng C, Brunette W, Borriello G. Open Data Kit: Tools to Build Information Services for Developing Regions. In: Proceedings of the 4th ACM/IEEE International Conference on Information and Communication

Technologies and Development. 2010 Presented at: ICTD'10; December 13-16, 2010; London, United Kingdom. [doi: 10.1145/2369220.2369236]
5. 360Works MirrorSync: FileMaker Sync for FileMaker Server, FileMaker Pro, and FileMaker Go on iPhone and iPad, FileMaker offline sync. URL: https://www.360works.com/filemaker-sync.html [accessed 2020-06-30]
6. Ruth C, Yu E, Huey S, Chitkushev L, Mehta S, Zhang G. ConnEDCt: A Development Framework for Mobile Electronic Data Capture in Disconnected Communities. In: 12th Annual International Conference on Computer Science and Education in Computer Science. 2016 Presented at: SIGCSE'16; July 1-4, 2016; Nürnberg, Germany URL: https://csecs.org/conference-schedule-2/
7. Claris FileMaker—tackle any task. URL: https://www.claris.com/filemaker/ [accessed 2020-07-15]
8. Cornu C, Donche A, Coffre C, le Gouge A, Rym B, Vaugier I, Comité de pilotage du groupe de travail harmonisation des procédures de réseau des CIC. [ECRIN standard requirements for good clinical practices-compliant data management in multinational clinical trials]. Therapie 2016 Oct 13:- epub ahead of print. [doi: 10.2515/therapie/2015042] [Medline: 27839710]
9. Computerized Systems Used in Clinical Investigations. US Food and Drug Administration. 2007. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computerized-systems-used-clinical-investigations [accessed 2020-06-26]
10. Bansal A, Chamberlain R, Karr S, Kwasa S, McLaughlin B, Nguyen B, et al. A 21 CFR part 11 compliant graphically based electronic system for clinical research documentation. J Med Syst 2012 Jun;36(3):1661-1672. [doi: 10.1007/s10916-010-9627-4] [Medline: 21107888]
11. CFR - Code of Federal Regulations Title 21. US Food and Drug Administration. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm [accessed 2020-06-26]
12. Mehta S, Finkelstein JL, Venkatramanan S, Huey SL, Udipi SA, Ghugre P, et al. Effect of iron and zinc-biofortified pearl millet consumption on growth and immune competence in children aged 12-18 months in India: study protocol for a randomised controlled trial. BMJ Open 2017 Nov 14;7(11):e017631 [FREE Full text] [doi: 10.1136/bmjopen-2017-017631] [Medline: 29138201]
13. Aroda VR, Sheehan PR, Vickery EM, Staten MA, LeBlanc ES, Phillips LS, D2d Research Group. Establishing an electronic health record-supported approach for outreach to and recruitment of persons at high risk of type 2 diabetes in clinical trials: the vitamin D and type 2 diabetes (D2d) study experience. Clin Trials 2019 Jun;16(3):306-315 [FREE Full text] [doi: 10.1177/1740774519839062] [Medline: 31007049]
14. Barton J, Young A, Lay M. Introduction of electronic data capture method using participant-completed online web-based follow up questionnaire in mail-based study achieves expected benefits and positive participant feedback. Trials 2015 Nov 16;16(S2):-. [doi: 10.1186/1745-6215-16-s2-p44]
15. Bingham CO, Gaich CL, DeLozier AM, Engstrom KD, Naegeli AN, de Bono S, et al. Use of daily electronic patient-reported outcome (PRO) diaries in randomized controlled trials for rheumatoid arthritis: rationale and implementation. Trials 2019 Mar 22;20(1):182 [FREE Full text] [doi: 10.1186/s13063-019-3272-0] [Medline: 30902094]
16. Drew CJ, Poile V, Trubey R, Watson G, Kelson M, Townson J, et al. Integrating technology into complex intervention trial processes: a case study. Trials 2016 Nov 17;17(1):551 [FREE Full text] [doi: 10.1186/s13063-016-1674-9] [Medline: 27855710]
17. Ene-Iordache B, Carminati S, Antiga L, Rubis N, Ruggenenti P, Remuzzi G, et al. Developing regulatory-compliant electronic case report forms for clinical trials: experience with the demand trial. J Am Med Inform Assoc 2009;16(3):404-408 [FREE Full text] [doi: 10.1197/jamia.M2787] [Medline: 19261946]
18. Haynes R. Minimising paper: electronic data capture methods in the 3c study. Trials 2013 Nov 29;14(S1):-. [doi: 10.1186/1745-6215-14-s1-p60]
19. Hutton D, Smith N, Cappel-Porter H, Saw C, Rogers C. Development of the 'GeneSYS' database system to support trial data capture and conduct. Trials 2013;14(Suppl 1):P66. [doi: 10.1186/1745-6215-14-s1-p66]
20. Lewis R, Batten L, Friend C, Webster-Smith M, Burnett S, Morden J, et al. Design considerations when transitioning from paper case report forms (CRFS) to electronic data capture (EDC). Trials 2015 Nov 16;16(S2):-. [doi: 10.1186/1745-6215-16-s2-p35]
21. Seebregts CJ, Zwarenstein M, Mathews C, Fairall L, Flisher AJ, Seebregts C, et al. Handheld computers for survey and trial data collection in resource-poor settings: development and evaluation of PDACT, a palm pilot interviewing system. Int J Med Inform 2009 Nov;78(11):721-731. [doi: 10.1016/j.ijmedinf.2008.10.006] [Medline: 19157967]
22. Webster-Smith M, Paulding C, Burnett S, Jeffs L, Ismail D, McNamara C, et al. Implementing electronic data capture (EDC) training for site staff. Trials 2015 Nov 16;16(S2):-. [doi: 10.1186/1745-6215-16-s2-p41]
23. Yu P, de Courten M, Pan E, Galea G, Pryor J. The development and evaluation of a PDA-based method for public health surveillance data collection in developing countries. Int J Med Inform 2009 Aug;78(8):532-542. [doi: 10.1016/j.ijmedinf.2009.03.002] [Medline: 19369114]
24. Zhang J, Sun L, Liu Y, Wang H, Sun N, Zhang P. Mobile device-based electronic data capture system used in a clinical randomized controlled trial: advantages and challenges. J Med Internet Res 2017 Mar 8;19(3):e66 [FREE Full text] [doi: 10.2196/jmir.6978] [Medline: 28274907]

25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
26. Elliot R, Sawkar M, Williams M. The Power of Mobile Platorms for Data Collecion. The Agriculture, Nutrition & Health (ANH) Academy. 2019. URL: https://www.anh-academy.org/sites/default/files/SightandLifeMagazine_2019_Data_in_Nutrition_ThePowerofMobilePlatformsforDataCollection_0.pdf [accessed 2020-06-26]

Abbreviations

CFR: Code of Federal Regulations
CRF: case report form
EDC: electronic data capture
FDA: Food and Drug Administration
MBFC: multiple biofortified food crops
QFS: quadruple fortified salt
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 05.03.20; peer-reviewed by J Buttolph, P Zhang, J Zhang; comments to author 12.04.20; revised version received 24.04.20; accepted 14.05.20; published 13.08.20.

Please cite as:
Ruth CJ, Huey SL, Krisher JT, Fothergill A, Gannon BM, Jones CE, Centeno-Tablante E, Hackl LS, Colt S, Finkelstein JL, Mehta S
An Electronic Data Capture Framework (ConnEDCt) for Global and Public Health Research: Design and Implementation
J Med Internet Res 2020;22(8):e18580
URL: https://www.jmir.org/2020/8/e18580
doi:10.2196/18580
PMID:32788154

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org